Series on Innovation and Knowledge Management - Vol. 5

Creating Collaborative Advantage Through Knowledge and Innovation

Editor

Suliman HAWAMDEH

Creating Collaborative Advantage Through
Knowledge and Innovation

Series on Innovation and Knowledge Management

Series Editor: Suliman Hawamdeh *(University of Oklahoma)*

ISSN: 1793-1533

Vol. 1 Managing Strategic Enterprise Systems and e-Goverment Initiatives in Asia: A Casebook
edited by Pan Shan-Ling *(National University of Singapore)*

Vol. 2 Knowledge Management: Through the Technology Glass
by Meliha Handzic *(University of New South Wales)*

Vol. 3 Governing and Managing Knowledge in Asia
edited by Thomas Menkhoff *(Singapore Management University)*, Hans-Dieter Evers *(Bonn University)* and Yue Wah Chay *(Nanyang Technological University)*

Vol. 4 Managing Emerging Technologies and Organizational Transformation in Asia: A Casebook
edited by Pan Shan-Ling *(National University of Singapore)*

Vol. 5 Creating Collaborative Advantage Through Knowledge and Innovation
edited by Suliman Hawamdeh *(University of Oklahoma)*

Series on Innovation and Knowledge Management - Vol. 5

Creating Collaborative Advantage Through Knowledge and Innovation

Editor

Suliman HAWAMDEH

Oklahoma University, USA

World Scientific

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

Library of Congress Cataloging-in-Publication Data

International Conference on Knowledge Management (3rd : 2006 : Greenwich (London, England))

Creating collaborative advantage through knowledge and innovation / edited by Suliman Hawamdeh.

p. cm. -- (Series on innovation and knowledge management ; vol. 5)

ISBN-13 978-981-270-451-1 -- ISBN-10 981-270-451-5 (alk. paper)

"A selection of papers from the 2006 International Conference on Knowledge Management held in Greenwich, London"--P. .

Includes bibliographical references and index.

1. Knowledge management--Congresses. 2. Knowledge workers--Social networks--Congresses. 3. Technological innovations--Management--Congresses. I. Al-Hawamdeh, Suliman. II. Title. III. Series.

HD30.2 .I5566 2006
658.4'038--dc22

200605250

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

Printed in Singapore.

CONTENTS

This page intentionally left blank

ACKNOWLEDGMENTS

I would like to take this opportunity to thank the Organizing Committee and the International Advisory Board of ICKM2006 for their contributions to the review process and selection of papers. In particular, I would like to thank Martyn Laycock and his team for organizing the event, and preparing an exciting program that kept people engaged, while facilitating an atmosphere to network and build friendships. Special thanks goes to the University of Greenwich Business School for hosting the Third International Conference on Knowledge Management (ICKM2006) on the historic Greenwich Maritime World Heritage Site on the banks of the River Thames London. Matthew Khoe from Greenwich Business School and his team are to be commended for their work on providing technical support and maintaining the website.

The accomplishment of this book relied on the excellent contributions of the authors. I would like to acknowledge with gratitude the insights and valuable contributions provided by the authors. Also, I would like to thank Michelle Farabough for editing the conference papers and coordinating this effort with authors and the publisher, Matthew Broaddus for helping with the paper review process and World Scientific staff for their support in producing the final product.

This page intentionally left blank

PREFACE

The field of knowledge management is relatively new and evolving. What makes it attractive to a wide range of audience is its relevance to today's competitive business environment, as well as its association with disciplines such as information technology, communication, cognitive science, information science, engineering, business, and management. Knowledge management can be viewed as an interdisciplinary subject dealing with various aspects of knowledge processes such as knowledge creation, discovery, capture, sharing, organization, retention, and utilization. Some of these processes are technical in nature and require technical skills that need to come from technical disciplines like information technology, information science, and engineering. Others are considered human centric and require soft skills that need to come from disciplines like communication, cognitive science, business, and management.

The divergence between a practitioner's and an academic's approach to knowledge management is very important to the development and future of the profession. Very often practitioners perceive academics as mostly dealing with basic research and theoretical work that might not have direct applications to real world problems. The ivory tower concept, which is often used to describe academia, symbolizes detachment and it is an indication of someone losing touch with reality and not being able to relate theories to practice. On the other hand, academics might sometimes perceive practitioners' work as being shallow and lacking the theoretical and scientific foundation. Such perceptions could be problematic for emerging disciplines that need to fuse theory with practice. In an emerging field like knowledge management and in the absence of a defined professional qualification, anyone can claim to be an expert in the field. It is important to understand that the role of academia is not only in eliminating illiteracy but also providing graduates with the foundational knowledge and the necessary skills that can gain them entry to their chosen area of practice.

For knowledge management as a discipline to succeed, it needs to draw upon the support of many theoretical and methodological areas with pragmatic considerations of expertise required to conduct business. Keeping that in mind, the International Conference on Knowledge Management (ICKM) started in 2002 with the objective of bringing the academics and practitioners together to share knowledge and exchange ideas. The conference's aim is to encourage collaboration and address issues relevant to today's pressing problems, while delivering tangible benefits to both communities. The outcome, which can be measured through the presentations, publications, and feedback, is a testimony of the benefits of having both communities working in concert.

The collection of papers included in this book from the 2006 International Conference on Knowledge Management, held in Greenwich, London, represents some of the best work by researchers and practitioners in the field of knowledge management. Their subject matter covers a wide range of topics, including: social network analysis and technologies; innovation and creativity; KM tools and technologies; collaboration and knowledge sharing; issues in KM education and training; knowledge discovery (data mining, data warehousing, intelligent agents); knowledge organization (meta data, taxonomies, ontology); and social and psychological dimensions. This book will appeal to information and knowledge management professionals, as well as academicians, practitioners, and researchers who are looking for a deeper understanding of knowledge management research and its practical applications.

Suliman Hawamdeh, Ph.D
Professor and Program Coordinator
University of Oklahoma

Chapter 1

THE BUSINESS TRANSACTION THEORY AND MORAL HAZARDS FOR KNOWLEDGE SHARING: AN EMPIRICAL STUDY

Franz Barachini
University of Technology Vienna & BIC-Austria
Institut für Informationssysteme
Favoritenstrasse 9-11
A-1040 Vienna, Austria
Barachini@bic-austria.at

Individuals don't offer information (knowledge) for free. Therefore, knowledge sharing can be regarded as a business transaction process. During this process humans use a tacit but probably unique function--independent from cultural roots--to evaluate the value of information. After conducting a comprehensive company survey in Europe, we found indicators supporting the business transaction theory. Additionally, we selected a subset of companies and asked employees their thoughts about the motivators for knowledge sharing and working performance. In so doing we performed a cluster analysis and mapped the answers to Alderfer's pyramid. Very important cultural-dependant moral hazards for knowledge sharing were detected.

1. Motivation

Knowledge management is not only an IT challenge; foremost it is discovering how to motivate people to share valuable information so that intellectual capital of a company can be leveraged. Bontis (2002), Edvinsson and Malone (1997) and Sveiby (1997) see intellectual capital as the "stock" of knowledge that exists in an organization at a particular point in time. Managing this stock remains a challenge, as there is the need to socialize and codify tacit knowledge. Furthermore, we found knowledge acquisition was only successful when people were willing to cooperate. Willingness to cooperate, in turn, is strongly dependent on the trust level (Huener et al., 1998) in an organization. And it is not only the trust level that is important; it is the value of the information itself that plays a major role during information (knowledge) exchange.

Barachini (2003) developed a thought model, which maps the information exchange process between humans to the investment processes of the modern portfolio theory. He argues that knowledge always has been the cornerstone for mankind to survive. Therefore, in his opinion, individuals don't offer information (knowledge) for free. To establish a successful knowledge-sharing culture an organization must especially consider trading aspects of modern portfolio theory and refrain from being exclusively dependent on trust, attitude, leadership, and group support. In the company survey presented herein we found indicators supporting the business transactions theory. We also identified moral hazards, which hamper knowledge exchange within a society. It is important to note that parts of the presented results strongly depend on European culture and cannot be generalized as such.

2. Background of the Business Transaction Theory

Barachini (2002) defined two types of information exchange. Type-1 is the immediate exchange of information in both directions. Thus, sender and receiver give information away. This type of duplex information exchange can be mapped to over-the-counter businesses transactions executed by banks.

Type-2 is more complicated because information flow is, first of all, unidirectional. This concept is better defined in two scenarios: 1) when we consider the fact that we earn money by way of our profession as e.g. a teacher or 2) when we consider that we offer information to individuals, investing in hopes to receive even more valuable information in return at some future date. Type-2 of information exchange can be mapped to the most prominent type of option contracts--the call option for stocks. This agreement gives the buyer the right to buy from the option writer a specific number of shares of a particular company at a specific purchase price at any time[1] up to and including a specific date.

[1] For US options only.

Figure 1 shows the P&L graph[2] of a buyer. The buyer of a call option will have to pay the writer a premium in order to get the writer to sign the contract. The fair value of an option can be evaluated by the binomial option-pricing model or by the more modern method from Black-Scholes (Sharpe et al., 1995):

Fair value = N(d1)*Ps – E*N(d2)/e^{RT}

Where: d1 = (ln(Ps/E) + (R + 0.5σ^2)T)/ σ*sqrtT, d2 = d1 - σ *sqrT

Ps = Current market price of underlying stock
E = Exercise price of option
R = Compound risk free rate of return
T = Time remaining before expiration
σ = Risk of the underlying stock
sqr = square root

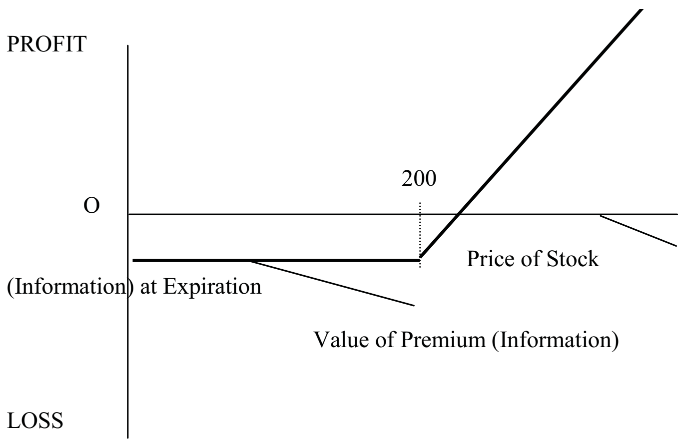


Figure 1. P&L graph for "buy a call".

[2] Profit and Loss.

Figure 1 relates the value of a call option with an exercise price of 200 to the price of the underlying stock of expiration. If the stock price is bellow 200, the option will be worthless when it expires, and the writer will gain the premium. If the price is above 200, the option can be exercised for 200 in order to obtain a security with a greater value than 200. As a result the option buyer will realize a net gain that will equal the difference between the securities market price and the 200-exercise price. However, in practice the calculations are even more complicated due to margin requirements, commission payments, and other market-making activities.

Type-2 information exchange describes the process by which one person (the buyer) gives information away, hoping to get even more valuable information in the future. The information offered to the writer has some value--the premium. The buyer invests in hopes he will receive in return another type of information that is at least as valuable as the information premium he gave. For our purposes, the underlying asset is not stock but again it consists of information. Following the analogy of this theory, then, the person who delivers information is the buyer of a call option.

The difficulty lies in determining how to evaluate a fair price for a piece of information which is yet unknown. The Black-Scholes formula is based on statistics, whereby the exercise price is known, the risk of the underlying common stock can be evaluated, and the option has a well-defined expiration date[3]. In the case of information brokerage, we don't know even the value of the underlying because it is an unknown piece of information that might be offered from the writer at a future time. In the Black-Scholes formula the current market price of the underlying stock can be evaluated. Since one type of information is evaluated differently from brain to brain, no objective evaluation can be performed for information generated by humans.

Thus, each of us uses our own evaluation function, which might be similar from brain to brain; however, due to different context knowledge, e.g. experience or intuition, the same piece of information is evaluated differently on an individual basis. Therefore statistics like those in the

[3] This is true for European options – US options can be exercised arbitrarily.

Black-Scholes formula cannot be applied immediately since the values of Ps, E, R and T represent individual functions. The parameter T is indeterminable since we don't know when and even if we will receive valuable information in the future. Thus, a fair price for information cannot be calculated. Nevertheless, the P&L statement of a call option can be used as a thought model when we talk about information exchange[4] between humans. By applying a very specific survey we hoped to find justifications for the business transaction theory.

3. The Method

We selected 150 companies in Europe[5] and asked each to select ten employees[6] to participate in an electronic questionnaire (see Figure 2).

CODE	QUESTION What is your motivation to exchange information with colleagues in your company? Please distribute scores between 0 (low) – 12 (high)	Score 0-12
Q1	Justification or refutation of personal perceptions	
Q2	More acknowledgement and better acceptance of my person and my ideas	
Q3	As part of a network I need to communicate (rumors, news, needs)	
Q4	I need it because of therapeutical reasons, will get sick otherwise	
Q5	I need it to learn from each other	
Q6	I need it because I have a desire to show off	
Q7	I am dependant on information and sometimes forced to use it	
Q8	To built up trust	
Q9	I am curios	
Q10	I want to reach my own goals	
Q11	I want that my group reaches its goals	

Figure 2. Questionnaire for the online survey.

[4] Type-2 information exchange.
[5] Germany, Austria, Switzerland.
[6] Management & Employees in total 1.500 persons.

Using the online survey, we asked 1,500 people to score on a continuum between 0 and 12 their response to eleven separate statements about the motivation for information exchange within companies.

In the second phase of research we created a focus group with participants randomly selected from 40 of the 150 companies used in the online survey. A structured focus group interview protocol was developed, and two researchers conducted each of the 40 direct interviews, soliciting answers to open questions. The motivation for the open interviews was twofold. Firstly, we reassured ourselves that the respondent understood the electronic questionnaire, and that our interpretation of their answer matched their intent. Secondly, we tried to identify motivators and hazards for the working performance of employees. In so doing, we performed a cluster analysis and mapped the answers to the Alderfer's pyramid. The results reflect the current fears and hopes of the Middle European culture in its worldwide context.

4. Results of the Survey

The results of the survey (Figure 3) show, that seven of the eleven statements were scored above the average level of six points. Figure 3 shows the means of the answers, and Figure 4 shows the variances of the results. According to this plot (Figure 4) we identified that there are exactly three statements with very low variance. Therefore we believed it worthwhile to discuss these three statements thoroughly during the interview phase.

We determined that "justification and refutation of perception", "reaching own goals", "learning from each other", and "building up trust" are the major motivations for information exchange--the latter previously discovered by Huener (1998). However, this result does not justify the business transaction theory. We needed, therefore, to extract the meaning of the statements by conducting interviews, hoping to identify interpretations supporting the business transaction theory.

During our interviews we found that the statement "reaching own goals" needed deeper discussion, especially as it relates to the business transaction theory.

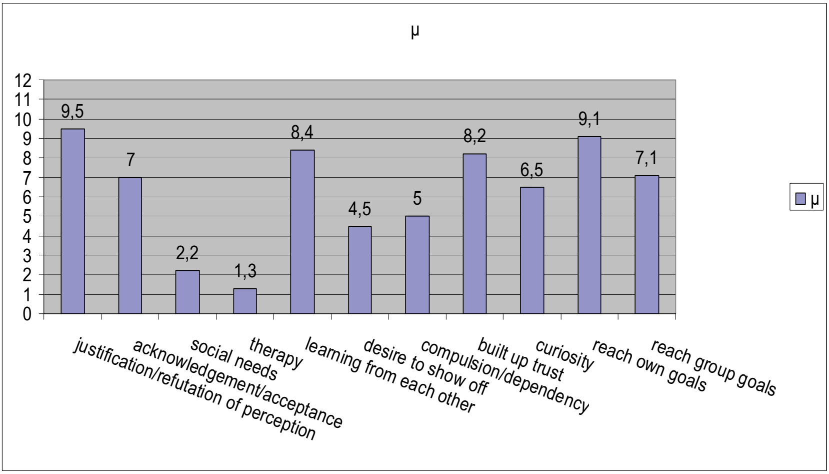


Figure 3. The mean of the answers.

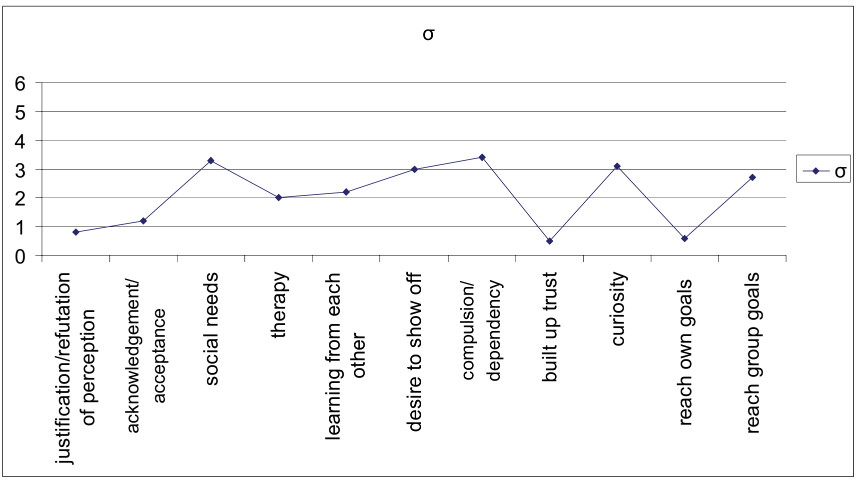


Figure 4. The variances of the answers.

There was common agreement among focus group participants that people's goals are linked with value. Thus, each goal has some personal value. Since most of the individual goals can only be achieved through information and knowledge sharing, it seems to be the summary[7] of all

[7] Or some mathematical function like integral or weighted summary.

types of information and their value e, which constitutes the individual value of the goals. The information trading process--with its asymmetric and individual evaluation of information--constitutes the cornerstone of this value chain.

The importance given by respondents to the statement “reaching own goals”, as well as the very low variance in scoring this statement, and most importantly the interpretation of such as described by focus group members gives us confidence that the business transaction theory is likely correct. Although we believe that differing cultures would probably favor other factors[8], we are convinced that the business transaction theory is valid and independent of cultural differences. To our knowledge, setting up goals is a cultural, independent human property. If we compare investigations about innovative online communities--as was performed with Niketalk [Füller et al., 2006] first of all seems to reflect that knowledge is exchanged for free. However, a deeper analysis shows that the main motivations to share knowledge in this case are the desire to help, striving for recognition from others, and deriving enjoyment from interaction. These factors in turn create satisfaction, which is of personal value to individuals. In this case the sole purpose of goal setting is fun.

The second part of our survey was devoted to the performance of employees. Some researchers see a connection between performance and knowledge sharing. Alternatively, Sveiby (2002) has shown that there is absolutely no empirical evidence that more knowledge sharing is creating more value than competition.

During our interviews we identified motivators as well as moral hazards, which hamper knowledge exchange within a society. We performed a cluster analysis from our interviews and mapped the answers to the Alderfer’s pyramid (Figure 5).

Like Hartmann (1964), we present a summary of existence needs, biosocial needs, cognitive needs, and psychosocial needs. For statistical relevance, we only present those extracted opinions, which are supported by more than 70% of focus group members. Compared to the online

[8] In Japan e.g. the factor “reach group goals” is probably more important than “reach own goals”.

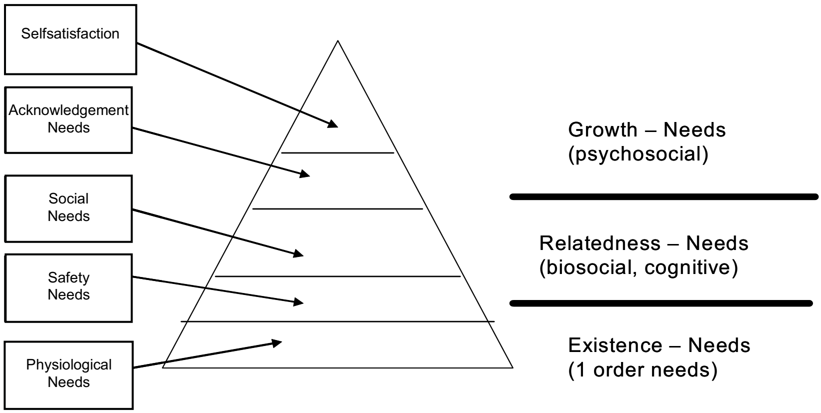


Figure 5. Alderfers pyramid.

survey, the random sample is rather small, and it represents a limited domain[9]. The results strongly depend on cultural roots and personal societal status.

The following common existence needs and motivators were identified:

- Participation affecting company results (success) is important:
  Success should be measured on individual and collective performance. Part of a salary should be dependent on the personal ability to cooperate. People felt that the European educational system is not successful in teaching cooperative working techniques.
- Salary variance between CEOs and workers is perceived as too great:
  Participants cite discomfort about the salary difference between CEOs and blue color workers[10]. Significant differences split the society and subsequently will not promote knowledge sharing between classes. This will in turn hamper economic growth.
- Fringe benefits are important.

[9] 40 employees.
[10] This is true also for white and gold colour workers.

- Both management and the "working class" need company binding programs. The longevity of employment directly affects the involvement in corporate knowledge processes.

We conclude that there is a substantial moral hazard for knowledge sharing. It is the salary and, thereof derived, as well as the power distance--as explained by Hofstede (2005)--which hampers knowledge exchange between humans. Moreover, there is a need for improved metrics to evaluate collective performance indicators.

The following common biosocial needs and motivators were identified:

- Dependable information is important:
  Honest, correct, and timely information is needed. Adherence to this principle prevents companies from being the object of rumors while supporting working morale in teams.
- Promotion of wellness is important:
  Wellness seems to be one of the major challenges for humans. Support for a variety of sport-related activities and healthy meals in addition to the corresponding education of such puts a company in pole position.
- Integration of elderly people is important:
  In contrast to e.g. China, Europe does not appreciate the accumulated know-how of elderly people. This is most probably due to the existing reward system in Europe, in which older people earn more money than younger, and very soon their pay, when value-compared, is too expensive.

The latter point needs an especially intensive consideration so that knowledge flow between generations can work properly; if not, reinventing the wheel is unavoidable.

The following common cognitive needs and motivators were identified:

- More knowledge sharing and incentives are desired:
  Too much competition does not promote knowledge sharing. Respondents felt that companies don't exploit all the available theoretical incentive methods. Many felt that knowledge sharing is not always believed to be positive.
- Better empowerment is beneficial:
  Empowerment was seen as the cornerstone for innovation. In this respect, people felt that learning is important. However, management in Europe has yet to develop the right attitude toward error acceptance. Making errors is still punished in some industries.
- Working morale and a productive atmosphere must be maintained:
  Respondents suggested that gaining e.g. 1 Euro through innovation in the production cycle could easily turn to loss due to inequitable foreign exchange rates[11]. Innovation does not pay off in such a scenario. Moreover, high taxes on labor and low taxes on assets erode working moral.

These cognitive needs, then, reflect the typical "winner takes all" principle of the European society. Likewise, they reflect the problems of high labor costs. Working morale is hampered and knowledge-sharing efforts are diminished by macro economic factors and political hazards.

Due to statistical relevance[12], it was not possible to find one single common motivator or morale hazard for the psychosocial needs.

5. Conclusion

Our comprehensive online survey, combined with personal interviews, supports the business transaction theory. According to this theory, knowledge sharing is a trading process. We are aware that survey results might vary greatly from culture to culture. Group goals might indeed be scored higher than individual goals in cultures separate from Middle

---

[11] Product export.
[12] The cluster analysis extracts only answers supported by more than 70% of the random sample.

Europe. However, it should be noted that goals are always linked with individual value, even those of online communities engaging in fun activities. Moreover, setting goals is a cultural independent human property. Since most of the goals can only be achieved through information sharing, it is the value of information, which plays a mayor role in the value chain. It is this piece of extracted common agreement, which makes the business transaction theory inviolable.

The knowledge sharing process and working morale is influenced by several motivators and morale hazards, which were detected during the interview phase. However, the presented results derived from the second part of the survey, namely personal interviews, are cultural dependent. These mirror a snapshot of Europe's current society.

Acknowledgments

I owe thanks to my students who helped in conducting the interviews. This research was performed with a grant of the 6th research framework of the European Union.

References

1. Barachini F. (2003). Frontiers for the codification of knowledge, *Journal of Information & Knowledge Management,* 001, 1-5.
2. Barachini F. (2202). *Macht in der Wissens- und Kompetenzgesellschaft, Shaker Verlag GmbH*, ISBN 3-8322-0656-6.
3. Bontis N., Crossan M., Hulland J., (2002). Managing an organizational learning system by aligning stocks and flows, *Journal of Management Studies,* 39(4), 437-469.
4. Edvinsson L., Malone M.,(1997). *Intellectual capital,* New York: Harper Business.
5. Füller J., Jawecki G., Mühlbacher H., (2006). Equipment-related knowledge creation in innovative online basketball communities, *The Future of Knowledge Management,* 161-183.
6. Hartmann N., (1964). Der Aufbau der realen Welt, Grundriss der allgemeinen Kategorienlehre, Berlin.
7. Huener L., von Krogh G.,& Roos J., (1998). Knowledge and concept of trust, in Krogh, Roos, Kleine (ed.) Knowing in firms, *Understanding, Managing and Measuring Knowledge,* London: Sage.
8. Hofstede G., & Hofstede G.J., (2205). *Cultures and organizations software of the mind,* New York: McGraw Hill.
9. Hofstede G., (2001). *Culture's consequences: comparing values, behaviors, institutions, and organizations across nations,* (2nd ed.). London: Sage.

10. Sharpe F.W., Alexander G.J., & Bailey J.V., (1995). *Investments.* Upper Saddle River, NJ: Prentice Hall.
11. Sveiby K.E., (1997). *The new organizational wealth: Managing and measuring knowledge-based assets.* New York: Berret-Koehler.
12. Sveiby K.E., & Simons R., (2002) Collaborative climate and effectiveness of knowledge work—an empirical study*, Journal of Knowledge Management,* 6(5).

This page intentionally left blank

Chapter 2

KNOWLEDGE MANAGEMENT TECHNIQUES FOR KNOW-HOW TRANSFER SYSTEMS DESIGN: THE CASE OF AN OIL COMPANY

Djilali Benmahamed
Telecommunication National Institute - Information Systems Department
9, Rue C. Fourier 91011 Evry - France
+33 160 764 765
djilali.benmahamed@int-evry.fr

Jean-Louis Ermine
Telecommunication National Institute - Information Systems Department
9, Rue C. Fourier 91011 Evry - France
+33 160 764 765
jean-louis.ermine@int-evry.fr

1. Introduction

In today's context (globalization, harsh competition, knowledge based economy, ever growing international mobility, etc.), the knowledge management in companies has become an extremely important issue At stake is the potential damage caused by the loss of a key competency or significant number of personnel in terms of planned departures or in terms of the most experimented staff. This reality has caused concern in an ever-increasing way, thus calling for the necessity to adopt a strategy of knowledge management. In this article, we define the elements of a strategy of know-how transfer for the petroleum Group Sonatrach[1]. This research work is based on practical application in the company and is financed by Sonatrach.

Sonatrach is the Algerian company for research, exploitation, transport by pipeline, transformation, and trading of hydrocarbons and their derivatives. It also operates in other sectors such as electricity

[1] www.sonatrach-dz.com.

production, new and renewable energies, and the desalination of seawater. Its activities represent approximately 30% of the GNP in Algeria. It employs more than 120,000 staff and workers.

The organization policy and operating principles adopted by the Sonatrach Group organize its activity around its core business with a reinforcement of the capacities of the top management in terms of development of the strategies and policies, an effective decentralization, and a simplification of its functioning. Operational activities involve those in the Group and the development of its business capacities both in Algeria and overseas with upstream and downstream activities, transport by pipelines, and commercialization.

Sonatrach top management has launched and supports the knowledge management project as a strategic project. This project is based on a global vision of the company, together with local and effective actions aiming at producing noticeable profits in the short run. It aims at preserving the strategic potential of knowledge that has been acquired throughout the years, but which has remained tacit for its holders.

The mapping studies, which were initiated during the early stages of the project, constitute the backbone for a future observatory of competencies in the Sonatrach Group. Similarly, knowledge management has proven to be a powerful and inevitable tool for the forthcoming Sonatrach Corporate University.

Our research aims first at showing the feasibility of a capitalization strategy and the transfer of know-how in oil industry domains as much as it aims at describing the conditions of success in such environments. The re-use of the strategy as well as of the devices and their deployment for every unit constitutes one of the expectations of the company.

From a scientific point of view, we try to validate a formal approach of know-how transfer based on strategy and on knowledge management and knowledge engineering techniques. The first result is based on an effective action plan which involves innovating concepts of knowledge management such as strategy maps, knowledge mapping, critical analysis, knowledge engineering, community of practice management, collaborative work, etc.

The scientific framework of this study is based essentially on some concepts such as strategy maps (Kaplan & Norton, 2004); knowledge maps and criticality analysis as developed by the Knowledge Management Club[2]; knowledge strategic alignment (Ermine, Boughzala & Tounkara, 2005); knowledge modeling as in the MASK method (Ermine, 2002) and (Ermine, 2003); as well as other knowledge engineering techniques (Graqc[3]); learning groups and communities of practice (Wenger, McDermott & Snyder, 2002); modeling of the elements and structures of the various learning units using the IMS specification (IMSLD, 2003; the computer-assisted human learning environments (Tchounikine, 2002); and E-learning. The synergy resulting from the interconnection of these concepts during the various stages of our research is the formal and innovating framework of the ongoing study.

Our exploratory study is interested in the operational activities within Sonatrach, particularly the upstream activity (research, exploitation, and production of hydrocarbons). It appears to us more valuable to carry out the study of this activity, with respect to the size and complexity of the company. We have thus decided to confine ourselves to targeted knowledge by selecting a structure, which constitutes a nodal point and an inevitable upstream activity, involving strategic and critical know-how. Thus, the PED (Petroleum Engineering & Development) department was selected as the test domain of our study.

This research project, within the PED, consists in designing and testing the KM methodology, based on the concepts mentioned above and directed by the Sonatrach strategy, whose aim is to ensure sharing and transfer of the most critical knowledge. During the initialization phase, we presented our project in order to make the actors at various levels aware of the importance of this research work and to ensure their active participation in this project.

The phase currently underway consists of designing, then implementing devices (e.g. knowledge servers and computer-assisted human learning, e-learning type) by capitalizing knowledge with a knowledge engineering method and carrying out, thereafter, a

[2] http://www.club-gc.asso.fr.

[3] French Group of Research in Knowledge Acquisition, http://www.irit.fr/GRACQ/.

pedagogical observation based on the acquired knowledge models. The groups of learners involved in the learning apparatus presented here, once knowledge is sharing, will evolve into communities of practice (Wenger et al., 2002).

2. The Method

The main principle of the method is to identify the best professional knowledge and practices, to formalize them into models, and to ensure the transfer of know-how. This requires the design of a knowledge repository (knowledge mapping), a strategic analysis, and knowledge alignment, then choosing, designing, and implementing the most suitable modes for diffusion, sharing, and acquisition of this knowledge.

The objectives are to select the processes of collective learning that have a capitalized set of know-how available and which have been recognized as crucial by this analysis. Figure 1 summarizes the main steps of our method, which, in addition to its strategic dimension, puts the knowledge holders (Knowledge Workers) at the centre of any process of thinking or acting.

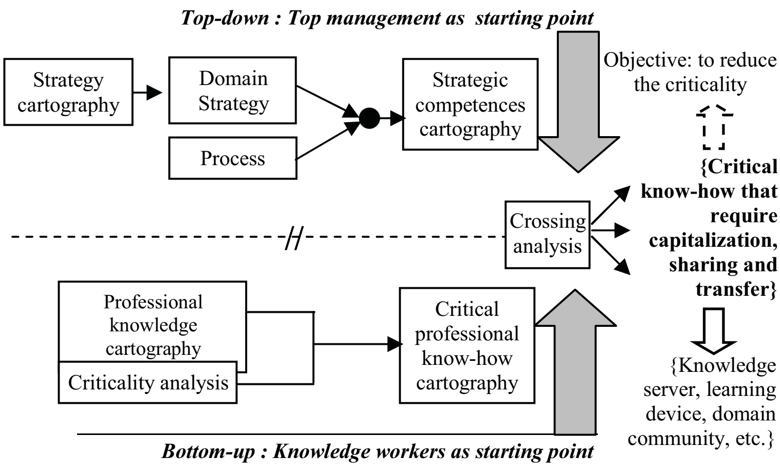


Figure 1. Main steps.

The practical work has been carried out on the basis of four steps: group working sessions involving those concerned with discussion on a particular issue; individual interviews with nearly 20 people concerned in various processes (knowledge actors or experts); interviews with managers to explain the strategy; and finally, readings of reference documents. Eighty percent of the interviewed knowledge actors have an average of 20-years experience in their competences domains with highly qualified professional profiles. The interviews and group working sessions took place at the PED and together with Sonatrach top managers.

During the interviews, some support tools have been used (such as evaluation grids for criticality of the KM club, profile cards, recordings, etc.) as well as interviewing techniques. The fact that the interviewees work for the company has facilitated the contacts and the interviews. During group working sessions, we used facilitation techniques (brainstorming) and video projections to help the group converge their views and ideas for a better knowledge sharing. Once formalized, all results have been validated with the participants.

3. Strategic Analysis

3.1. *Strategy Mapping*

The strategic orientation of a company is the crucial factor for its performance (Atkinson, 1990). In fact, whatever the type of strategy chosen, as may be the case for generic strategies or development strategies, the company may win a competitive advantage resulting in sales increase, profits, or outputs (Jouirou & Kalika, 2004). The strategy mapping of the company was carried with the objective of strategic alignment. This alignment aims at matching the needed strategy for knowledge management with that of the company, specifically in terms of professional knowledge.

After a number of meetings, consultations, interviews, and the reading of reference documents, we have extracted a number of contextual factors that represent parameters for Sonatrach strategy. The current context involves a situation that is increasingly competitive, however, it also offers various opportunities as: the possibility of exploiting the convergence gas–electricity, the possibility of export of gas towards markets where the price level is high, and/or focus towards markets where there is reserves depletion, etc. Therefore, this constitutes both elements of strategy for the Sonatrach Group (aiming at improving the strategy of exploration and the consolidation of its reserves potential) as much as it constitutes assets (competitive attitude, exploitable NG/LNG flexibility, potentially rapid increase in production, etc.).

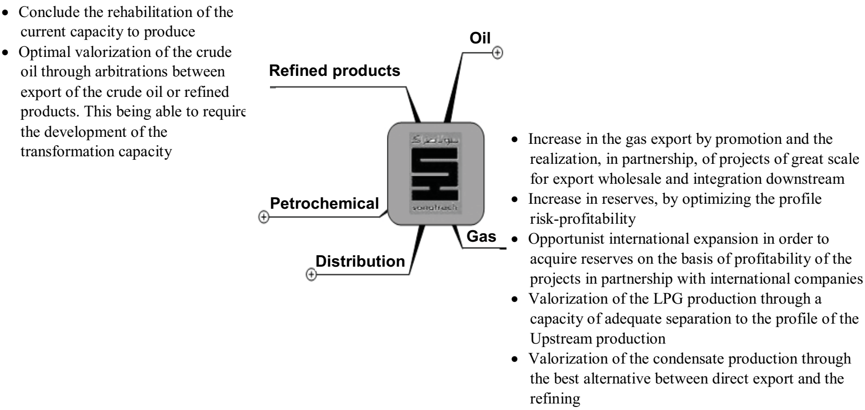


Figure 2. Some elements of the strategy oriented business.

At the end of this first phase, we have obtained a consolidation of the strategic objectives of the Sonatrach Group. This analysis is represented as a Kaplan-Norton's "strategy map" (2004), which may be summarized as follows:

1. Development of the reserves level (discoveries and re-estimation),
2. Realization of production targets of 1,5 MMbbl per day in 2010,
3. Attainment of exports previsions of 85 bcm per year of natural gas by 2010,
4. Improvement of the LPG and Condensate,
5. Optimal exploitation of refining assets, distribution and petro-chemistry,

6. Encouragement in partnership to reduce risks,
7. Development at the international level on all segments of the chain.

3.2. *Impact of Sonatrach Strategy on the PED Department*

3.2.1. *The PED, a Significant Reservoir of Strategic Know-how*

The PED (Petroleum Engineering & Development) is a division that deals with upstream within Sonatrach Company. Involved in nearly the complete chain value E&P (Engineering and Production), the PED is the division of upstream that handles the largest variety of petrotechnical data. One of the major roles of the PED is to gather and store the relevant data generated by other divisions. This position puts the PED at the centre of the interactions with various structures of upstream. Thus, the PED has to fulfil missions such as:

- Basic engineering studies and definition of the options for development in each domain,
- Planning and following operations (drilling and workover) and production,
- Technological survey and implementation of new technologies (shorts radius, horizontal drilling, etc.),
- Design and definition of development plans and reserves exploitation (operated by Sonatrach and in association or combined actions),
- Realization of technical and economic studies according to the company policy to improve existing or discovered reserves,
- Evaluation of the reserves in all oilfields throughout the Algerian territory, thus setting production and injection previsions on the basis of the reserves situation, of the oilfields development level, and installation capacities,
- Estimation of opportunities as to the acquisition and assets development through self-effort and/or joint projects in Algeria and overseas.

The PED department is a nodal and most-important structure for upstream within Sonatrach. It constitutes a significant bank of strategic know-how. This is ample reason for the choice of this department for the purpose of testing the present research project.

3.2.2. *Strategic Competences for PED*

The mapping of Sonatrach strategy (Figure 2) has been declined on the particular case of PED. Using these results, we have tried to highlight the competences, which fit the vision of PED managers. We thus performed several meetings and interviews to carry out the mapping of strategic competences in the PED. Figure 3 illustrates the results.

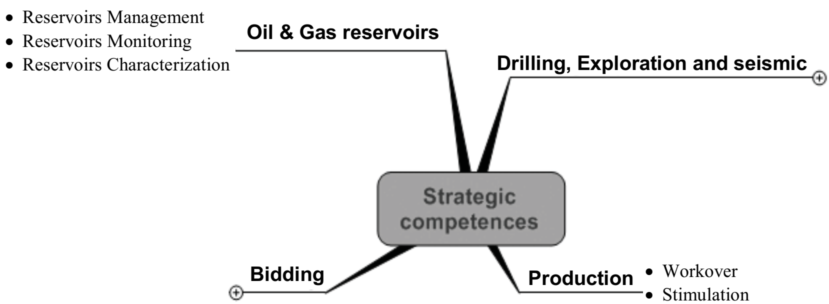


Figure 3. PED strategic competences.

This set of competences is attached to the strategy of the domain, and it identifies "core competences" as listed in Prahalad & Hamel (1990). The strategic analysis has put forward the strategic and necessary competencies that allow attainment of the corporate objectives.

One of the main factors for success in a knowledge management method is the active participation of the knowledge workers, who remain the fundamental resources of any operational device of knowledge processing (Oswaldo & Matta, 2005). The strategic vision is, however, not sufficient to ensure the success of the method. It is also necessary to match this strategic vision with that of the operational knowledge workers. This is, in fact, the objective of the following step, i.e., that of the professional knowledge analysis. The purpose of this analysis is to identify critical know-how, according to the professionals.

4. Professional Knowledge Analysis

4.1. *Knowledge Mapping and Criticality Analysis*

Building a map and carrying out a criticality analysis require a methodological process (Aubertin, 2005). The method for knowledge mapping used here is defined in Ermine et al. (2005).

The criticality of a domain is defined as risks or opportunities estimation in a given sector of the company (Aubertin, 2005). We have chosen a grid of criticality constructed by the French Knowledge Management Club[4]. Figure 4 represents an overview of these criteria.

Once the mapping of the knowledge domains had been set up, we carried out the estimation of these criteria for each knowledge domain.

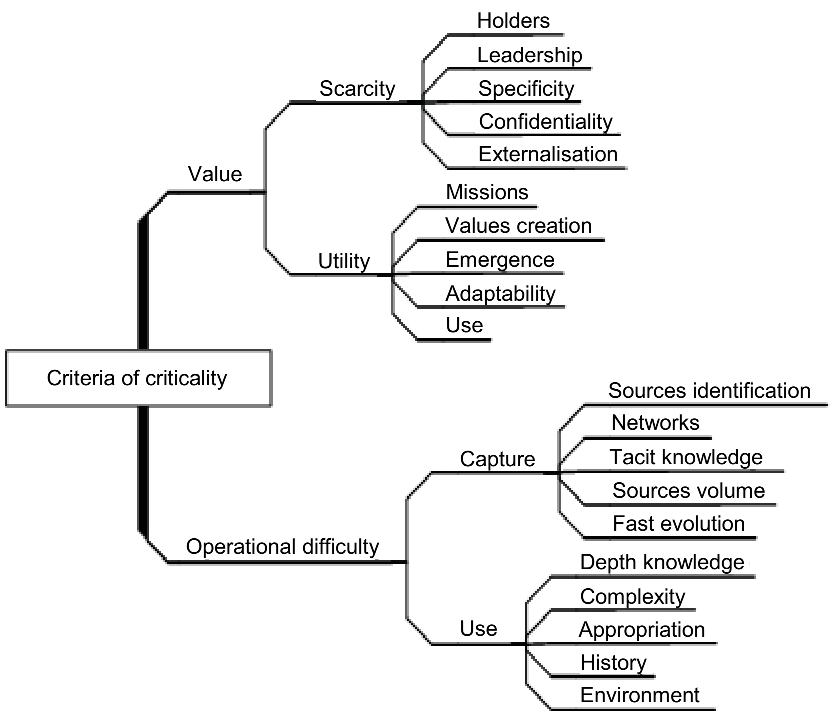


Figure 4. Criteria for the evaluation of criticality.

[4] http://www.club-gc.asso.fr.

4.2. *The Results*

During this phase, our interlocutors were professional experts (knowledge workers) at PED. The adopted principle was to gather the various activities in the knowledge domains, to elaborate them by means of a critical representation, then to supplement and validate in an iterative way the resulting mapping produced by these experts.

A map became elaborated, as the interviews were ongoing. An appropriation of the questioning by the interviewees strengthened progressively. The various map versions were validated. This iterative validation took the form of a co-building in order to guarantee a maximal building together, collaboration, and appropriation by the interviewees. Once this stage had been finalized, we obtained the map given in Figure 5.

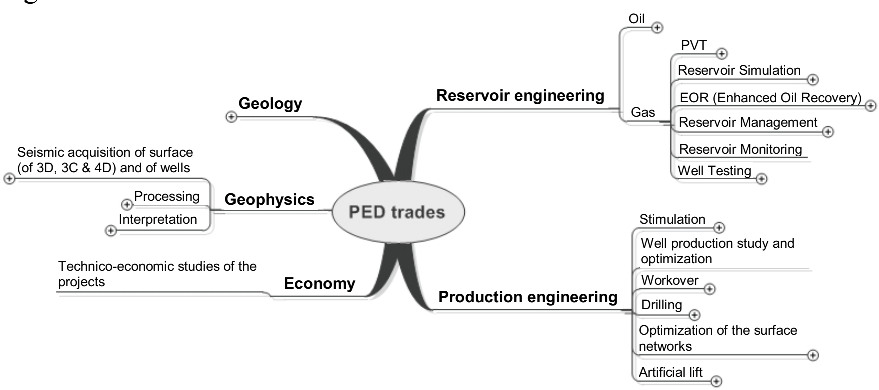


Figure 5. PED knowledge map.

This map represents a description of PED know-how at a supra (Meta) level (Prax, 2005). It provides an addressing system of know-how, which facilitates knowledge access.

On the basis of this map, a criticality study was carried out. The evaluation of the criticality of a domain consists in attributing a grade according to each criterion of the analysis grid for each domain. The more critical the domain, the higher the grade. Each domain was evaluated independently from the others. The restitution of the results relating to each domain is synthesized graphically in a radar diagram.

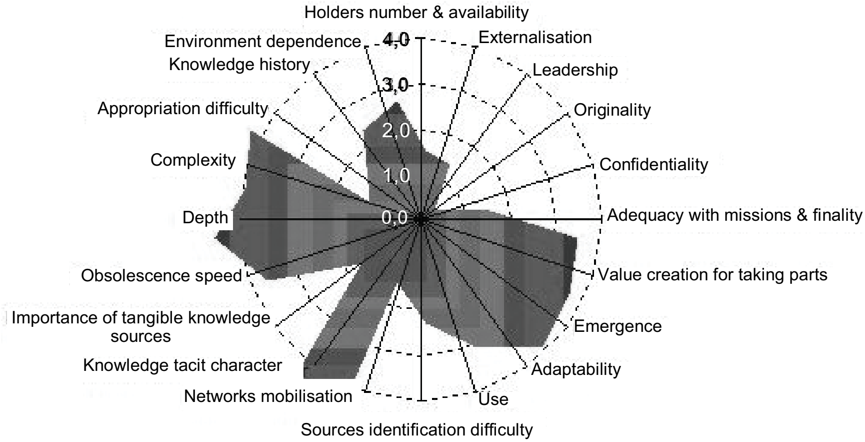


Figure 6. Criticality radar diagram of "reservoir modelling" know-how.

The results of this criticality study allowed us to establish a map where the most critical domains are identified. The critical domains in Figure 6 are highlighted according to their color: red, orange and green.

This visualization was very useful when presented to the managers. It has the advantage of being simple and easy to trace in the sense that any element can be justified by a complete analysis file (interviews verbatim, notational system, synthetic leaflets, radars, etc.).

The work was carried together with the interviewers to guarantee appropriation of method and results. At this stage of the KM process, we have obtained two results:

- A map of strategic competences involving the management and obtained on the basis of the company strategy analysis, and
- A map of critical knowledge, obtained by an analysis involving the knowledgeable professionals.

Without being contradictory, these two mappings represent different points of view. The next step, strategic alignment, aims at synthesizing these points of view.

5. Strategic Alignment

At this stage, there is a need to compare critical knowledge and strategic competences. In other words, we have to match the results displayed in

Figure 3 with those displayed in Figure 5. To achieve this, we suggest a cross analysis on two levels.

A first filter on the main branches of the two maps (Figures 3 and 5) allows us to link knowledge domains to strategic competences. The table in Figure 7 illustrates this first level of our cross analysis.

In the second step, each identified link (highlighted by an X in Figure 7) is identified by decomposing the main branches in the mappings. Below is an example of the analysis of the links between “Engineering Reservoirs X Hydrocarbons reservoirs”, which is

Domains	Geophysics	Geology	Economy	Reservoirs engineering	Production engineering
Hydrocarbons reservoirs		**X**	**X**	**X**	
Production				**X**	**X**
Drilling, exploration & seismic	**X**	**X**			
Bidding			**X**	**X**	

Figure 7. Cross analysis of knowledge domains.

summarized in the table of Figure 8. From a qualitative point of view, the results represent a set of identified knowledge considered at the same time critical by the professionals and responding to the required competences of the strategy.

Let us recall that we are concerned here only with the most critical knowledge. It is thus necessary to weight the analysis with the criticality factor of the knowledge. Thus, in this example, the know-how appearing on the last line will not be considered because it is not critical. Lines 1 to 4 represent know-how, which is both critical and responds to strategic competences. It is, however, only line 2, which represents a critical know-how involved in the maximum of strategic competences. Our analysis, then, makes it possible to select this know-how, “reserves simulation”, as the most critical know-how and the one that responds most to strategic requirements.

Hyd. Res. / Res. Eng.	Reservoir Characterization	Reservoir Monitoring	Reservoir Management
PVT	**X**		
Reservoir Simulation	**X**		**X**
EOR			**X**
Reservoir Management			**X**
Well testing	**X**	**X**	**X**

Figure 8. Link Analysis for “Engineering Reservoirs X Hydrocarbons Reservoirs”.

The whole set of critical know-how that has been selected is reliable and may be verified because it is the traceable result of a global methodology (strategy mapping, knowledge mapping, criticality analysis, strategic alignment, etc.) that involves every kind of the company actors. It allows discriminating knowledge domains within a very important initial knowledge capital (in that case, only 15% has been retained from initially identified knowledge), on the basis of objective criteria that are open to debate.

The result of this strategic alignment of critical know-how constitutes the basis for the next step (currently in progress). This critical knowledge is characterized by professional skills and logics, and within this division they are using a great number of knowledge sources. The criticality analysis has revealed the very tacit nature of a large number of know-how. In order to test the next step of our methodology, we have decided to focus on this tacit know-how, though this does not represent in itself the only problem that has been identified. A domain has been selected; experts have been identified; and capitalization has been initiated.

6. Capitalization of Tacit Know-how

Knowledge codification is a method, which is often used for processing knowledge, acquiring it, designing learning environments, and setting up the processes by which knowledge based organization becomes a learning organization. This can be achieved by knowledge modeling methods, which identify and structure knowledge in a diagrammatic representation in order to make it visible, easy to handle, comprehensible, and transferable (Paquette, 2002).

The capitalization process includes several stages centered on the strategic knowledge notion, which involves the location or identification of the sources of knowledge, their formalization, organization, storage, distribution, and maintenance. For each one of these stages, there are methods, tools, and computerized-products, besides or in place of less industrial and more traditional techniques.

Knowledge capitalization is not a goal in itself. It is a permanent issue that is omnipresent in each employee activities. This process is well illustrated on the diagram proposed by Grundstein and Zacklad (2001). As a method of capitalization, we have chosen the MASK method (Ermine, 2003) and (Ermine, 2002). This method for knowledge management is currently in its third generation. For nearly 15 years, it has been continuously refined and improved through numerous projects that have been carried out in companies of various sectors and different sizes.

The result obtained initially in a MASK project is a set of models that formalize knowledge and which were elaborated through interviews with the knowledge holders. These are sufficient enough if one wishes to proceed to an immediate data-processing application (a computerized decision-making system, a database, etc.).

The MASK models, supplemented by full information and documents, cards, files, etc. which are relevant, constitute the “Knowledge Book” of the given domain. The concept of Knowledge Book is a continually growing notion within the knowledge management framework, and it has come to be a very important concept. It capitalizes and diffuses a whole set of knowledge in a given area; it represents knowledge structure; and it indexes the documents concerning the activity (descriptive cards, memos, publications, hyperlinks, etc.) together with multi-media contents (video, images, sounds, etc.). It also provides a strong basis for any operational project on knowledge processing (Ermine, 2003).

The first possible use of a Knowledge Book is its access by the knowledge workers in a given intranet “knowledge space”, which is dedicated and integrated into the information system of the company. It is what we also call a “knowledge server” or “knowledge portal”.

In addition, the MASK models represent a strong interest for teaching scenarios in order to describe the steps for the sharing and the acquisition of the modeled knowledge. This allows determining the contents of the learning devices (Benmahamed, Ermine & Tchounikine, 2005).

Knowledge servers and computer-assisted human learning environments--e-learning type--represent in fact the technical devices for the know-how transfer we are suggesting in our methodology. The designing and application of these transfer devices and the definition of their contents, are to be treated in the present research.

In the objective of professional learning, we are interested in the Knowledge Book representing know-how and best practices, as provider of the essential content for training devices. Hence, we distinguish ourselves from the classical methodology on traditional training (Brown & Campione, 1994), (Fleer, 1992), (Soller, 1992) and (Strommen & Loncoln, 1992), which is centred more on the learner than on expert knowledge. In what follows, we shall indicate some hints for future research and developments in the KM process in the company.

7. Designing Transfer Devices

We are currently focussing on the most critical domains and aiming at establishing a coherent set of methods and tools in order to ensure the transfer. Rather than considering disparate devices, which would look more like an array of tools than a concerted methodology (Ermine, 2003), our target is to develop a complete and coherent working plan.

Several solutions are being investigated: sharing devices under the form of collaborative spaces, online discussion, and knowledge and transfer servers, especially through training of this know-how by using E-learning technologies according to clearly defined norms and standards, that can later on be deployed (scalability).

In that purpose, we will exploit and re-use our already completed work (Benmahamed et al., 2005) for the definition of learning contents, starting from the knowledge included in MASK models (Ermine, 2003) and (Ermine, 2002). This work articulates knowledge engineering techniques that the MASK models encapsulate, and those of teaching management that can be exploitable through teaching scenarios. Once the targeted know-how has been made explicit and capitalized, we suggest ways for moving from these MASK models to teaching scenarios whose elements are described according to the description language IMS–Learning Design. This use of normalization reinforces the setting of scenarios and responds to a vision of standardization and re-investment of the contents of the transferring by learning devices.

This is a new approach, which appears to be effective and gives the possibility to develop realistic learning activities within the professional context. Indeed, knowledge to be learned corresponds to practices deriving directly from the concerned professional activities. Figure 9 describes the principle for the elaboration of learning contents, starting from the knowledge encapsulated in the MASK models and other complementary sources. This allows pointing the contribution of the Knowledge Book to the definition of learning devices, and particularly, the central role of the models, supplemented by other elements.

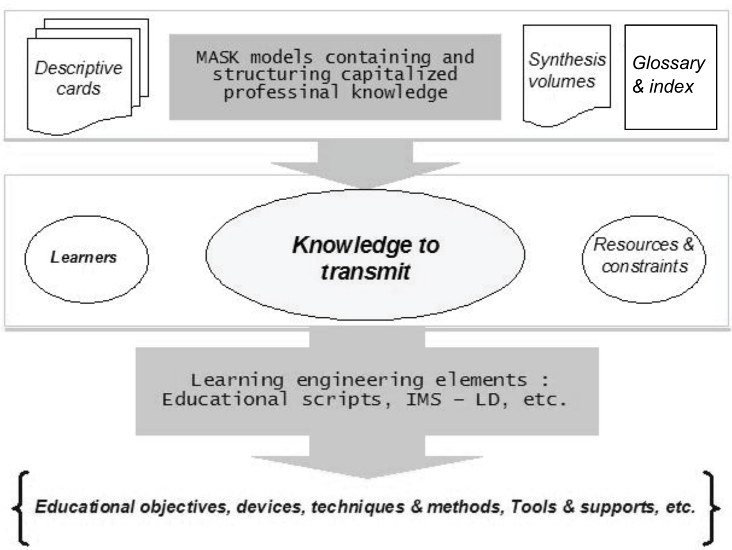


Figure 9. Principle of learning contents elaboration.

The teaching design must take into account not only the contents and the tasks to be achieved, but also the "learning relations", which will bring up to date (update) learning, as advocated in Mayes (2003). Examined will be the importance of online supervision, which is based on the suggestions in the works by Fowler and Mayes (2000) on the importance of the dialogical communication in the building of knowledge.

Synergies between these tools and the process of capitalization and sharing constitute the basis of our research. This study will also have to consider the emergence of a learning community that activates around the designed devices and its progressive evolution towards a community of practice as defined in (Wenger et al., 2002).

8. Conclusion

Both from a scientific and an industrial point of view, the objective of the project is to design and test a complete knowledge management method for the transfer of professional know-how. On the basis of what has been achieved so far, the method is articulated in five steps:

- Stage 1: Strategic Analysis,
- Stage 2: Critical knowledge Analysis,
- Stage 3: Strategic Alignment,
- Stage 4: Capitalization of Tacit Know-how,
- Stage 5: Design of Knowledge Servers and Computer Assisted Human Learning Environments (e-learning like).

Each stage is based upon established theoretical bases. (See Introduction.) Those bases are classical in management science and knowledge engineering and have been improved and adapted to the issues raised by the research project.

The project is currently in the mid-achievement. Stage 4 is now classical. For a substantial part of Stage 5, the theoretical framework has already been conducted in (Benmahamed et al., 2005). Research in progress will thus lead us to develop knowledge models and learning models, as well as to design knowledge servers and Computer Assisted Human Learning Environments as an e-leaning platform.

We have particularly stressed the fact that the project proceeds in the most possible participative way, as this is a key factor of success. The professionals are indeed the main producers and consumers of know-how. Their participation as co-designers is a must for the appropriation of the methodology. This active participation approach for the pilot project, which is currently undertaken, involving both top managers and operational knowledge workers, has largely reduced resistance to sharing and elicitation of knowledge.

In our mapping approach, we aimed at the location of knowledge (review of knowledge areas, identification of expertise holders, etc.), evaluation and analysis of criticality of the knowledge capital (audit), and the visualization of critical knowledge and its alignment to Sonatrach strategy. This made it possible to identify know-how on which this strategy had some impact, and therefore to identify the knowledge domains which have to be sustained and/or developed by actions of transfer via capitalization and learning in particular. The next step is to demonstrate how the technology can support this transfer.

References

1. Atkinson, R.A. (1990). Strategic planning: The motivations for strategic planning. *Journal of Information Systems Management,* (Fall), 53-56.
2. Aubertin G. (2005). Cartographier les connaissances critiques : une démarche stratégique pour l'entreprise (Ed. Lavoisier). *Management des connaissances en entreprise.*
3. Aubertin G. (2005). Knowledge mapping: A strategic entry point to knowledge management. *Trends in Applied Knowledge Management*, Ed. Hermes.
4. Benmahamed D., Ermine J-L & Tchounikine P. (2005). From Mask KM methodology to learning activities described with IMS-LD. *Professional KM international conference, Workshop on Learner-Oriented Knowledge Management & KM-Oriented E-Learning* (LOKMOL), Kaiserslauten (Allemagne). (Ed. by Springer Verlag) – ISBN 3-00-016020-5.
5. Brown, A. L. & Campione, J. C. (1994) *Classroom lessons: Integrating cognitive theory and classroom practice*, Cambridge, MA: MIT Press/Bradford Books.
6. Ermine J-L. (2003). *La gestion des connaissances,* (Ed. Hermès, Paris).
7. Ermine J-L. (2002). *Initiation à la méthode MASK,* CD-ROM de l'Université de Technologie de Troyes.
8. Ermine J-L, Boughzala I. et Tounkara T. (2005). Using cartography to sustain inter-generation knowledge transfer: The M3C methodology. In *2nd International Conference on Intellectual Management, Knowledge Management and Organisational Learning.* American University in Dubai, U.A.E.

9. Fleer M. (1992). Identifying teacher-child interaction which scaffolds scientific thinking in young children. *Science Education, 76(4),* 373-397.
10. Fowler C.J.H & Mayes, J.T. (2000). Learning relationships: from theory to design. *The Changing Face of Learning Technology.* Cardiff: University of Wales Press.
11. Grundstein M. & Zacklad M. (2001). *Ingénierie et capitalisation des connaissance.* (Ed. Hermes Sciences publication).
12. IMS Learning Design Version 1 Final Specification approved by the IMS Technical Board in February 2003, http://www.imsglobal.org/learningdesign/index.cfm.
13. Jouirou N. & Kalika M. (2004). L'alignement stratégique déterminant de la performance (étude empirique sur les PME) In *9ème colloque de l'Association Informatique et Management*: AIM2004.
14. Kaplan R. S. & Norton D. P. (2004). *Strategy maps: Converting intangible assets into tangible outcomes.* Boston: Harvard Business School Press.
15. Kuipers, B. (1983). *The cognitive map: Could it have been any other way? Spatial orientation: theory, research, and application*. (Eds. Herbert L. Pick, Jr. & Linda P. Acredolo). New York: Plenum, 345-359.
16. Mayes T. (2003). *Learning technology and learning relationships.* Unpublished on line presentation, Clasgow Caledonian University, UK.
17. Paquette G. (2002). *L'ingénierie pédagogique pour construire l'apprentissage en réseau.* Presses de l'université de Québec.
18. Prahalad, C.K. & Hamel, G. (1990). The core competence of the corporation, *Harvard Business Review, 68(3),* 79.
19. Prax J-Y. (2005). *Le manuel du Knowledge Management, une approche de 2ème génération.* (Ed. Dunod, Paris).
20. Oswaldo Castillo N. & Matta N. (2005). A knowledge acquisition system for the french textile and apparel institute. *Proceedings of the Knowledge-Based Intelligent Information & Engineering Systems,* Melbourne, Australia, 960-966.
21. Soller A.L. (1992). Supporting social interaction in an intelligent collaborative learning system. *International Journal of Artificial Intelligence in Education, 12(1),* 40-62.
22. Strommen E.F. & Loncoln B. (1992). Constructivism, technology, and the future of classroom learning. Unpublished online *Institute for learning technologies*, Columbia University.
23. Tchounikine P. (2002). Quelques éléments sur la conception et l'ingénierie des EIAH, In *2ème assises nationales su GDR I3*, Nancy, France, 233-246.
24. Wenger E., McDermott R. & Snyder W-M. (2002). *Cultivating communities of practice.* Boston: Harvard Business School Press.

This page intentionally left blank

Chapter 3

PURSUING THE HOLY GRAIL OF KNOWLEDGE-BASED ECONOMY ASSESSMENT: THE CASE OF SINGAPORE

Alton Y.K Chua, Abdus Sattar Chaudhry, Fong Pin Fen
Division of Information Studies, School of Communication and Information
Nanyang Technological University, Singapore
65-6790-5810
altonchua@ntu.edu.sg

While the knowledge-based economy (KBE) and its effects are often recognized and discussed, there is a lack of consensus within the international community about how an economy's progress in the knowledge arena is definitively measured. This paper focuses on the context of Singapore and seeks to develop a set of indicators that could be used to track Singapore's economic development as a KBE. To achieve this objective, major KBE assessment models were reviewed and conflated into an interim set of KBE indicators for Singapore. Thereafter, the appropriateness of the indicators was subject to experts' opinions solicited from the Singapore Economic Development (EDB). The findings yielded specific recommendations to help build capabilities in the areas of knowledge creation, knowledge acquisition, knowledge dissemination, and knowledge application for the economy of Singapore.

1. Introduction

For many years, international bodies, such as the Organisation for Economic Co-operation and Development (OECD), the Asia-Pacific Economic Cooperation (APEC), and the Australian Bureau of Statistics (ABS), have been attracted to the notion of a knowledge-based economy (KBE). This interest is largely fueled by the recognition that knowledge is the key driver of productivity and economic growth. OECD defines a KBE as an economy, which is "directly based on the production and use of knowledge and information" (Organisation for Economic Co-operation and Development, 1996). Along a similar vein, the APEC economic committee refers to a KBE as "an economy in which the production, distribution, and use of knowledge is the main driver of growth, wealth creation, and employment across all industries" (Asia-Pacific Economic Cooperation, 2000). These conceptions of KBE suggest that the basis for economies to grow successfully lies in the creation, acquisition, dissemination, and effective use of knowledge. Requirements to build, maintain, and grow a KBE include an economic government which offers incentives for the use of knowledge, the presence of an educated and skilled workforce, a dynamic information and communication infrastructure, as well as a system of innovation comprised of knowledge-intensive organizations, such as research centers, universities, and think-tanks (Sigurdson, 2000).

While the KBE and its effects are often recognized and discussed, there is a lack of consensus within the international community on how an economy's progress in the knowledge arena is definitively measured. This difficulty stems not only from the fuzzy nature of knowledge but also the fact that different economies are too nuanced among themselves for any generic set of indicators to be universally applicable. Furthermore, as social, economic, and political situations are constantly evolving, indicators used in the past may not always remain relevant. Thus, the need for a set of appropriate indicators to accurately reflect the economy's progress is a perennial one. This is also the 'Holy Grail' for which economies aspiring to remain or become a KBE must incessantly pursue.

This paper focuses on the context of Singapore and seeks to develop a set of indicators that could be used to track Singapore's journey as a KBE. To achieve this objective, major KBE assessment models were reviewed and conflated into an interim set of KBE indicators for Singapore. Thereafter, the appropriateness of the indicators was subjected to experts' opinions solicited from the Singapore Economic Development (EDB), one of the main agencies responsible for spearheading the economic development of Singapore.

This paper contributes to the understanding of knowledge measurement in two ways. One, the scope of most extant knowledge management (KM) studies related to measurement is confined at the organizational level. This paper sheds light on measurement issues that are not often discussed in the literature by considering the economy as the unit of analysis. Two, even though the findings are intended to serve Singaporean-specific needs, they hold relevance for economies of comparable size and stage of development. Furthermore, scholars may use the conceptual model as the basis to develop other KBE assessment frameworks for economies of their interest.

The remainder of the paper is organized as follows. The next section reviews the major KBE assessment models hitherto developed and culminates in a set of interim KBE indicators for Singapore. Thereafter, the methodology section details the data collection method and participants in this study. Following that, the findings section discusses the indicators, which are inappropriate for Singapore, as well as proposes new indicators deemed relevant. The concluding section presents a set of KBE indicators for Singapore and highlights the implications of this study.

2. Literature Review

Attempts to measure knowledge have always been difficult. For one, knowledge possesses characteristics including subjectivity, transferability, embeddedness, self-reinforcement, spontaneity, and perishability--all of which are markedly different from those commonly associated with tangible assets such as land and equipment (Kluge, et al., 2001). Nonetheless, knowledge measurement remains an important priority in knowledge-intensive environments, because measurement serves as the basis on which knowledge processes can be controlled, evaluated, and improved (Ahmed, et al., 1999). Furthermore, measurement accentuates critical knowledge assets and allows increment in the value of knowledge assets to be calibrated. In the long run, measurement helps foster a performance-oriented culture (Kannan & Aulbur, 2004).

A slew of knowledge measurement models such as Skandia Navigator (Edvinsson & Malone, 1997), Balanced Scorecard (Kaplan and Norton, 1992), the American Productivity and Quality Center's Knowledge Management Assessment Tool (Hiebeler, 1996), and Knowledge Management Maturity Model (Kochikar, 2000) have emerged in the knowledge management literature. However, these models are more suited for organizations rather than economies. Thus, an important agenda of this paper is to illuminate knowledge measurement models that are specifically designed to be used at the national level. Among these include the Organisation for Economic Co-operation and Development (OECD) Scorecard, Asia-Pacific Economic Corporation (APEC) framework, Australian Bureau of Statistics (ABS) framework, World Bank Development Report, New Economy Index, and the Economic Survey of Singapore.

To measure the extent to which an economy is knowledge-based, the OECD Scorecard specifies indicators related to five areas, namely investment in intangibles, investment in ICT, investment in Science and Technology, internationalization of technology, and trends in international trade and foreign investment (Organisation for Economic Co-operation and Development, 2001). The indicators appear broad-based, as they were developed to accommodate the needs of all OECD member countries. Furthermore, they show a bias towards technology and unduly amplify the importance of high-technology industries in the economy. By definition, a KBE comprises any industries whose main factor for production is knowledge. Hence, traditional industries such as transport, agriculture, construction, and mining would need to be reckoned.

The APEC framework was intended to promote the creation, dissemination, and effective use of knowledge among APEC economies (Asia-Pacific Economic Cooperation, 2000). It identifies four dimensions for KBE, namely business environment, ICT infrastructure, human resource development, and innovation system dimensions. In comparison to the OECD model, the APEC framework presents a more holistic view of the KBE. Nonetheless, its emphasis on research and development (R&D) as the only form of innovation in a KBE results in a disproportionately high number of R&D indicators.

The ABS framework was developed to measure knowledge in the Australian economy and society (Australian Bureau of Statistics, 2001). While the framework was drawn from the earlier OECD and APEC frameworks, it recognizes the reciprocal influences between the economy and the society. Hence the ABS framework comprises five dimensions, namely innovation and entrepreneurship, human and social capital, the role of ICT, the economy, society and environment, and finally, economic and social impacts. Although some of the indicators are similar to those in the OECD and APEC models, the majority measure the effects of knowledge on society rather than the knowledge of the economy per se.

The World Bank Development Report identifies two types of knowledge critical for economies seeking to become KBE (World Bank, 1999). One is the knowledge about technology, and the other is the knowledge about attributes, such as the quality of a product, diligence of a worker, or credibility of a firm. The report also proposes national policies to increase both types of knowledge through knowledge creation, knowledge acquisition, knowledge absorption, and knowledge communication.

The New Economy Index was developed to benchmark the economic transformation in the United States (Atkinson and Coduri, 1999). It comprises five characteristics of the KBE, namely quality of knowledge jobs, globalization, economic dynamism and competition, transformation into a digital economy, and technological innovation capacity. Given the specific era and purpose for which it was intended, some of the indicators are no longer relevant to the United States today. These indicators include "internet and computers use by farmers" and "number of .com domain names registered".

The Economic Survey of Singapore was developed from the World Bank Development Report to assess the stage of Singapore's economic development (Toh, et al., 2002). It identifies four main knowledge activities, namely knowledge creation, acquisition, dissemination, and application. Such a classification scheme is not only knowledge-centric but covers indicators suitable for Singapore. However, there are a few indicators, which appear too narrowly focused, such as one that measures the number of patents registered only in the United States.

While all the above models and frameworks seek to capture the essence of the KBE, each is purpose-designed and targets a specific economy or group of related economies. Nonetheless, collectively, they form a useful resource from which a context-sensitive KBE assessment framework can be built. On the basis of the above frameworks, an interim set of KBE indicators for Singapore shown in Table 1 is proposed. The development of these indicators was guided by two considerations. One is relevance and timeliness. The indicators must aptly measure the current and near-future economic activities in Singapore. Two is parsimony. The list of the indicators cannot be too lengthy or complicated to be understood.

Table 1. An interim set of KBE indicators for Singapore.

Knowledge Activities	Indicators
Knowledge Creation	• % of GDP spent on R&D • Researchers per capita • Number of patents registered
Knowledge Acquisition	• International Mobility of workers • Number of head and regional offices in Singapore • Foreign Direct Investment Flows
Knowledge Dissemination	• ICT spending as a percentage of GDP • Internet access cost as a % of per capita GDP • % of workforce with at least secondary school education
Knowledge Application	• % of workforce with university education • Presence of venture capital funds • World Competitiveness Yearbook rating of entrepreneurship

3. Methodology

Experts' opinions on the appropriateness of the interim set of KBE indicators for Singapore were solicited from the Singapore Economic Development Board (EDB). Two focus group meetings, each comprising six participants, were conducted. The participants were drawn from the Cluster Development and Corporate Divisions of EDB and chosen on the basis of their familiarity with and involvement in promoting economic activities in strategic industries--including electronics, chemicals, information communications and media, logistics, transportation and services--targeted by the government. These participants are also keenly aware of global trends, emerging industries, and the major initiatives Singapore is undertaking to boost and sustain the country's economic growth.

The focus group meetings at the EDB premises lasted an hour and a half. At the start of each meeting, the purpose of the study and definitions of terms to be used were explained. The participants were also informed that the entire proceedings would be audio taped, but they were assured of anonymity in the report arising from this study. Following that, the interim set of KBE indicators was presented. Comments were sought from every participant, as an individual indicator was presented in succession. Before each knowledge activity--which comprises a triad of indicators--saw closure, a quick summary was made to ensure that the views of the participants were accurately understood and captured. At appropriate juncture, notes were taken to record agreements as well as points of contention among the participants.

4. Findings

4.1. *Knowledge Creation Indicators*

The participants unanimously agreed that all three knowledge creation indicators were valid but restricted in scope. The indicators seemed to favor high-technology sectors and hence unable to capture knowledge created in the non-technical sectors. In the tourism industry, for example, money invested in conducting market surveys to spawn innovative services is usually not recognized as a form of R&D expenditure. Furthermore, other than researchers, many occupational roles such as engineers and management staff are responsible to create new processes, systems, designs, and business models. Table 2 summarizes the participants' opinions on the knowledge creation indicators.

As one of the long-term measures to enhancing Singapore's capabilities for knowledge creation, the participants concurred with the Ministry of Education's move to foster innovation in students through "Thinking Schools, Learning Nation" vision (Goh, 1997). The participants also proposed pro-business initiatives such as incubation schemes to help promising local start-ups raise funds.

Table 2. Discussion on the knowledge creation indicators.

Original Indicators	Comments	Proposed Indicators	Rationale
% of GDP spent on R&D	Valid but restricted in scope	% of GDP contributed by different industries	Measures the number of innovative products and services developed in Singapore
Number of researchers per capita	Valid but restricted in scope	Number of knowledge managers	Measures the number of workers managing the aftermath of knowledge creation
Number of patents registered	Valid but restricted in scope	Number of Singapore brands registered	An indication of the knowledge content in the products created in Singapore

4.2. *Knowledge Acquisition Indicators*

The participants agreed that the knowledge acquisition indicators were generally valid but suggested some refinements to increase the indicators' correlation to knowledge acquisition outcomes. For example, while the number of head and regional offices in Singapore provides an indication of how much firm-specific knowledge brought in by the Multinational Corporations (MNCs) could be acquired by local companies, such an indicator could be extended to include the number of mergers and acquisitions activities in the economy. Also, instead of measuring the foreign direct investment flows, participants suggested considering the amount of money used to purchase foreign patents, franchises, licenses, and rights for use in Singapore companies for the purpose of improving processes or products. Three additional indicators to assess knowledge acquisition activities in Singapore were also offered, namely "the total monetary value of intellectual property owned by Singapore firms", "companies' total investment in training", and "the number of hours spent on training per worker per year". Table 3 summarizes the participants' opinions on the knowledge acquisition indicators.

Table 3. Discussion on the knowledge acquisition indicators.

Original Indicators	Comments	Proposed Indicators	Rationale
International mobility of workers	Valid	-	-
Number of Head and Regional Offices in Singapore	Valid but restricted in scope	Number of mergers and acquisitions activities	Mergers & acquisitions allow local companies to acquire knowledge, expertise and resources to which it would otherwise have no access
Foreign Direct Investment Flows	Valid but restricted in scope	Amount of money used to purchase foreign patents, franchises, licenses and rights for use in Singapore	Measure the amount of technology or know-how that is acquired from outside Singapore
		Total monetary value of Intellectual property owned by Singapore firms	The total monetary value spent on purchasing foreign patents, franchises and licenses gives an indication of the knowledge acquired by the economy
		Companies' total investment in training	Training is a process of acquiring new knowledge or know-how
		Number of hours spent on training per worker per year	Increase in number of hours spend on training reflect the companies' emphasis on knowledge acquisition

Suggestions to improve knowledge acquisition in the economy include national-level efforts to build strong support networks such as industry groups and trade associations to promote lifelong learning. As workers' skills and knowledge are being upgraded through continuous training and development, Singapore would be poised to attract more foreign enterprises from knowledge-intensive industries.

4.3. *Knowledge Dissemination Indicators*

The participants pointed out that the indicator "ICT spending as a % of GDP" did not clearly reflect the economic situation in Singapore. They argued that the bulk of ICT spending could be derived from ICT infrastructure development initiated by the government or government-linked firms such as Info-Comm Development Authority of Singapore (IDA) and Singapore Telecommunications (SingTel). Once the infrastructure has matured, development costs borne by the government would reduce while maintenance costs to be incurred by businesses would increase. Thus, both government and business spending should be monitored separately. Two indicators, namely "government spending on ICT" and "business spending on ICT" were suggested. The participants also questioned the validity of the indicator "internet access cost as a % of per capita GDP" as the affordability of the Internet is not a critical concern for most families in Singapore. The "% of internet penetration in homes" was instead proposed as an indicator for knowledge dissemination. The participants were unconvinced that the percentage of the workforce having at least a secondary educational qualification level correlates well with knowledge dissemination. Rather, several indicators were proposed, namely "major newspapers circulation figures", "the number of seminars, conventions and conferences" conducted by professional bodies, as well as "the extent of deregulations in ICT and the media". Table 4 summarizes the participants' opinions on the knowledge dissemination indicators.

The participants asserted that the current business environment requires organizations to deal with information overload rather than information shortage. A key recommendation was thus to promote the

Table 4. Discussion on the knowledge dissemination indicators.

Original Indicators	Comments	Proposed Indicators	Rationale
ICT Spending as a % of GDP	Not Valid	Government spending on ICT	Government spending reflects spending on ICT infrastructure for the economy
		Business spending on ICT	Business spending reflects spending on ICT products for knowledge dissemination
Internet Access cost as a % of per capita GDP	Not valid	% of Internet penetration in homes	Internet penetration rather than affordability is currently more relevant for Singapore
% of workforce with at least secondary school education	Not valid	Major newspapers circulation figures	Newspapers is an important medium for knowledge dissemination
		Number of seminars, conventions and conferences	Specialised knowledge is disseminated to people who have an interest in the domain
		The extent of deregulations in ICT and the media	Measure the ease of information flow

ability to discriminate noise from useful information and to strategically transform information into competitive advantage for the economy.

4.4. *Knowledge Application Indicators*

The participants commented that the knowledge application indicators were too restrictive in scope. For example, while the percentage of the university graduates in the workforce could indicate the prevalence and intensity of knowledge application in the economy, cases where innovative business models developed by successful entrepreneurs with only high school education would not have been tracked. Given that the current economic emphasis in Singapore is on entrepreneurship and wealth creation, the "number of start-ups in Singapore" was proposed as an indicator. Next, the "presence of venture capital funds" merely indicates a ready pool of finances in Singapore but does not imply actual investment activities. Moreover, there could be foreign-based venture capital funds investing in Singapore-based companies located overseas. Thus, the indicator "the value of funds invested in Singapore-based companies" was suggested instead. The participants also recommended three other indicators, namely "the monetary value of Intellectual Property (IP) commercialized by companies in Singapore", "the number of companies applying for the Initial Public Offering in Singapore", and "the return of investment (ROI) spent on training". Table 5 summarizes the participants' opinions on the knowledge application indicators.

Table 5. Discussion on the knowledge application indicators.

Original Indicators	Comments	Proposed Indicators	Rationale
% of workforce with university education	Valid but restricted in scope	Number of start-ups	Measures the entrepreneurial activities in Singapore
Presence of venture capital funds	Not valid	Monetary value of funds invested in Singapore-based companies	Measure the total amount of money invested in the economy
World Competitiveness Yearbook rating of entrepreneurship	Valid	-	-

Table 5. (*Continued*)

Original Indicators	Comments	Proposed Indicators	Rationale
		Monetary value of Intellectual Property (IP) commercialized by companies in Singapore	Measure the extent companies apply IP to generate wealth
		Number of companies applying for Initial Public Offering in Singapore	Measure the expansion of companies in Singapore which indirectly measures the effectiveness of knowledge application
		Return of investment (ROI) spent on training	Measure the knowledge application capability of the workforce

The participants commented the lack of connectivity between knowledge creation and knowledge application activities as the main weakness in the Singapore economy. For example, they cited cases in which some start-up firms possessed good ideas but did not have access to people with the experience to help commercialize those ideas. Thus, participants offered a recommendation to create platforms for successful businessmen and businesswomen to act as mentors to budding entrepreneurs.

5. Conclusion

This paper has reviewed several major KBE assessment models with the intention to develop an interim set of KBE indicators for Singapore. With inputs solicited from experts from the Economic Development Board of Singapore, the indicators were refined to be more encompassing, as well as better reflect the prevailing local economic situation. For example, the indicators now balance the importance of both high technology and non-technical sectors. Merger and acquisition activities, the monetary value of intellectual property of Singapore-based companies, and the percentage of Internet penetration in homes are also included in the measurement. Collectively, the indicators serve as the blueprint to help track Singapore's economic development as a KBE. Even so, the longevity of these indicators cannot be presumed. As the world globalizes, technologies advance, and economic forces shift, these indicators have to be periodically reviewed, modified, or even replaced with new ones.

In addition to the development of the KBE indicators, this paper also presents a range of recommendations to enhance Singapore's capabilities in the areas of knowledge creation, knowledge acquisition, knowledge dissemination, and knowledge application. Among them include pro-business initiatives, building strong support networks to promote lifelong learning, and creating platforms for entrepreneurship mentoring.

Hopefully, the modest contribution of this paper in the form of the above KBE indicators could be used as the starting point for scholars to develop other KBE assessment frameworks for economies of their interest. For economic policy makers, the KBE indicators can be used not only for the economy but also to support sectoral-level analysis. In this way, the relative contribution of different sectors to the overall KBE score of an economy can be determined. As greater interest on knowledge measurement is spurred, more economies may be guided in their paths in remaining or becoming a KBE.

Reference

1. Ahmed, P.K., Lim, K.K., & Zairi, M. (1999). Measurement practice for knowledge management. *The Journal of Workplace Learning, 11*(8), 304-311.
2. Asia-Pacific Economic Cooperation, (2000). *Towards knowledge based economies in APEC.* APEC Secretariat.
3. Atkinson, R.D., Coduri, R. (1999). *The new economy index: understanding America's economic transformation.* Washington DC: Progressive Policy Institute.

4. Australian Bureau of Statistics, (2001). *Measuring the knowledge-based economy: a statistical framework for measuring knowledge in the Australia economy and society.* Retrieved on December 20, 2004, from http://www.unescap.org/stat/cos12/wgse12/wgse12-06.asp.
5. Edvinsson, L. and Malone, M. (1997*). Intellectual capital: realising your company's true value by finding its hidden brainpower*. New York: Harper Collins.
6. Goh, C.T. (1997). Shaping our future: "Thinking Schools" and a "Learning Nation". *Speeches*, *21*(3), 12-20. Singapore: Ministry of Information and the Arts.
7. Hiebeler, R. J. (1996). Benchmarking: knowledge management. *Strategy & Leadership, 24* (2), 22-29.
8. Kannan, G. & Aulbur, W.G. (2004). Intellectual capital: measurement effectiveness. *Journal of Intellectual Capital, 5*(3), 389-413.
9. Kaplan, R.S. and Norton, D.P. (1992). The balanced scorecard measures that drive performance. *Harvard Business Review, 70*(1), 71-79.
10. Kluge, J., Stein, W. & Licht,T. (2001). *Knowledge unplugged: the McKinsey & Company global survey on knowledge management,* 63-67, New York: Palgrave.
11. Kochikar, V.P., (2000, September 12-15). *The knowledge management maturity model: a staged framework for leveraging knowledge*. Paper presented at KM World 2000, Santa Clara, CA.
12. Organisation for Economic Co-operation and Development, (1996). *The knowledge-based economy science, technology and industry outlook.* Paris: OECD.
13. Organisation for Economic Co-operation and Development, (2001). *The new economy: beyond the hype.* Final Report on the OECD Growth Project. Paris: OECD.
14. Sigurdson, J. (2000). *Singapore means to turn into a knowledge-based economy,* The Stockholm School of Economics.
15. Toh, M.H., Tang, H.C. & Choo, A. (2002, third quarter). *Mapping Singapore's knowledge-based economy,* Economic Survey of Singapore.
16. World Bank, (1999). *World development report 1998/99: knowledge for development.* Washington: World Bank. Retrieved November 12, 2004, from http://www.worldbank.org/wdr/wdr98/contents.html.

This page intentionally left blank

Chapter 4

A SHANNON'S THEORY OF KNOWLEDGE

Benoît Le Blanc
Institut de Cognitique
Université de Bordeaux 2, 33076 Bordeaux cedex, France
leblanc@idc.u-bordeaux2.fr

Jean-Louis Ermine
Institut National des Télécommunications
91011 Evry cedex, France
jean-louis.ermine@int-evry.fr

Our so-called "Information Society" is gradually changing into a "Knowledge Society". The Shannon's Theory provided a formal and mathematical framework for information. It was very fruitful for avoiding ambiguity on the concept of information. This paper proposes a transposition of this theory for knowledge. From the three axes of a formal model designed for knowledge engineering (information, sense and context), three quantitative measures are proposed to get a measure of the quantity of knowledge of a system. This notion permits to consider applications as the cognitive measure of a web site, of a knowledge community (community of practice…).

1. Introduction

There is now a tremendous focus and lot of energies spent around the notion of knowledge. Of course, meanings around this "buzz word" are extremely various, but it is meaningful of a fundamental change of our societies. We can distinguish several meanings around the concept of knowledge.

- Knowledge Society
- This is a concept popularised by nations, international organisations… (Anonymous, 2000). [ICT] (Norris, 2004)

(Department of Economic and Social Affairs [DESA], 2005). (Mansell & Wehn, 1998; Corniou, 2002; Anonymous, 2001).

- Knowledge Based Economy
- This is a new theory in economics developing the notion of the knowledge as an economic good; Precursors are Nelsons (1959), Kenneth Arrow (1962) Herbert Simon (1982), and especially F. Machlup (1984). The modern theory is exposed in the pioneer book of D. Foray (2004) and experienced in Organisation for Economic Co-operation and Development [OECD] publication (Anonymous, 2004). See also in France the Commissariat Général au Plan [CGP] (Anonymous, 2003).
- Knowledge Management (KM)
- This is a now very increasing domain that becomes complex, see for instance (Bollinger, 2001).
- Knowledge Engineering
- Knowledge Engineering is design methodology for knowledge-based systems (Schreiber, 1999; Studer, 1998) (Dieng et al., 2000).
- Information vs. Knowledge

We may conclude that in the concepts of Knowledge Society, Knowledge Economy, Knowledge Management, Knowledge Engineering, ICT is always strongly present. But ICT is dedicated to information processing.

For more than 50 years, information is a very well known and very definite object, notably by the theory of Shannon that gives an operative mathematical definition that solved the ambiguousness problem concerning its nature (Shannon, 1949). What about knowledge? How can one say that ICT processes knowledge? Can we solve the ambiguousness between information and knowledge, and bring a formal answer, even though necessarily partial?

The Shannon's theory of information is not only a very powerful technical tool, but also a very powerful metaphoric tool, (when used scarcely) (Moles, 1975; Bougnoux, 1993; Eco, 1972). An extension of the Shannon's theory to knowledge would give a formal relationship between knowledge and information, while providing, in the same way, the fertile metaphors, if they are used with discernment. We propose a sketch of what could be a Shannon's theory of knowledge.

2. Knowledge Corpus

2.1. *A Formal Model of Knowledge Corpus in KM*

This model (called AIK model) is an attempt to provide sound basis for the definition of the knowledge capital of organised system. It is described in (Ermine, 2005). This is a mathematical formalism, based on set and morphism theory, general enough to include most of the basic well known concepts of Knowledge Management (Le Moigne, 1990; Morin, 1986; Nonaka & Takeuchi, 1995; Wenger, 1998; Drucker, 1959; etc.). The formal model is summarised in the diagram of Figure 1.

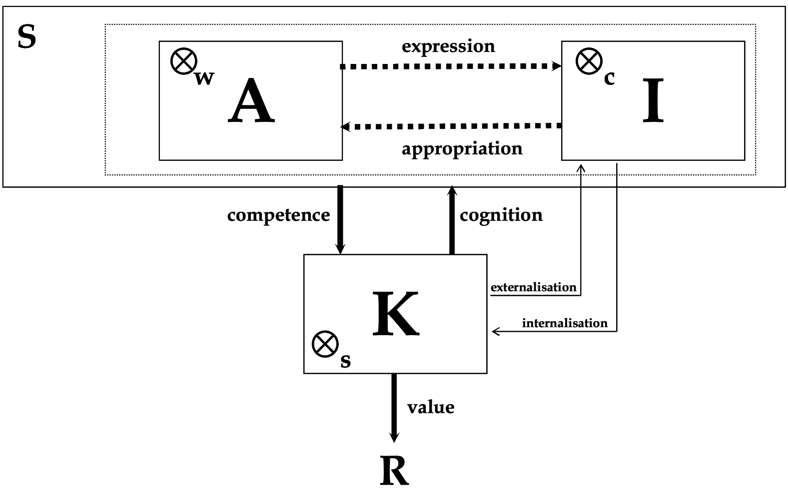


Figure 1. A formal model for KMS.

S is the Knowledge Management System composed of the systems of knowledge workers networks A, and the Information System I. K is the Knowledge Capital. $\otimes_w$ is the Wenger's operator of knowledge community aggregation, $\otimes_c$ and $\otimes_s$ are the combination and socialisation Nonaka's operators, there are also the externalisation and internalisation Nonaka's functions. Competence and Cognition are the cognitive functions of the KMS, according to Edgar Morin's theory of knowledge. Expression and appropriation are for knowledge workers, in their relationship with information system.

The value function is a real valuation function that permits to evaluate the added value brought by the KMS.

In the AIK model the function value is essential. It may be expressed as the added value in term of "knowledge capital"; it is the hypothesis of knowledge economy (OECD, 2004) as a strategic assessment function, which gives a grade on a scale of "criticality" (or of risk assessment of the risk) for the knowledge capital (Aubertin, 2006).

The proposed approach is quite different. It starts from the elementary idea, for example, that a data base is made to collect data, and that the value function grows if the quantity of information accumulated in this basis grows. However one knows, thanks to the theory of information, to give a meaningful measure (otherwise applicable) of the quantity of information, calculated in bytes for example. Can one have a similar measure for a knowledge capital? We try here to give some answers to this question.

2.2. *A Formal Model for Knowledge Corpus in Knowledge Engineering*

The Knowledge Macroscope is a tool to structure the knowledge capital of an organised system. That is a kind of knowledge theory that involves a lot of different aspects that have been studied on knowledge and information through times. It has been fully described and justified in (Ermine, 1996, 2002, 2003).

It is based on the "semiotic hypothesis", considering that knowledge is perceived as a sign, which includes *information* (what is the form, encoded or perceived, of the sign sent to my perception?), *sense* (what semantic representation is generated by the information in my mind?), and *context* (what environment is influencing the sense generated from the information?). Knowledge is information that makes sense in a given context. If the notion of information is clear, referring to Shannon's theory, the notion of sense (signification or semantic), and the notion of context are far less clear. Then we have to give "measures" for what is supposed to be sense or context. It is not a simple task. We will give some first propositions in that direction.

2.3. *Definitions and Notations*

According to semiotic theory, as quoted above, knowledge can be split into three parts, corresponding to the three points of views of information, sense and context:
I is the Information space, S_e the semantic (sense) space, and C_o the space of context. We have:

$$K = I \times S_e \times C_o$$

$$\forall k \in K, \ \exists \ i \in I, s \in S_e, c \in C_o \quad : \ k = (i,s,c)$$

Then, for $k \in K$, we can define three real valued fucntions:

$$\mathrm{Val}_I : I \longrightarrow \mathbf{R}$$

$$\mathrm{Val}_S : S_e \longrightarrow \mathbf{R}$$

$$\mathrm{Val}_C : C_o \longrightarrow \mathbf{R}$$

When valuating respectively the value of information, the semantic value, and the contextual value of knowledge, by composition we have the global value of knowledge by:

$$\mathrm{Val}(k) = F(\mathrm{Val}_I (k), \mathrm{Val}_S (k), \mathrm{Val}_C (k))$$

We generalise that definition of knowledge value to a knowledge set, we will call "Knowledge Corpus":

Definition: A Knowledge Corpus H is an element of $\mathcal{P}(K)$ (subsets of K).
And by extension, we define:

$$\mathrm{Val}_X(H) \mid \int_{h\subset H} \mathrm{Val}_X(h)dm$$

for a measure m such that t(he integral is convergent).

Then we can develop a "limited expansion" of the Val function, Val = $F(\mathrm{Val}_I , \mathrm{Val}_S , \mathrm{Val}_C)$ that we will limit to first order (with an unknown operator $\otimes$ to be defined):

$$\mathrm{Val}(H) = \mathrm{Val}_I(H) \otimes \mathrm{Val}_S(H) \otimes \mathrm{Val}_C(H) + o(H)$$

We will simplify by vanishing the residual term, and we will suppose that:

$$Val(H) = Val_I(H) \otimes Val_S(H) \otimes Val_C(H)$$

Intuitively, this means to suppose that information, sense and context may vary independently: one can imagine a knowledge corpus that has a lot of information and few sense (a telephonic directory, for example), a lot of sense and little information (a proverb, for example), a lot of sense but little context –to the sense context of usage, cf. Infra– (an intimate diary, or some blogs, for example), little sense and a lot of context (a tenacious and rife rumour, for example) etc.

3. The Measure of the Quantity of Information of a Corpus

The Shannon's theory of information permits to define what a quantity of information is (Shannon, 1948; Shannon & Weaver, 1949). This theory is a probabilistic point of view on information produced by a system.

During the communication process, the receptor is waiting of a certain message. Let's take the case of a traffic light. When a person looks at this light, he already has an idea of messages transmitted by this light. A priori, he is unaware of what message is precisely going to be transmitted. However, thanks to his experience, he expects to receive some messages with different probabilities.

If we consider that all messages that can be transmitted potentially by the traffic light had the same probability, the probability of each of them would P = 1/8; the 8 possible cases being simultaneous switch on of 0 lamp (1 case), 1 lamp (3 cases), 2 lamps (3 cases) or 3 lamps (1 case). As this value 8 also can be also written as 2^3 and that it is a binary coding (lamps are either on or off), the quantity of information associated to a traffic light, without other precision, is 3. Intuitively it means that there are three pieces of information, every lamp being considered independent of the two others.

In the real word, it is of course different. Only four cases occur (let's take a virtual example in France, it is different in other countries): the red lamp on (45% of cases), the orange lamp on (9.5% of cases), of the green lamp on (45% of cases), or no lamp on (0.5% of cases). This example reflects the reality of information systems of for which the equally probable is a configuration... very unlikely! The quantity of Q information of a message m with a probability of occurrence P is given by the formula (called entropy formula):

$$Q(m) = -P*\log_2(P)$$

In that case, the calculation gives: Q(m) = 1,4

One sees on this example that there is a one to one relation between the uncertainty, in terms of probability, of a receptor, relatively to a determined message, and the quantity of information contained in this message. The notion of quantity of information is replaced then advantageously by the notion of entropy that is a mean information quantity in the sense of probabilities, calculated on the set of messages. So if p_i is the probability of occurrence of the message m_i, the entropy is by definition the mathematical variance:

$$H = -p_1\log(p_1)-p_2\log(p_2) -......$$

This notion requires an important commentary. The more entropy is low, the more informative is the system. In fact the more a message is unlikely, the more it is informative (the message of the assassination of a president is more informative that the message of the fact that there is no snow in Paris in summer!). Then entropy takes the same signification than information, as the possibility of choice for information source, or mean of occurrence probabilities of a set of messages.

In information theory, the introduction of the entropy function was a considerable innovation that was incredibly fruitful. In fact, Shannon introduced this function for simple reasons of regularity, as soon as we have a distribution of probabilities on a set of events (Shannon & Weaver, 1949).

The usual Shannon's entropy is the requested function for $Val_I(H)$ for a knowledge corpus H.

4. The Measure of the Quantity of Sense of a Corpus

4.1. *Definitions and Notations*

With the "bit" ("binary" unit or "binary digit"), we have defined an elementary information unit.

The elementary unit of sense, in linguistics, is called "seme", it is often represented by one significant term, framed of signs /, in order to distinguishes it from the common word (for example /beautiful /, /feminine /, /white / etc…). The semantics of a knowledge corpus is represented therefore by the semes that it contains, but also by semantic links that link these semes. It is that that one calls a semantic network. Works on semantic networks are plethoric, in cognitive sciences, in linguistics or in artificial intelligence. One can define an elementary unit of sense therefore like a set of two semes linked by semantic link. The sense of a corpus K is defined by combinations of elementary units of sense.

The semes of K constitute a finite set of elements (elements or significant terms) S.

Let Ind a function, called indexation function:

$$\mathrm{Ind} : \mathcal{P}(K) \longrightarrow \mathcal{P}(S)$$

that associates to every knowledge corpus H, element of $\mathcal{P}(K)$, a set of semes, element of $\mathcal{P}(S)$.

A semantic graph is a set (V,E) where V is a subset of elements in S, called nodes of the graph, and E a subset of elements of $V \times V$, called vertices or links.

A graph, (we consider only finite graph, with node numbered from 1 to n) is defined by its incidence matrix $P = [p_{i,j}]$: This is a square matrix (n,n), such that $p_{i,j} = 0$ if there is no link between the node i and j, and $p_{i,j} = 1$ otherwise.

A path of length n in a graph is a sequence $(s_0, \ldots, s_n)$, such that (s_i,s_{i+1}) is a vertex of the graph for every i ; s_0 is the origin of the path, s_n is the end.

A graph is called connected if any two nodes may be linked with a path.

We denote $p_{i,j}(n)$ the number of path of length n that starts from the node i and ends at node j . This is also the coefficient at line i and column j of the matrix P^n : $p_{i,j}(n) = (P^n)_{i,j}$.

A function of semantic graph construction is a function:

$$\Gamma : \mathcal{P}(K) \longrightarrow S \times S^2$$

that associates to every knowledge corpus H a semantic graph Γ(H), such that $Pr_1(\Gamma(H)) = Ind(H)$. (This is a semantic graph where all the nodes constitute the set of semes indexing H, Ind(H)).

4.2. *Quantitative Characterization of a Semantic Graph: Gurevich Entropy*

The semantic graph of a knowledge corpus characterizes "the semantic path", the "semantic random walks" that are possible in the corpus. Hence, the topology of the graph characterizes the semantic complexity of the corpus.

Very similar to the information theory, there exists a very developed theory that characterises the random walks in the graph (hence the semantic random walks in the semantic graphs), this is the theory of graph entropy (Simonyi, 1995).

We will give an approach developed by Gurevich (1969, as cited in Ruette, 2001).

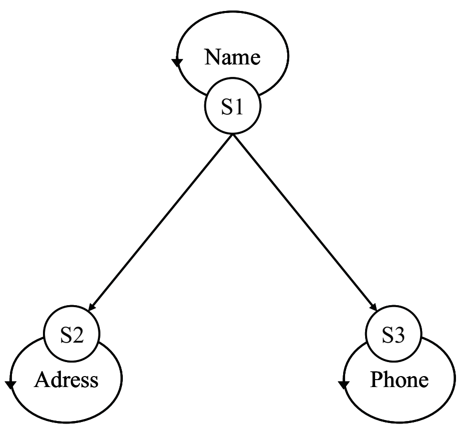


Figure 2. Semantic graph of a phone book.

Let G a graph, P its incidence matrix. Let recall that $p_{i,j}(n) = (P^n)_{i,j}$.

Gurevich entropy is defined as:

$$h(G) = \limsup_{n \to +\infty} 1/n \log(p_{i,j}(n)) = \limsup_{n \to +\infty} 1/n \log((P^n)_{i,j})$$

For a finite connected graph, this number does not depend on i or j. It represents the exponential growth rate of the number of path with fixed extremities.

To better understand that notion of graph entropy, let's take a very simple example: a phone book. The semantics of a phone book very simple, and given by the graph of Figure 2.

Its incidence matrix is:
We can easily calculate for the incidence matrix P that $P^n = (I+A)^n = I + nA$, hence $p_{i,j}(n) = 0$, 1 or n, then h(G) = 0, the entropy of that graph is null. Intuitively, it is true that the semantics of a phone book is very poor!

If we try to add a few semantics more, by designing an inversed phone book, where you can find the name from the phone number, we obtain a new semantic graph. The graph is only augmented with a inverse link from the S3 node to the S1 node. The new calculation shows that $P^n = 2^{n-1} (A + A^2)$ Hence $p_{i,j}(n) = 2^{n-1}$ or 0, and then:

$$h(G) = \lim_{n \to +\infty} (n\text{-}1/n) \log(2) = \log(2) = 1$$

The semantics of the new phone book has increased of one bit of sense!

The Gurevich entropy is the requested function for $Val_S(H)$ for a knowledge corpus H.

5. Measure of the Context of Usage of a Knowledge Corpus

The third part of our knowledge unit is about the measure of its use. We consider this measure of the context as the use that makes people concerned in a knowledge corpus, but also in relations that settle between these people. This idea considers that a pertinent knowledge will be shared between its holder and his knowledgeable neighbourhood or that conversely a weakly distributed knowledge corresponds to something obsolete, uninteresting or inappropriate. We are conscious to overlook, by this approach, all powerful knowledge that remains kept secret. On the other hand, our model may perfectly apply to knowledge that one tries to distribute, via publication media, paper or electronic, via the media for verbal communication, etc. and that won't have any echoes if they don't cause a lot of interest in knowledgeable communities.

The context of use of a knowledge corpus or its diffusion potential of diffusion is studied here through the existing acquaintance network between elements, holders and users of knowledge, individuals, groups, or systems. This kind of network corresponds to the social networks studied in psycho-sociology and more lately in graph theory.

5.1. *Social Networks*

The social networks--called "small worlds"--were initiated by the American psycho-sociologist Stanley Milgram (1967). He postulated then that every person is linked to any other individual while achieving an average of six jumps, materializing thus the theory of "six degrees of separation", invented by the Hungarian writer Frigyes Karinthy in a novel of 1929 entitled "Chains". These experiences have been refuted lately by J. Kleinfeld (2002). Nevertheless, with his very simple and attractive protocol, Milgram shows that it is possible to find an experimental measure of the distance between two random people, by counting the number of necessary mediators to establish a chain between them.

This idea has been studied for specific populations, as for the one of mathematician researchers in mathematical or Hollywood actors. In this case the connectivity between two people corresponds to the realization of a common task: the writing of a scientific article or the apparition in a same movie. See for instance the web site of the Erdös Number Project is: http://www.oakland.edu/enp/. or: http://smallworld.columbia.edu/.

5.2. *Hierarchical Small World Networks*

The underlying model for all these social networks is a graph where each individual (the nodes) is considered regarding his connections (links). From this graph structure, it is possible to calculate some formal values (Hayes, 2000).

According to a similar approach, they are recent studies on others kinds of networks: subway stations, phone connections, flight connections, but also neurons in simple organisms, diffusion of epidemics, Web pages or specific sites etc. In short, all observable network type, produced naturally or generated by a human activity, may be a subject for this kind of measures (Gaume, 2004). It is remarkable that all these studies assign to all these very various graphs the similar properties. One designed a specific class for these graphs: the Hierarchical Small-World Networks [HSWN] (Albert & Barabási, 2002). For example, Adamic and al. advanced in 1999 that the Web is of that type (Adamic et al., 2000) and that average 10 clicks only separated any pair of pages, very weak values compared to billions of pages that compose the Web.

To study HSWNs, Watts and Strogartz (1998) proposed to compare them on the one hand to the random graphs[1] and on the other hand to regular graphs[2]. The two studied parameters are the distance between nodes of the network (global diameter of the graph) and the level of nodes clustering (local coefficient of clustering).

Watts and Strogartz proved that small-world networks have the surprising property to be locally dense (as the regular graphs) and to have a relatively short path length (as the random graphs). A small-world graph is thus halfway between a random graph and a regular graph, thus combining properties of local regularity and global disorder.

5.3. *The Scale-free Networks*

While working on the more general problem of the construction and the organization of networks, Barabási and Bonabeau (2003) added a third relative metrics to the hierarchical distribution of links on these networks. Indeed the growth observed on certain kind of small-world networks (as the Web) shows a property of preferential attachment that cannot be modelled by the simple mean of path lengths, and clustering of

---

[1] An random graph is a graph with N nodes connected by n links, chosen randomly with probability p among the between uncertainly among the N(N - 1)/2 possible links. This graph has p.N(N - 1)/2 links randomly distributed.

[2] the regular graphs have their nodes distributed regularly and each node is linked to its i nearest neighbours, “i” being here the degree (distributed regularly) of connection of every node.

nodes. This new property considers the fact that a new node will stand preferentially connected to nodes that are already greatly connected. This model of accumulated advantage produces a network where most of nodes have few links and some nodes have a lot. The distribution of this nodal connectivity (the probability to have k neighbours) is described by a power law, whereas for the random graphs this probability is described by a Poisson's law. According to Barabasi the coefficient of this power law is a strong characteristic of the network. For the small-world networks, this coefficient is always a number between 2 and 3.

5.4. *Measure of the Context of a Knowledge Corpus*

The scale-free networks seem to be a good candidate to model the social networks that use a given knowledge corpus. It is on the basis of this distribution of connection degrees within a network of users that we propose to construct the measure of the context of a knowledge corpus.

We start from the graph formed by the users of a knowledge corpus, and take as connection between two users the representation of an existing link (an e-mail exchange for example). The obtained network is the network of the users of the considered knowledge corpus. According to the above theory, it is a scale-free network. It is then possible to characterize this graph with the distribution law of distribution of the connectivity degree of the different nodes.

If p(k) is the probability for a node of having k neighbours, the usual Poisson's law gives:

$$p(k) \sim g \exp(-k).$$

For a scale-free network, that law is simplified. The power law gives:

$$p(k) \sim g\, k^{a}\ .$$

Then we can easily define and calculate entropy for the scale-free network with the formula:

$$\mathrm{Val}_C(H) = -\sum_k p(k) \log(p(k))$$

This entropy is a good characteristic of the network of the users of the knowledge corpus. We will take it as the measure of the context of this corpus.

This entropy is the requested function for $Val_C(H)$ for a knowledge corpus H.

6. Perspectives

We now have a proposition for calculating the entropy of a knowledge corpus, by calculating respectively the information entropy (Shannon's entropy), the semantic entropy (Gurevich entropy of the semantic graph of the corpus), and the context entropy (scale-free characteristic of the user network of the corpus). The combination of those three entropies is not yet clear, and is currently under research, to define global knowledge entropy of the corpus.

The definition of a measure of the quantity of knowledge of a corpus is not only a theoretical objective. Defining a quantity of knowledge of a corpus can bring to numerous innovations, as by example:

- "Scoring" for information retrieval. Search engines, from a research using key words, classify corpora found according to their relevance. This relevance can be calculated on the content (occurrence of terms, for example), on the contextual value of the site (number of connections, for example). The finer quantification of the content in semantic term, or indicators on its context of usage, would permit to give a classification far more interesting.
- Improving the content of a document. Information theory permits to find an optimal coding of an informational corpus. Analogically, if one has a measure of the quantity of knowledge of a corpus, of a document, for example, it may lead to rewrite the original document, so that it would be better understood regarding the signification of reference.
- Supervising knowledge communities. The context of a knowledge corpus is essentially valued by its context of usage. Refined indicators users communities for a knowledge corpus permits to better know these communities and to facilitate their improvement.

These some examples show all the potential that one may have with the notion of measure of the quantity of knowledge of a corpus.

References

1. Adamic, L.A., Huberman, B.A., Barabási, A.L., Albert, R., Jeong, H., & Bianconi, G. (2000). Power law distribution of the World Wide Web. *Science*, *287,* 2115a.
2. Albert, R., & Barabási, A.L. (2002). Statistical mechanics of complex networks. *Review of modern physics, 74,* 47-97.
3. Anonymous. (2000). *Conclusions de la Présidence Conseil européen de Lisbonne 23 et 24 mars 2000*. Public report DOC/00/8, European Communauty, Brxelle, Belgique, from http://europa.eu.int.
4. Anonymous. (2001). La société du savoir. *Sciences Humaines Hors-Série, 32*.
5. Anonymous. (2003). *La France dans l'économie du savoir: pour une dynamique collective*. Public report, Commissariat Général au Plan [CGP], La Documentation Française, Paris, France.
6. Anonymous. (2004). *Gestion des connaissances: mesurer la gestion des connaissances dans le secteur commercial, premiers résultats*. Public report, OCDE, Paris, France.
7. Arrow, K. (1962). Economic Welfare and the allocation of resource of inventions. In Nelson (Ed.), *The rate and direction of inventive activity: Economic and social factors*. Princeton, NJ: Princeton University Press.
8. Aubertin, G. (2006). Knowledge mapping: a strategic entry point to knowledge management. In *Trends in Applied Knowledge Management*. US: Hermes Penton, to appear.
9. Barabasi, A.L., & Bonabeau, E. (2003). Scale-free networks. *Scientific American, 288(5).*
10. Bollinger, A.S., & Smith, R.D. (2001). Managing organizational knowledge as a strategic asset. *Journal of Knowledge Management, 5(1).*
11. Bougnoux, D. (1993). *Sciences de l'information et de la communication*. Paris, F: Larousse.
12. Corniou, J.P. (2002). *La société de la connaissance: nouvel enjeu pour les organisations*. Paris, France: Hermès science publications.
13. Department of Economic and Social Affairs [DESA]. (2005). *Understanding Knowledge Societies*. Public report ST/ESA/PAD/SER.E/66, United Nations, New-York.
14. Dieng, R., Corby, O., Giboin, A., Goliebowska, J., Matta, N., & Ribière, M. (2000). *Méthodes et outils pour la gestion des connaissances*. Paris, France: Dunod.
15. Drucker, P. (1959). *Landmarks of tomorrow*. New-York: Harper and Brothers.
16. Eco, U. (1972). *La structure absente, Introduction à la recherche sémiotique*. Paris, France: Mercure de France.
17. Ermine, J-L. (1996). *Les systèmes de connaissances*. Paris, France: Hermès sciences publications.
18. Ermine, J-L. (2002). *Initiation à la méthode MASK (CD-ROM).* Troyes, France: UTT Press.
19. Ermine, J-L. (2003). *La gestion des connaissances*. Paris, France: Hermès sciences publications.

20. Ermine, J-L. (2005). A theoretical and formal model for a Knowledge Management System. In Dan Remenyi (Ed.), *Second International Conference on Intellectual Capital and Knowledge Management,* 187-199, Dubai.
21. Foray, D. (2004). *The Economics of Knowledge*. Cambridge, MA: The MIT press.
22. Gaume, B. (2004). Balades aléatoires dans les Petits Mondes lexicaux. *I3 Information, Interaction, Intelligence, 4(2).*
23. Gurevich, B.M. (1969). Topological entropy of enumerable Markov chains (Russian). *Dokl. Akad. Nauk SSSR*, 187, 715-718. English translation *Soviet Math. Dokl*, 10(4), 911–915.
24. Hayes, B. (2000). Graph Theory in Practice. *American Scientist, 88(1),* 9-13.
25. Kleinfeld, J.S. (2002). Could it be a big world after all? The six degrees of separation myth. *Society, 39(61).*
26. Le Moigne, J-L. (1990). *La modélisation des systèmes complexes*. Paris, France: Afcet Systèmes, Dunod.
27. Machlup, F. (1984). *Knowledge, its creation, distribution and economic significance, vol III*. Princeton, NJ: Princeton University Press.
28. Mansell, R., & Wehn, U. (1998). *Knowledge societies information technology for sustainable development*. Oxford, UK: Oxford University Press.
29. Milgram, S. (1967). The small world problem. *Psychol. Today, 1,* 60-67.
30. Moles, A. (1975). *Préface à "Théorie mathématique de la communication" par C. Shannon et W. Weaver*. Paris, France: Retz-CEPL.
31. Morin, E. (1986). *La méthode, 3. La connaissance de la connaissance*. Paris, France: Editions du seuil.
32. Nelson, R.R. (1959). The simple economics of basic scientific research. *Journal of Political Economy*.
33. Nonaka, I., & Takeuchi H. (1995). *The knowledge-creating company*. Oxford, UK: Oxford University Press.
34. Norris, P. (2004). *Building Knowledge Societies*. Public report, UNESCO, New-York.
35. Ruette, S. (2001). *Chaos en dynamique topologique, en particulier sur l'intervalle, mesures d'entropie maximale*. Doctoral dissertation, Université Aix-Marseille II, Marseille, France.
36. Schreiber, G., Akkermans, H., Anjewierden, A., de Hoog, R., Shadbolt, N., de Velde, W.V., & Wielinga, B. (1999). *Knowledge engineering and management: The CommonKADS Methodology*. Cambridge, MA: The MIT Press.
37. Shannon, C.E. (1948). A mathematical theory of communication. *The Bell System Technical Journal, 27,* 379-423, 623-656.
38. Shannon, C.E., & Weaver, W. (1949). *The mathematical theory of information*. Urbana, IL: University of Illinois Press.
39. Simon, H. (1982). Models of Bounded Rationality. *Behavioural Economics and Business Organisation, 2.*
40. Simonyi, G. (1995). Graph entropy, a survey, combinatorial optimization. *DIMACS, Series on Discrete Math. and Computer Science, 20,* 399-441 (W. Cook, L. Lovasz, P. Seymour eds.), AMS.
41. Studer, R., Benjamins, R.V., and Fensel, D. (1998). Knowledge engineering: principles and methods. *Data & Knowledge Engineering, 25,* 161-197.

42. Watts, D.J., & Strogatz, S.H. (1998). Collective dynamics of "small-world" networks. *Nature, 393,* 440- 442.
43. Wenger, E. (1998). *Communities of practice: Learning, meaning and identity*. New York: Cambridge University Press.

This page intentionally left blank

Chapter 5

"WORKING IN PARALLEL": THEMES IN KNOWLEDGE MANAGEMENT AND INFORMATION BEHAVIOR

Sue Halbwirth
Information & Knowledge Management Program, Faculty of Humanities & Social Sciences University of Technology, Sydney
PO Box 123 Broadway NSW 2007 AUSTRALIA
Sue.Halbwirth@uts.edu.au

Michael Olsson
Information & Knowledge Management Program, Faculty of Humanities & Social Sciences University of Technology, Sydney
PO Box 123 Broadway NSW 2007 AUSTRALIA
Michael.Olsson@uts.edu.au

1. Introduction

This paper brings together approaches, theories and research from two complementary fields: knowledge management and information behavior research.

Against a background of knowledge management in Australia, the paper describes Standards Australia's recently published *AS 5037-2005 Knowledge Management--A Guide* (of which one author was committee chair) as an exemplar of the ways in which knowledge management theory and practice have evolved in recent times. This evolution is mirrored in a review of the literature of the field and manifests a growing recognition of the complex social nature of organizational knowledge cultures.

This is followed by a discussion of developments in the field of information behavior research. In particular, we highlight that, as in knowledge management, an important trend in the field has been a growing awareness of the shortcomings of a focus on individual cognition and the emergence of a range of socio-cultural approaches to understanding information behavior. These include a range of social constructivist perspectives, which make social/discursive context the central focus of theoretical attention.

From this we identify areas of significant common interest between the fields of knowledge management and information behavior research. A case study of the development of a knowledge management research project linking universities and industry partners to explore organizational knowledge cultures highlights a range of theoretical and methodological challenges and opportunities. The research project (case study) draws on the divergent expertise of the authors--one with a background in information behavior research; the other with extensive experience in the practice and teaching of knowledge management. The case study highlights that the information behavior and knowledge management communities have much to learn from--and teach--each other.

The authors hope this paper will contribute to further conversations and encourage other similar collaborative projects.

2. Knowledge Management in Australia

In Australia, as in other developed nations, the knowledge based services sector is a key driver of national wealth. Australian organizations in the private, public and community sectors recognize the importance of knowledge as a resource, an asset and a form of competitive advantage. The concept of knowledge management has developed as a practice: an approach to managing organizations. It is the topic of discussion in academic, business, government, and not-for-profit sectors, and in Australia, continues to fuel an active conference/workshop circuit, community based forums, journal articles, a growing pool of practical case studies, and academic research.

The last four years have seen a maturing of knowledge management as both a management approach and as a research topic. There has been a move away from the 1990 to 2000 techno-centric approaches, which saw major investments in “knowledge systems” to a framework, which relates to building organizational cultures conducive to knowledge creation, sharing and use. The goal is to have sustainable knowledge enabled organizations that are agile and adaptive to changing environments.

Hasan and Handzic (2003), in their collation of recent knowledge management research in Australia, explore ideas of models of KM, generation, transfer, and utilization of knowledge, socio-technological enablers, measurement, and integration with business. The aim of the text is to highlight the relationship between research and practice in knowledge management. The final chapters consider issues and challenges for knowledge management in Australia, and conclude:

> *There is widespread recognition in the KM community of the importance of suitable research methods… …the question of which research methods are the most appropriate for knowledge management research still remains unanswered. This is because KM draws upon diverse research traditions. There is also a lot of tension and misunderstanding between proponents of different paradigms.* (Hasan & Handzic, 2003, 550)

The authors argue that it is indeed the potential of the application of diverse research methods within knowledge management that will ultimately strengthen the field.

3. Knowledge Management Standard

A major development within knowledge management in Australia has been the release of the world's first national standard in the field. Standards Australia is recognized through a Memorandum of Understanding with the Federal Government as the peak non-government standards development body in Australia. In October 2005, Standards Australia concluded a five-year journey by releasing *Knowledge Management—A Guide*, a national standard that describes an approach for Australian organizations to effectively leverage their knowledge to innovate, learn and respond to changes in an ever-increasing competitive market place.

In 2000 Standards Australia assessed that knowledge management was a topic of interest with the emergence of pockets of understanding and growing expertise in a few organizations, while at the time, increased confusion in many others. Standards Australia began an informal consultative process to produce a handbook on knowledge management with the objectives of providing clarity and adding value to the Australian knowledge management space. This handbook, *A framework for succeeding in the knowledge era*, (Standards Australia, 2001) was well received and raised the level of debate within knowledge management communities in Australia.

In late 2001 Standards Australia established a technical committee (MB-007) to further develop thinking in knowledge management and move towards the development of an Australian knowledge management standard. Standards Australia practices a consensus based development process; and therefore, the knowledge management technical committee consists of representatives from a diverse set of organizations, including academic institutions, professional and industry bodies, as well as government. The intent was that the committee reflect the diversity and multidisciplinary nature of the field of knowledge management. It was recognized that this standard would represent a new type of standard--one based around describing and clarifying, rather than about prescription and compliance.

February 2003 saw the release of an interim standard *AS 5037 (Int)*. A proactive process to collect feedback from the public on the document followed this release.

While this document began the development of a generic model to describe knowledge management, feedback suggested areas for improvement: the need for more guidance on how to implement knowledge activities, a continuance of the philosophies of embracing diversity of practice, and the idea that "one size does not fit all". The model presented in the interim standard (*AS5037-2003 int*) was perceived by some professionals as "too simplistic" and "rigid".

While acknowledging the role of organizational culture and capability as a key contributor to successful knowledge management, the interim standard was described in some feedback as “mechanistic” and “too linear” in its approach. The interim standard reflects an understanding of knowledge management at a time of change within knowledge management. The publication of the interim standard coincided with a maturing of the approach both in terms of the practice and intellectual thinking. There was a sense in the knowledge management community that the final standard and the visual representation of a “model” needed to “go further”.

Overall, the willingness of those in the knowledge management community to provide their ideas and opinions was encouraging, and the feedback gathered informed the development of the final standard.

In 2005 the final Standard was released, and the major changes in the text highlighted a more fluid, contextual and socio-cultural vision of knowledge management.

The major changes in the revision were:

- increased emphasis on how to understand whether an organization is ready to adopt and/or expand knowledge management activities;
- recognition that organizations are knowledge ecosystems--a complex set of relationships existing between people, process, technology, and content;
- detailed guidance on how to implement the Standard within the context of an organization's environment; and
- identification of emerging issues and trends in knowledge management.

The 2005 Standard aims to:

- provide an easy-to-read, non-prescriptive guide on knowledge management;
- help individuals and organizations deepen their understanding of knowledge management concepts;
- assist organizations in their efforts to understand the environment best suited for enabling knowledge activities; and
- offer a scalable and flexible framework for designing, planning, implementing, and assessing knowledge interventions.

An area of debate in the development of the standard was the phrasing of the definitions of knowledge and knowledge management. For the purposes of the document, these key terms are defined. There is, however, a disclaimer that recognizes the contextual nature of both knowledge and knowledge management and encourages each reader to develop definitions that suit the organizational context in which they are applying knowledge management.

For the purpose of the Standard the following applies:

> *Knowledge management is a trans disciplinary approach to achieving organisational outcomes and learning through maximizing the use of knowledge. It involves the design, review and implementation of both social and technological activities and processes to improve creating, sharing and applying or using knowledge.*
>
> *Knowledge management is concerned with innovation and sharing behaviours, managing complexity and ambiguity through knowledge networks and connections, exploring smart processes, and deploying people-centric technologies.* (AS5037-2005, p.2)

The Standard has two major themes that show clearly the socio-cultural emphasis.

One is the recognition that an organization is a knowledge 'ecosystem', which is characterized by connections and pathways between people, process, content, and technology. The second is the way in which knowledge is activated in each organization and driven by the context, culture and strategic intent of the organization. This highlights that the socio-cultural dimensions of organizations have elements of uniqueness and, therefore, are a determinant for knowledge interventions, and indeed their success and failure.

Figure 1 is adapted from the visual representation of the knowledge ecosystem available in the standard (AS5037-2005). Figure 1 illustrates some of the core components of the ecosystem.

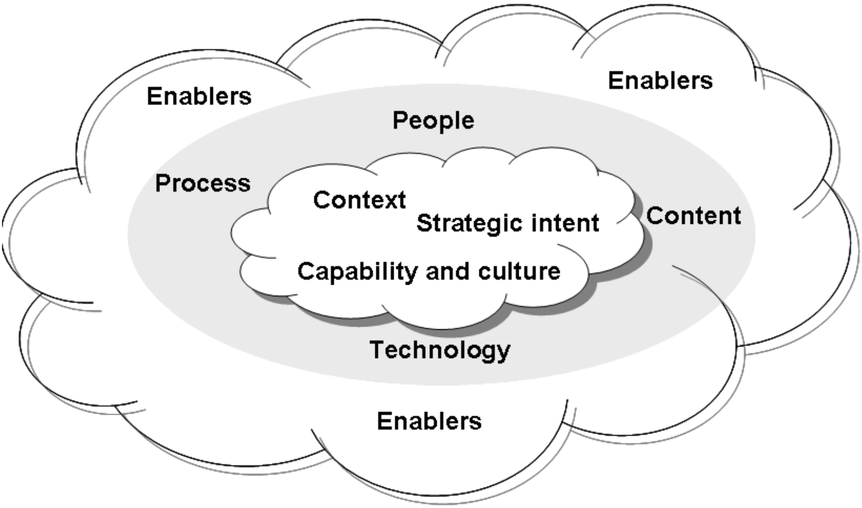


Figure 1. The Knowledge Ecosystem (adapted from AS-5037-2005).

It is important to gain an understanding of the interrelationships within the ecosystem. For example, starting from the outer ring with the introduction of a technology enabler such as intranet is not recommended. "Technology and content that is devoid of content will not deliver a holistic knowledge initiative" (AS5037-2005: 9).

The core of the knowledge ecosystem is

> *"organisational outcomes. These outcomes flow from the contextual environment (culture and strategic intent) and the manner in which an organization operates within the business environment. A major aim of knowledge management is to stimulate and enhance collective organizational skills and competencies."* (AS5037-2005: 9)

To ensure a balanced approach to knowledge management, the Standard provides a three-step methodology that encourages the experimentation and prototyping of interventions. This methodology of Map/Build/Operationalize is supported by a range of enablers.

The suggested methodology is cyclical and includes three phases:

- Mapping context and culture
- Building experiences and linkages
- Operationalizing initiatives and capabilities

The standard also includes practical notes from knowledge management implementations, a description of a range of enablers, measurement and evaluation, and a section which covers six emerging areas: complexity, innovation, the creative economy, sustainability, working in a global culture, and technology.

This new standard gives senior executives and business and government leaders a flexible and iterative approach to ensure the knowledge in their organization is created, shared and applied to grow and strengthen the organization. According to John Tucker, CEO Standards Australia, "this is an important document; it will help leaders make better sense of the world they operate in, and it will help them do their business smarter" (Standards Australia, 2005b).

Future publications from Standards Australia will develop and further explain the content of the Standard for specific sectors. A guide for knowledge management in small medium enterprises is planned for release in 2006. The Standards Australia Knowledge Management Committee believes that the 2005 Standard is a 'living document' that will continue to grow and change and reflect Australian knowledge management practice and thinking.

Knowledge management, like the approach it describes, is fluid and adaptive to the environment in which it is actioned. Therefore, there are a number of themes, which could be described as issues or challenges, with which both practitioners and researchers are grappling as they continue on the knowledge management journey.

4. Knowledge Management Literature

A scan of trends and challenges in the knowledge management literature highlights the emergence of a socio-cultural perspective in the knowledge management sphere (e.g. Snowden, 2002; Wenger et al, 2002), along with a growing recognition of the shortcomings of many earlier knowledge management approaches. Whilst, as Wilson (2002) and Snowden (2002) have highlighted, first generation knowledge management was epistemologically naïve, grounded in an implicit assumption that knowledge was a 'thing' to be captured, recent developments are marked by a growing recognition of the complex social nature of organizational knowledge cultures.

Since its branding in the early 1990's (Prusak, 2001) knowledge management has, as a term, fuelled debate and scepticism. McKinlay states:

> *...KM cannot simply be dismissed as a passing fad. Inevitably, KM will prove to be ephemeral, but the underlying objective of harnessing employee knowledge and creativity will remain of critical importance.* (McKinlay, 2002:76)

While proponents of knowledge management might dispute the 'ephemeral' tag, there is an acceptance of the limitations of the term "knowledge management". In the last few years practitioners have preferred to use the concepts, among others, of enabling and/or facilitating knowledge yet the brand 'knowledge management' remains entrenched.

One key issue that any academic or professional endeavour in this area needs to address is the most fundamental of all issues: is it possible to manage, or even research, knowledge? Wilson (2002) in his review of the knowledge management literature to date, highlighted the theoretical nature of much knowledge management research and practice. In doing so, he cast some much-needed critical light onto the question of the theoretical and epistemological underpinnings of knowledge management.

Wilson (2002) and Snowden (2002) have both highlighted the epistemological naivety of most early knowledge management approaches. Whilst the knowledge management literature in general has had remarkably little explicit discussion of the nature of knowledge, Snowden (2003) has pointed out that:

> *...mainstream* [knowledge management] *theory and practice have adopted a Kantian epistemology in which knowledge is perceived as a thing, something absolute, awaiting discovery through scientific investigation*. (Snowden, 2002, 101)

5. Knowledge Management: The Socio-Cultural Perspective

The recent knowledge management literature has been marked by a growing awareness of the limitations of a techno-centric view of knowledge management. An important feature of this has been a growing appreciation of the importance of context. There is now a widespread acceptance among knowledge management practitioners that both knowledge and knowledge management are different for each organization. There is no one way "to do" knowledge management--no 'one size fits all' solution.

This has led to the development of an increasingly influential *socio-cultural perspective*, or rather perspectives: a range of approaches to knowledge management research and practice that recognize the central role of social factors, such as organizational culture and inter-personal interactions, for knowledge creation, dissemination and use. Socio-cultural approaches to knowledge management are both theoretically and methodologically diverse, drawing on ideas and approaches from a variety of other disciplines.

Evidence for the growth of the socio-cultural perspective can be seen in the growing influence of social network analysis (e.g. Liebowitz, 2005; Schönström, 2005) and communities of practice (Brown & Duguid, 1991; Wenger et al., 2002). Both these approaches are focussed on the central importance of informal communication and social interaction between an organization's members.

The growing recognition that an effective knowledge management strategy needs to facilitate an organizational culture which facilitates creativity and knowledge sharing has led to the development of a range of approaches for examining and 'mapping' an organization's knowledge culture. These range from the structural equation modelling of Lopez et al. (2004) to Boreham & Morgan's (2004) socio-cultural analysis of organizational learning.

The growing sophistication of socio-cultural perspective can be seen in the development of approaches grounded in complexity theory (Snowden, 2002; Snowden & Kurtz, 2003; Sbarcea, 2003). Approaches such as Snowden's 'Cynefin' framework move away from the simplistic linear thinking that has contributed to KM implementations to date. They recognize that organizations need to be viewed as complex adaptive social systems.

Schultze (1999) offered a well thought out analysis of the shortcomings of the prevailing functionalist paradigm, as well as the interpretive assumptions underpinning the influential work of writers such as Brown & Duguid (1991) and Weick (1995). She argues that both of these approaches neglect key aspects of the role of social context in shaping knowledge cultures, in particular that they pay little attention to the role of power relations for knowledge management and the practices of knowing. Schultze argues for the adoption of a critical paradigm drawing on the discourse analytic theories of Foucault.

Post-modern theorists, such as Foucault (1972; 1980) and Lyotard (1984), have as yet had relatively little impact on knowledge management, although there are some notable exceptions such as Chay-Németh (2002) and Sbarcea (2003). However, with the growing influence of socio-cultural approaches and an increasing recognition of the importance of power-knowledge relations, we argue that poststructuralist perspectives can make a valuable contribution to knowledge management research and practice.

This change in knowledge management towards a socio-cultural perspective is well illustrated in the previous discussion about the development of Standards Australia's *AS 5037-2005 Knowledge Management--A Guide*.

6. Information Behavior Research

Information behavior research is a field whose historical origins can be traced to library and information systems evaluation research (Wilson 2000). However, over the last two decades and with the increasing acceptance of a “user-centred paradigm” (Dervin & Nilan, 1986), research has moved away from a narrow focus on information systems use towards a more holistic person-centred investigation of the ways in which people need, seek, interpret, understand, and use information. Contemporary information behavior research is a diverse and multi-disciplinary field, drawing theoretical and methodological insights from a range of other disciplines including philosophy, sociology, cognitive science, communication theory, and linguistics.

It is important to note that in contrast to the knowledge management literature where it has been common-place to associate the word ‘information’ with artefacts (documents, records, literature, etc. e.g. the contents of information systems), information behavior research is heavily influenced by cognitivist and constructivist approaches (e.g. Brookes, 1980; Dervin & Nilan, 1986) and has adopted a much broader definition of the term. Information is seen as ‘that which informs’, whether a text, a lecture, a conversation, or personal reflection. In other words, information behavior researchers have for many years been examining many of the same phenomena that are of central interest to knowledge managers.

7. Parallel Growth: Socio-Cultural Perspective in Information Behavior Research

The shift in focus in knowledge management away from mentalist ‘knowledge capturing’ towards a greater interest in culture and social context, parallels a somewhat earlier (and consequently better established) shift in information behavior research:

> *Approaches to studying information behaviour that focus on social context emerged slowly during the early 1990s and are becoming more prominent. ...social approaches were developed to address information behaviour phenomena that lie outside the realm of cognitive frameworks.* (Pettigrew et al., 2001, 54).

These social approaches to the study of information behavior have included phenomenological and phenomenographic work by e.g. Wilson (2003) and Limberg (1999); Pettgrew's 'Information Grounds' (1999); and social network analysis research as undertaken by e.g. Williamson (1998) and Sonnenwald (1999).

Further, the last decade has seen the emergence of social constructivist approaches to information behavior research, including Chatman's 'life in a small world' and 'life in the round' (1991; 1999); the more recent developments of Dervin's Sense-Making (1999); Savolainen's (1995) use of Bourdieu's 'Mastery of Life' and the discourse analytic work of Talja (2001), Given (2003), McKenzie (2003) and Olsson (2004; 2005). These approaches consider social context not only as a factor influencing the individual information user's cognitive processes, but as the primary focus of theoretical attention.

Social constructivists reject both the positivist/Kantian notion that information systems/artefacts can capture and record an objective reality, as well as the Cartesian separation of the physical and mental spheres (Frohmann, 1992). They argue that both information artefacts and individuals' sense-making processes should be seen not in terms of an objective/subjective divide, but rather as 'intersubjective':

> *Our experience of the world, upon which our thoughts about the world are based, is intersubjective because we experience the world with and through others. Whatever meaning we create has its roots in human action, and the totality of social artefacts and cultural objects is grounded in human activity.* (Wilson, 2003, 71)

This approach has therefore focussed its attention on exploring the role of socio-cultural factors such as shared practices, values and beliefs in shaping people's relationship with information. Influenced by Foucault's (1980) theory of 'power/knowledge' (*pouvoir/savoir*), Dervin (1999) and Olsson (2004) have called for greater recognition of the role of power relations in shaping information behavior.

8. Common Ground

So we are now at a point where parallel development means that members of the knowledge management and information behavior research communities have more in common than ever before. Furthermore, both communities have a great deal to offer one another.

The theoretical perspectives and empirical approaches developed by information behavior researchers can make a major contribution to overcoming the shortcomings outlined by Snowden (2002) and Hasan & Handzic (2003), and are well suited to examining a range of issues--such as sense-making, collaboration and informal information/knowledge sharing practices--of strong interest to knowledge management practitioners.

Collaboration with the knowledge management community offers information behavior researchers the opportunity to further develop their ideas and research methods in a variety of different organizational settings and to expand their understanding of the role of context and information/knowledge culture/s by working with a practitioner community, which shares their interests. And, in contrast with information behavior research's perennial struggle for acceptance among IT professionals and systems designers, knowledge managers are already aware that there is more to the information/knowledge world than information systems!

9. Collaborative KM-IBR Research: The University of Technology Sydney (UTS) 'Knowledge Cultures' Project

The Information and Knowledge Management Program at UTS is currently involved in the development of a research project to develop an innovative and practical methodology for assessing and potentially benchmarking the knowledge focus in organizations.

Currently, in both the theory and practice of knowledge management there is a significant gap in tools and techniques for organizations to understand, assess and benchmark their knowledge focus. The challenge for organizations is to understand and assess their knowledge environment/ecosystem. Reflecting on case studies and comparative data will assist them in making decisions about where to direct effort in knowledge management strategies.

The challenge in the research is to combine a quantitative survey tool with a qualitative approach that considers the analysis of case studies of knowledge interventions within an organizational construct. Understandings and methodologies from both knowledge management and information behavior will be used. By adopting a multi-faceted research approach, the research aims to look beyond 'canonical work roles' (Brown & Duguid, 1991) and organizational policy to examine the everyday experience--the information/knowledge life-world (Chatman, 1999)--of members of the partner organization/s. The research aims to identify patterns and groupings, within a variety of organizational contexts, as to what constitutes a 'knowledge culture'. While it can be argued that every 'culture' is unique, the proposed research seeks to identify the patterns that emerge and give insights into the likely outcomes from knowledge interventions. (British Standards Institute, 2003).

The research approach will consider the socio-cultural elements holistically in terms of the organizational context. The methodology developed will seek to explore:

- Communications--modes, stories and meanings
- Practices, behavior, actions--'the way things are done'
- Material culture, objects artefacts, symbols
- Communities, networks, alliance, associations, relationships
- Understanding time and space
- (adapted from British Standards Institute, 2003: 5-6)

In adopting a longitudinal approach over time, the effect of knowledge activities can be "measured" and indicative effects within the knowledge culture can be identified.

The research will focus on developing an understanding of the partner organization's knowledge cultures by tapping into the 'insider' knowledge of participants. This approach will enable the study to examine not only organizational procedures and systems for knowledge sharing, creation and use, but also the less visible (but vitally important) area of informal communication/knowledge sharing amongst and between communities of practice. The research will not only examine participants' actions/work practices, but also the shared values and beliefs that underpin their behavior, and the cultural practices within the organization by which knowledge culture/s are shared, contested and changed.

Methodologically, the research draws on the research traditions of both information studies and knowledge management and adopts/adapts methods and approaches from communication research, discourse analysis, sociology, and ethnographic research. A key aspect of the methodology will be the use of the Sense-Making, Time-Line and Life-Line interviewing techniques developed by Dervin and her collaborators as a means of developing a holistic understanding of the communication/information life-worlds of participants (Dervin 1992, 1999).

10. Conclusion

With their parallel development of socio-cultural perspectives, the areas of common interest--and thus the opportunities for mutually beneficial cooperation between knowledge managers and information behavior researchers--have never been greater. We hope that in presenting this paper at the International Conference on Knowledge Management, the field's premier conference, that others might also be encouraged to 'bridge the divide' and develop joint projects that will increase the understanding (dare we say, the knowledge?) of both communities.

References

1. British Standards Institute (2003). *Managing culture and knowledge: a guide to good practice* (PD 7501) London: British Standards Institute.
2. Brookes, B. (1980). The foundations of information science. Part 1: philosophical aspects. *Journal of Information Science*, *2*, 125-133.
3. Brown, J. S. & Duguid, P. (1991). Organizational learning and communities-of-practice: towards a unified view of working, learning and innovation. *Organization Science*, *2* (1), 40-57.
4. Chatman, E. (1991). Life in a small world: applicability of gratification theory to information-seeking behavior. *Journal of the American Society for Information Science*. *42*, 438-449.

5. Chatman, E. A. (1999). A theory of life in the round. *Journal of the American Society for Information Science, 50*, 207-217.
6. Chay-Németh, C. (2002). Power-maps for organizational change: Foucault's archaeology as a hypothetical model for mapping strategies of power and resistance. *Journal of Information & Knowledge Management*, *2*(1): 47-52.
7. Dervin, B. (1992). From the Mind's Eye of the User: the sense-Making qualitative-quantitative methodology. In: J. Glazier & R. Powell. *Qualitative Research in Information Management.* Englewood, CO.: Libraries Unlimited: 61-84.
8. Dervin, B. (1999). On studying information seeking and use methodologically: The implications of connecting metatheory to method. *Information Processing & Management,* 35, 727-750.
9. Dervin, B. & Nilan, M. (1986). Information needs and uses. *Annual Review of Information Science and Technology*, *21*, 3-33.
10. Foucault, M. (1972). *The Archaeology of Knowledge*. London: Tavistock.
11. Foucault, M. (1980). *Power/Knowledge: Selected Interviews and Other Writings 1972-1977*. London: Harvester Press.
12. Frohmann, B. (1992). The power of images: a discourse analysis of the cognitive viewpoint. *Journal of Documentation*, *48*, 365-386.
13. Given, Lisa (2002). Discursive constructions in the university context: social positioning theory & nature of undergraduates information behaviours. *The New Review of Information Behaviour Research*, *3*, 127-142.
14. Hasan, H. and Handzic, M. (2003). *Australian studies in knowledge management.* Wollongong: University of Wollongong Press.
15. Liebowitz, J. (2005). Linking social network analysis with the analytic hierarchy process for knowledge mapping in organizations. *Journal of Knowledge Management, 9*(1), 76-86.
16. Lyotard (1984) *The postmodern condition: a report on knowledge.* Minneapolis: University of Minnesota Press.
17. López, S. P, Peón, J.M.M. & Ordás, C.J.V. (2004) Managing knowledge: the link between culture and organizational learning. *Journal of Knowledge Management*, *8*(6), 93-104.
18. McAdam, R. and McCreedy, S. (2000) A critique of knowledge management: using a social constructionist model. *New Technology, Work and Employment*, *15*(2), 155-168.
19. McKenzie, P. J. (2002). Connecting with information sources: how accounts of information seeking take discursive action. *The New Review of Information Behaviour Research*. *3*, 161-174.
20. McKinlay, A. (2002). The limits of knowledge management. *New Techonology, Work and Employment*, *17*(2): 76-88.
21. Olsson, M. (2004, September 21-24). Understanding Users: Context, Communication and Construction Paper presented at the ALIA 2004 Biennial Conference Gold Coast Convention & Exhibition Centre. Retrieved March 25, 2006, from http://conferences.alia.org.au/alia2004/pdfs/olsson.m.paper.pdf.
22. Olsson, M. (2005) Meaning and authority: the social construction of an 'author' among information behaviour researchers. *Information Research*, 10(2) paper No. 219. Retrieved March 25, 2006, from http://InformationR.net/ir/10-2/paper219.html.

23. Pettigrew, K. (1999). Waiting for chiropody: contextual results from an ethnographic study of the information behaviour among attendees at community clinics. *Information Processing and Management*, *35*, 801-807.
24. Pettigrew, K. E., Fidel, F. & Bruce, H (2001). Conceptual frameworks in information behaviour. *Annual Review of Information Science and Technology*, 43-78.
25. Prusak, L. (2001). Where did knowledge management come from? *IBM Systems Journal*, *40*(4), 1002- 1007.
26. Quinn, B., Anderson, P. and Finkelstein, S. (1996). Managing professional intellect: making the most of the best. *Harvard Business Review*, March-April, 71-80.
27. Savolainen, R. (1995). Everyday life information seeking: approaching information seeking in the context of "Way of Life". *Library and Information Science Research*, *17*, 259-294.
28. Sbarcea, K.V. (2003). Knowledge management at the edge of chaos: using complexity science to manage organisations in the postmodern era. Retrieved February 15, 2006, from http://www.thinkingshift.com/page.php?key=17.
29. Schönström, M. (2005). Creating knowledge networks: lessons from practice. *Journal of Knowledge Management,* 9(6), 1729.
30. Schultze, U. (1999). Investigating contradictions in knowledge management. Retrieved February 15, 2006, from http://is.lse.as.uk/helsinki/schultze.pdf.
31. Senge, P. (1990). *The fifth discipline: the art and practice of the learning organization.* New York: Doubleday.
32. Snowden, D. (2002). Complex acts of knowing: paradox and descriptive self-awareness. *Journal of Knowledge Management*, 6(1), 100-111.
33. Snowden, D.J. & Kurtx, C.F. (2003). The new dynamics of strategy: sense-making in a complex and complicated world. *IBM Systems Journal.* 42(3), 462-483.
34. Sonnenwald, D. H. (1999). Evolving perspectives of human information behaviour: contexts, situations, social networks and information horizons. *Exploring the Contexts of Information Behaviour.* T. D. Wilson and D. K. Allen. London: Taylor Graham**,** 176-190.
35. Standards Australia (2001). *Knowledge management: a framework for succeeding I the knowledge era*. (HB 275-2001). Sydney: Standards Australia.
36. Standards Australia (2003). *Knowledge management* (AS 5037-2003 int.). Sydney: Standards Australia.
37. Standards Australia (2004). *Knowledge management terminology and readings—an Australian guide.* (HB 189—2004). Sydney: Standards Australia.
38. Standards Australia (2005a). *Knowledge management--a guide* (AS 5037-2005). Sydney: Standards Australia. Available from: www.standards.org.au.
39. Standards Australia (2005b). *Knowledge management standard media release,* November 2005. Sydney: Standards Australia.
40. Talja, S. (2001). *Music, culture and the library: an analysis of discourses.* Lanham, MD: Scarecrow Press.
41. Wenger, E., Mc Dermott, R. and Snyder, W. (2002). *Cultivating communities of practice.* Boston: Harvard Business School Press.
42. Williamson, K. (1998). Discovered by chance: the role of incidental information acquisition in an ecological model of information use. *Library & Information Science Research*, 20(1), 23-40.

43. Wilson, T.D. (2002). The nonsense of 'knowledge management'. *Information Research,* 8(1), paper No. 144. Retrieved February 15, 2006, from http://InformationR.net/ir/8-1/paper144.html.
44. Wilson, T.D. (2003). Alfred Schutz, phenomenology and research methodology for information behaviour research. *New Review of Information Behaviour Research, 3,* 71-81.

This page intentionally left blank

Chapter 6

ALTRUISTIC LEADERSHIP: THE POWER OF KNOWLEDGE CREATION

Maria Sarabia
Department of Business Administration
University of Cantabria
Avda. de los Castros, s/n
39005 Santander-Cantabria, Spain
sarabiam@unican.es

José M. Sarabia
Department of Economics
University of Cantabria
Avda. de los Castros, s/n
39005 Santander-Cantabria, Spain
sarabiaj@unican.es

Knowledge creation is presented as a strong change of power within an organization structure. A knowledge framework must be built in order to understand how leadership arises. Leadership is built upon a four-phase process of organizational knowledge creation (Nonaka & Takeuchi, 1995) as well as both an ontological (levels of knowledge) and an epistemological (tacit and explicit knowledge) dimension. Thus, two leaderships are identified: change and altruistic. This paper establishes a relationship between Herbert Simon's sociability concept and Nonaka and Takeuchi's socialization process in order to define an altruistic behavior in organizational leadership. Organization needs altruistic leaders in order to create a balanced firm once knowledge creation has been developed by the change leader. The need for altruistic leadership is the consequence of the power of knowledge creation.

1. Introduction: Building Knowledge Framework

Different economic theories have studied knowledge as an interesting power factor. According to Penrose (1959), "economists have, of course, always recognized the dominant role that increasing knowledge plays in economic process”. How to acquire and utilize knowledge is viewed from different perspectives. For example, Marshall (1965), a classical economist, held that capital is formed by organization and knowledge to a great extent: “knowledge is our most powerful engine of production”.

The Austrian school of economics by Hayek and Schumpeter showed knowledge in economic affairs. Hayek (1945) classified knowledge into scientific or context-specific knowledge, while Schumpeter (1951) emphasized the importance of combining explicit knowledge. In fact, Schumpeter pointed out combinations of knowledge from new products, production methods, and organizations.

Penrose (1959) focused on the growth of individual firms through their use of mental models to appraise strengths and weaknesses. In this way, a firm could find these images in its experience and knowledge. At this moment, knowledge is related to the growth of the firm, but is not included in the organizational mechanism through which a firm’s members can process knowledge.

Nelson and Winter (1977, 1982) defined the concept of knowledge repository. Such knowledge was also recognized as the essence of innovation, but was not linked at that moment to the creation of technological knowledge in organizational processes.

The evolution of knowledge concept has tried to find a scientific line (Taylor, 1911) that will reduce knowledge into rules for daily work, a humanistic line (Mayo, 1933), and developing social human skills to facilitate organizational relationships.

The scientific and humanistic management views were synthesized by Barnard (1938), who emphasized the importance of behavioral knowledge in management processes. Polanyi (1966) overemphasized this behavioral knowledge, or non-linguistic mental process, defining a tacit viewpoint of knowledge. Simon, inspired by previous work in conjunction with Barnard and March (1958), in which they built a scientific theory of problem solving and decision-making based on the concept of bounded rationality, included human thought in his computer model. Simon (1973) further argued that knowledge is used in deciding a course of action and in consequences in each formulated strategy used by executive managers. Human potential for creating knowledge was neglected for the moment.

Following the evolution to the present knowledge concept, Porter (1980, 1985) developed a framework for analyzing competitive advantages in the firms thanks to his famous five-force model and his value chain model. Both models assumed the relevance of knowledge in organizational strategy. But Drucker (1993) was the visionary who suggested the term of knowledge society and the role of a knowledge worker. In this sense, Quinn (1992) established the key points for the configuration of intangible values (technological know-how).

Knowledge and a firm's capacity for learning represent the cure for many organizations, which suffer an accelerated technological change. An organization's need to adapt themselves to change is defended by Argyris and Schön (1978) with two kinds of learning: single-loop and double loop. Senge (1990) also proposed the learning organization as a new paradigm. At the same moment, Prahalad and Hamel (1990) offered a new approach based on resources as competencies, capabilities, skills and strategic assets. They defined the sustainable competitive advantage over core competences of the firm.

Culture is observed by building a knowledge framework and linking it with learning. Schein (1985) argued that culture is a learned product of group experience. Thus, Pfeffer (1981) defined organizations as systems of shared meanings and beliefs. Following Nonanka and Takeuchi, (1995) organizational culture is observed in beliefs and knowledge shared by members of a firm. As we seek to discover how organizations create new products and new internal processes, a new concept is defined as very important: knowledge. How organizations create new knowledge to make such creations possible, presents a fundamental need for the firm (Nonaka & Takeuchi, 1995).

This paper will, then, look at Nonaka and Takeuchi's view of knowledge creation and leadership building based on the analysis of created knowledge, and it will present the altruistic leadership as a new behavior.

2. Organizational Knowledge Creation: Nonaka and Takeuchi's Theory

Nonaka and Takeuchi's (1995) knowledge creation takes place at three levels: the individual, the group, and the organizational levels. Their viewpoint is very important in order to understand how knowledge can build leadership. In reviewing their theory, the two forms of knowledge interaction and the levels of knowledge creation must be analyzed in the knowledge creation process.

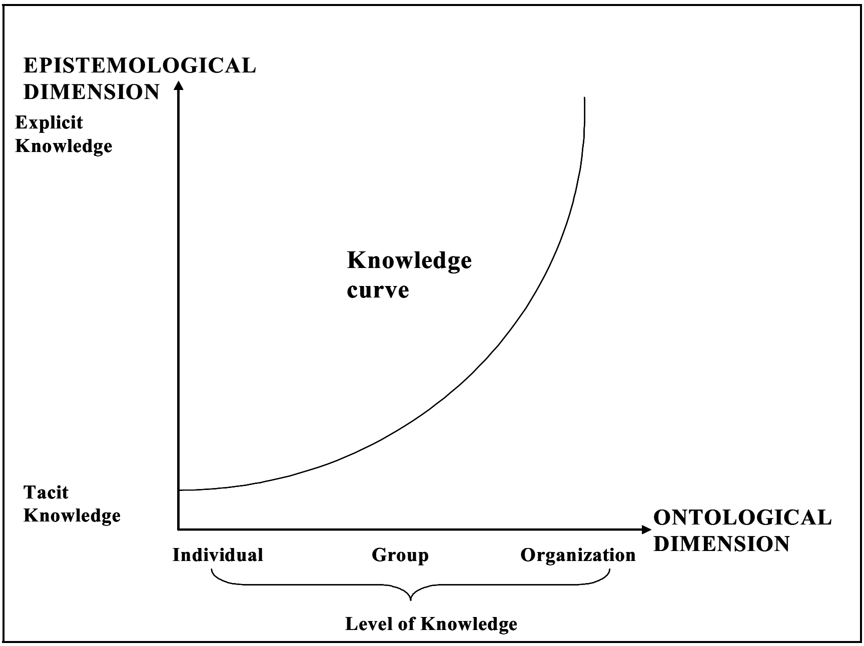


Figure 1. Two dimension of knowledge creation. Adapted from Nonaka & Takeuchi, 1995.

Tacit and explicit knowledge are represented through the different modes of knowledge interaction. These knowledge modes are presented by the epistemological dimension and the ontological dimension. In the epistemological dimension, knowledge derives from its distinction between tacit and explicit. In the ontological dimension, organizational knowledge creation is understood as a process that amplifies the knowledge created by individuals and crystallizes it in the knowledge network (see Figure 1). Thus, a knowledge curve results from the action of the two dimensions and explains how knowledge is created from individual to organization, from tacit to explicit.

One of the four modes of knowledge conversion represents how personal, context-specific or tacit knowledge can be changed to formal, codified or explicit knowledge. This process is explained by Nonaka and Takeuchi as knowledge interplay (see Figure 2). Thus, socialization (tacit to tacit), externalization (tacit to explicit), combination (explicit to explicit), and internalization (explicit to tacit) are the four ways of interaction in knowledge interplay.

But knowledge creation is not yet defined with the two forms of knowledge interaction and the levels of knowledge creation. The process through which knowledge is created over time within an organization requires a four steps structure (see Figure 3) that comprises: sharing knowledge (it is in the individuals), creating concepts (tacit into explicit), justifying the concepts (for verifying their viability), and building an archetype (from prototype to model).

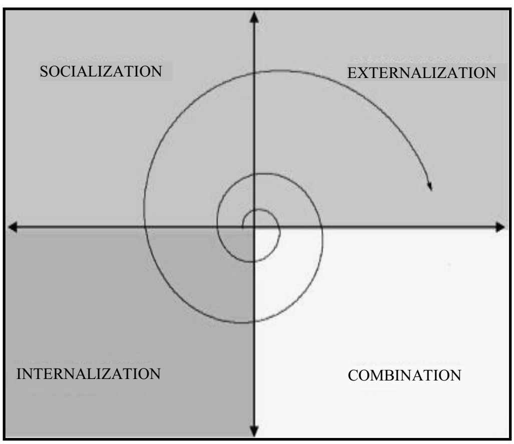


Figure 2. Knowledge interplay. Adapted from Nonaka & Takeuchi, 1995.

Figure 3. Four-phase process of organizational knowledge creation. Adapted from Nonaka & Takeuchi, 1995.

3. Building Leadership from Knowledge Creation

Learning and culture have been discussed due to their relationship with knowledge: Garvin (1993) asserts that a learning organization is an expert in creating, acquiring, and transmitting knowledge. Yeung et al. (1999) also establishes that an organization learns when the knowledge of each individual who is part of the group is shared. So, learning needs knowledge for it to be understood.

Culture is the other concept related to knowledge and to learning. According to Nonaka and Takeuchi (1995), knowledge is acquired in culture form from previous organizations. Furthermore, organizational culture represents knowledge shared by members of the organization. So, culture defines a way of learning from older generations. DeLong and Fahey (2000) identify the relationship between culture and learning as culture defining what knowledge is outstanding, the relationships between the levels of individual and organizational knowledge needing culture for their viability, and the context of social interaction that learning needs being built through culture.

The relationship between knowledge, learning, and culture could be defined as step-stacked, where knowledge represents the first and fundamental step, because learning needs knowledge and culture is the result of knowledge repository.

Schein (1985) argued that managing organizational culture is a leader's major contribution to an organization. It is reasonable to consider a relationship between culture and leadership, because, not only do leaders create, maintain, and even change culture, but on occasion it shapes a leader's behavior (Francesco & Gold, 1998).

For example, Daimler-Benz AG, a major German corporation, suffered important losses in 1993 because of an economic recession in 1992-93 and a belated response to increased global competition. Edzard Reuter, CEO of Daimler-Benz initiated significant changes at the company, including reducing the workforce, and manufacturing automobiles outside of Germany to create a global presence. All of these moves surprised organizational members. These changes required restructuring the organization and creating a new organizational culture within a national culture that resisted change (Francesco & Gold, 1998).

By building leadership from knowledge, it is possible to identify a stack of knowledge, which accounts for a developing leader's behavior (see Figure 4). According to Nonaka and Takeuchi, (1995) on the

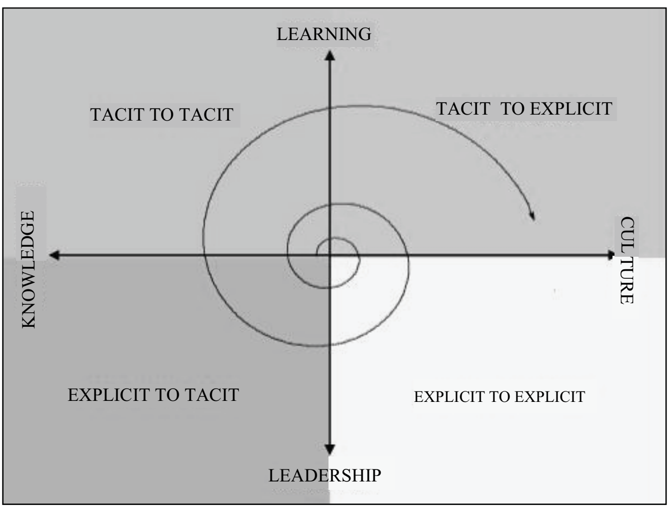


Figure 4. Knowledge stack from knowledge interplay. Adapted from Nonaka & Takeuchi, 1995; Sarabia et al., 2006.

four-phase process of knowledge management, leadership can be defined as sharing knowledge (organizations do not create knowledge but individuals create it in dialogue environment), creating concepts (the first dialogue becomes an exchange of experiences or learning), justifying the concepts (concepts which learning generates are selected by culture, which justifies the viability of each one) and building an archetype (a justified concept becomes tangible and specific and leadership is the consequence of interaction between knowledge, learning and culture) (Sarabia et al., 2006).

In this sense, when the four-phase process of knowledge management, the two forms of knowledge interaction and the levels of knowledge creation are linked, leadership is built as a result of stack of knowledge (see Figure 5).

Following Nonaka and Takeuchi (1995), the knowledge creation process must be understood in a temporal space because knowledge is created over time within the organization. Therefore, leadership created from knowledge must be understood in a particular context where the

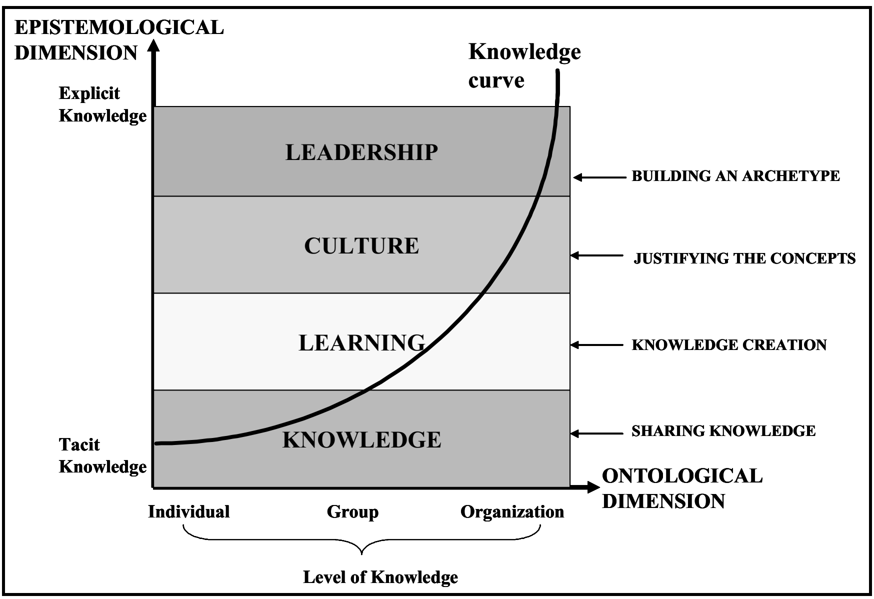


Figure 5. Leadership from knowledge creation. Adapted from Nonaka & Takeuchi, 1995.

organization assumes its knowledge, maintains it, and allows different leadership behaviours to take place within the organization (Sarabia, 2007).

4. Altruistic Leadership: A New Behavior

Simon (1990) defines altruism as behavior that increases the reproductive fitness of others at the expense of the fitness of the altruist. Simon's model of altruism is based on the assumptions of bounded rationality and docility. His viewpoint of human problem-solving processes and rational choices was analyzed in the first section of this paper, and tries to identify knowledge concepts in Simon's information paradigm. It is interesting how Simon argued that the business organization reacts to the environment by adjusting the information-processing structure. However, according to Nonaka and Takeuchi (1995), he missed the idea that the organization acts on the environment, processing and creating information and knowledge by itself.

On the other hand, docility is defined as the tendency to accept and believe information obtained through social channels as a major basis of choice (Simon, 1997). Docility and/or sociability are terms used to define a docile person as one that readily learns in a social setting and tends to acquire socially approved behaviors and beliefs (Knudsen, 2003).

Sociability is related to reliable authority. Individual behavior is learned from other individuals (Simon, 1993). According to Nonaka and Takeuchi (1995) and their knowledge interplay, the knowledge interaction between tacit and explicit, Simon's sociability could be compared to the socialization way of converting knowledge from tacit to tacit. Once socialization is reached, an individual can acquire tacit knowledge directly from others without using language, because the key to acquiring tacit knowledge is experience. The relationship between Simon's sociability and Nonaka and Takeuchi's socialization allow there to be altruistic leadership.

Knowledge is created over time within an organization and its interactions with learning and culture take place in change of leadership and altruistic leadership (see Figure 6). Change leadership defines how a

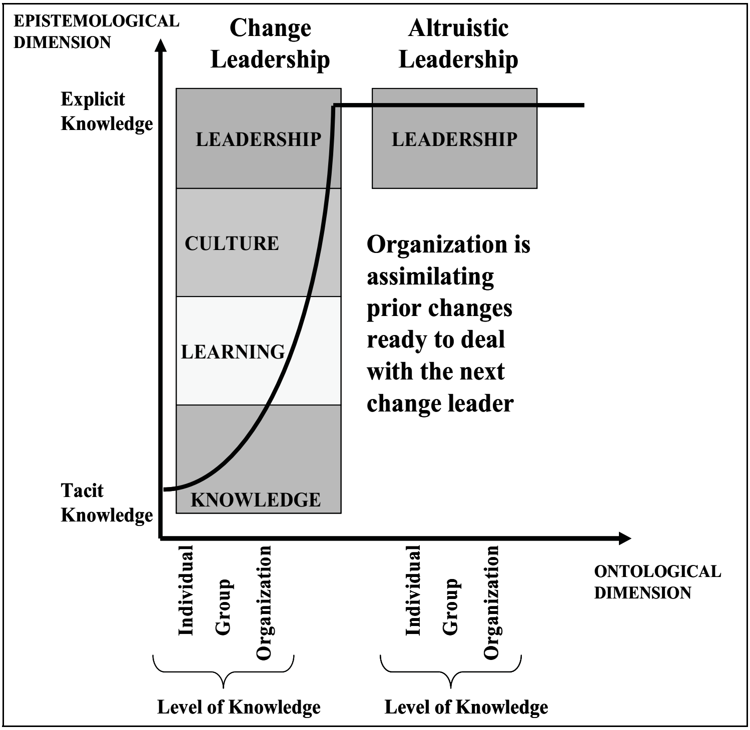


Figure 6. Leadership from knowledge creation: Change and altruistic leaderships. Adapted from Nonaka & Takeuchi, 1995.

leader is able to manage organizational knowledge successfully and also create a successful leadership cycle when problems are unsolvable by the prior knowledge stack. Organization needs change leaders but there is also a need for another kind of leadership, such as altruistic leadership.

Altruistic leadership defines how a leader has to sacrifice his knowledge creation and change cycle due to the organization needing time to assimilate new knowledge, for creating routines for learning and for establishing deep culture, which will possibly develop change leadership. This change may well happen when new and unsolved problems are faced by organization. So, altruistic leadership can only be identified after change leadership when organization is processing new knowledge, new learning and new culture. During altruistic leadership, organization is being impregnated by knowledge created from prior change leadership.

So the characteristics of altruistic leadership are (see Figure 7 and Figure 8): (1) Change leadership must take place before altruistic leadership; (2) altruistic leaders offer balance to an organizational structure that has previously suffered internal changes from new knowledge creation; and (3) two consecutive altruistic leadership cycles do not allow new knowledge, new learning, new culture to develop and so there is no change leader. This business behavior where two altruistic leaderships are consecutive, would represent "paralysis by analysis" because the explicit knowledge would prevail for too much time (Nonaka & Takeuchi, 1995). So, altruistic leaders are important for organization at specific moments.

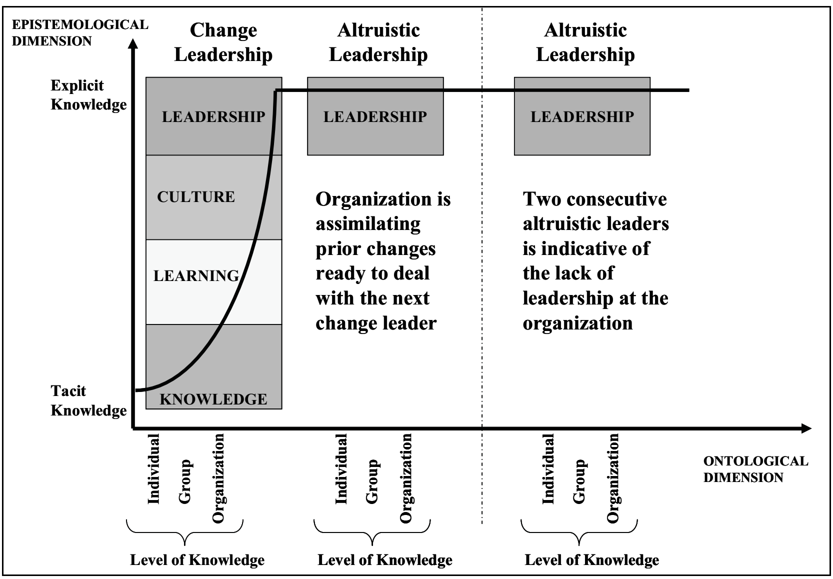


Figure 7. Leadership from knowledge creation: Two consecutive altruistic leadership. Adapted from Nonaka & Takeuchi, 1995.

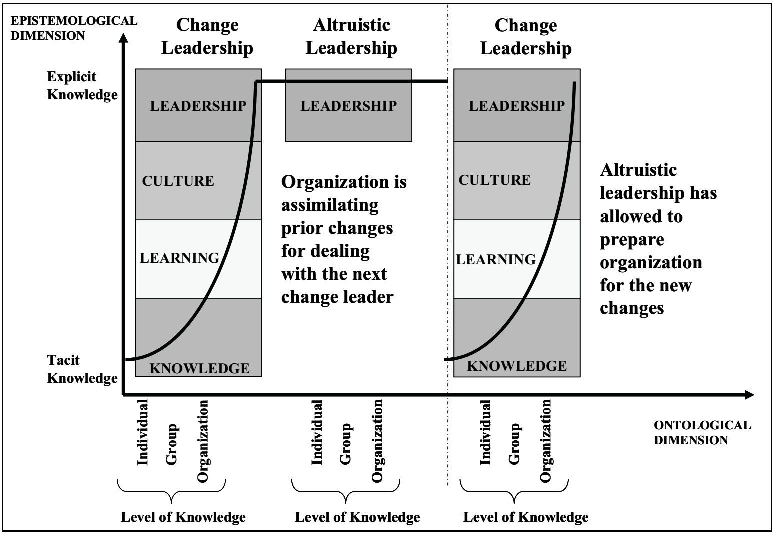


Figure 8. Leadership from knowledge creation: Combination of change and altruistic leaderships. Adapted from Nonaka & Takeuchi, 1995.

5. Conclusions

A Knowledge framework is presented in order to introduce the importance of knowledge creation within an organization. In this sense, Nonaka and Takeuchi's theory plays a central role in knowledge creation, which is the responsibility of leadership building.

Knowledge creation presents the dimensions of epistemological and ontological. The epistemological dimension defines the interaction between tacit and explicit knowledge. The ontological dimension establishes individual, group, and organization as the three levels of knowledge spread. By linking both dimensions to the four-phase process of organizational knowledge creation, leadership is built from the relationship between knowledge, learning, and culture (knowledge stack). Thus, change and altruistic are identified as the two styles of leadership. Change leadership appears when an organization faces new problems, which are unsolvable by the prior knowledge stack. On the other hand, altruistic leadership is built from Simon's sociability and Nonaka and Takeuchi's socialization. Altruistic leadership allows an organization to assimilate prior changes in knowledge stacks made by change leadership. An altruistic leader sacrifices his knowledge creation in order to balance the organization in order to deal with the next change leader.

The power of knowledge creation is so deep, so strong, and so capable of changing organizational structure that altruistic leadership is necessary before a new change leadership cycle can be faced.

References

1. Argyris, C., & Schön, D.A. (1978). *Organizational learning*. Reading, MA: Addison-Wesley.
2. Barnard, C.I. (1938). *The functions of the executive*. Cambridge, MA: Harvard University Press.
3. DeLong, D.W., & Fahey, L. (2000). Diagnosing cultural barriers to knowledge management. *Academy of Management Executive*, *14(4),* 113-127.
4. Drucker, P.F. (1993). The new productivity challenge. *Harvard Business Review*, Nov.-Dec.: 69-79.
5. Francesco, A.M., & Gold, B.A. (1998). *International organizational behavior*. New Jersey: Prentice Hall.
6. Garvin, D.A. (1993). Building a learning organization. *Harvard Business Review*, *71(4),* 78-91.
7. Knudsen, T. (2003). Simon's selection theory: Why docility evolves to breed successful altruism. *Journal of Economic Psychology, 24,* 229-244.
8. March, J.G., & Simon, H.A. (1958). *Organizations*. New York: John Wiley & Sons.
9. Marshall, A. (1965). *Principles of economics*. London: Macmillan.
10. Mayo, E. (1933). *The human problems of an industrial civilization.* New York: Macmillan.
11. Nelson, R.R., & Winter, S.G. (1969). *An evolutionary theory of economic change*. Cambridge, MA: Harvard University Press.
12. Nelson, R.R., & Winter, S.G. (1977). In search of a useful theory of innovation. *Research Policy, 6(1),* 36-77.
13. Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company*. New York: Oxford University Press.
14. Penrose, E.T. (1959). *The theory of the growth of the firm*. Oxford: Basil Blackwell.
15. Pfeffer, J. (1981). Management as symbolic action. In L.L. Cummings, and B.M. Staw, (Eds.), *Research in Organizational Behavior* (Vol. 3, pp. 1-52). Greenwich, CT: JAI Press.
16. Polanyi, M. (1966). *The tacit dimension*. Chicago: Chicago University Press.
17. Porter, M.E.(1980). *Competitive strategy*. New York: The Free Press.
18. Porter, M.E. (1985). *Competitive advantage*. New York: The Free Press.
19. Prahalad, C.K., & Hamel, G. (1990). The core competence of the corporation. *Harvard Business Review*, May-June, 79-91.

20. Quinn, J.B. (1992). *Intelligent enterprise: A knowledge and service based paradigm for industry.* New York: The Free Press.
21. Sarabia, M., Corro, J., & Sarabia, J.M. (2006). Japanese knowledge creation and the fundamental illusion theory: A fresh look. *Journal of Information and Knowledge Management*, *5(1),* 1-11.
22. Sarabia, M. (2007). Knowledge leadership cycles: an approach from Nonaka´s viewpoint. *Journal of Knowledge Management*, *11(4),* forthcoming.
23. Schein, E.H. (1985). *Organizational culture and leadership*. San Francisco, CA: Jossey-Bass.
24. Schumpeter, J.A. (1951). *The theory of economic development*. Cambridge, MA: Harvard University Press.
25. Senge, E.H. (1990). *The fifth discipline: The art and practice of the learning organization.* New York: Doubleday.
26. Simon, H.A. (1973). Applying information technology to organization design. *Public Administration Review*, *33*, 268-278.
27. Simon, H.A. (1990). A mechanism for social selection and successful altruism. *Science*, *250* (December), 1665-1668.
28. Simon, H.A. (1993). Altruism and economics. *The American Economic Review*, *83(2),* 156-161.
29. Simon, H.A. (1997). Models of bounded rationality. *Empirically grounded economic reason.* (Vol. 3). Cambridge, MA: The MIT Press.
30. Taylor, F.W. (1911). *The principles of scientific management*. New York: Harper and Brothers.
31. Yeung, A.K., Ulrich, D.O. Nason, S.W., & Von Glinow, M.A. (1999). *Organizational learning capability*. New York: Oxford University Press.

Chapter 7

A FRAMEWORK FOR MEASURING THE VALUE CREATION OF KNOWLEDGE MANAGEMENT PROGRAMS: FROM THE PERSPECTIVE OF ORGANIZATIONAL KNOWLEDGE SHARING CAPABILITY

Kaisi Chang
Decision Support Group
Univerisity of Cambridge
kc285@cam.ac.uk

Tony Holden
Decision Support Group
University of Cambridge
th103@hermes.cam.ac.uk

The current thriving interests in introducing knowledge management (KM) initiatives have been accompanied by the limited understanding of the link between KM and business value. This paper begins with a critical review of conventional approaches for measuring the value creation of KM programs. We argue that current research uses operational performance improvement as the primary focus for measurement, and that this method is subject to various limitations. A conceptual framework for measuring the value creation—one based on organizational knowledge sharing capability--is then proposed. Based on previous theoretical and empirical research, our study shows how each of the four dimensions of the framework is cultivated through the implementation of KM programs. We then demonstrate how business value can be generated through the improvement of each dimension. Thus, the framework is seen as the vehicle through which KM programs bring value and can also be used as a tool for tracking the implementation of knowledge management programs.

1. Introduction

Although the passing down of knowledge has been carried out for years, it was not until the 1990s that a phenomenal increase in research of KM took place. One of the key arguments to explain the thriving research of KM is that the new knowledge-based economy is fundamentally different from the economic structures of the past, and the "new rules" of competition require new approaches for management (Arthur, 1999). Consistent with this view, many organizations have established their own KM programs. According to KPMG's survey in 2000, 68% of companies either had or were setting up their own KM programs (KPMG, 2000).

However, a basic research query remains largely unanswered: how does KM program contribute to the business performance, and what impact does KM have on organizations that deliver these achievements? It has been observed that KM benefits companies through saving costs, increasing organizational capacity, providing better customer service, and reducing cycle time (du Plessis, 2005; PricewaterhouseCoopers, 1999). Whereas there still lacks a comprehensive framework to explain and evaluate the value-creation of KM programs.

To address this problem we present herein our research experience with a value creation framework for KM programs that builds on previous theoretical and empirical research to provide a conceptual model from the knowledge sharing perspective.

2. Literature Review

2.1. *Management Practices under the Name of "Knowledge Management"*

The variety of management practices that are defined under the name of "knowledge management" is impressive. KM has been defined as "identifying and leveraging the collective knowledge in an organization to help the organization compete" (Alavi & Leidner, 2001). With the broad scope of this working definition, most of the management practices are more or less concerned with KM. However, in the organizational context, it has been observed that the term "knowledge management" has been in many cases biased towards IT (Wilson, 2002). The belief that IT system can separate knowledge from the knower, ensuring wider accessibility and sustainability, is widespread.

Nonetheless, it has been observed that this approach is subject to many limitations. (Brown & Duguid, 2000; Swan, Newell, & Robertson, 2000)

3. Perspectives Towards KM

The perspective that people adopt towards KM determines the value proposition of the management practices. And the most intuitive view towards KM originates from the resource-based perspective of organizations. The measurement for knowledge as assets is motivated by the significant gap between the capital market value and the net asset value of companies (Lev, 2001). Researchers have acknowledged that knowledge, which has surpassed physical and financial assets as the foundational resource for generating profit and sustaining competitive edge, accounts for the huge gap.

Consistent with this view, the concept of Intellectual Capital (IC) is proposed in supplement to the traditional accounting approaches (Biosot, 1998; Lev, 2001). From the resource-based perspective, the objective of KM program is translated into either the increase of IC or the better utilization of IC (Gupta, Pike, & Roos, 2002). It is suggested that managers should focus more on leveraging knowledge-based assets in order to achieve financial goals (Biosot, 1998). In recognition of the importance of IC, Dow carried out a full analysis of the patents it held in 1997, thus identifying key knowledge assets and saving the company $40 million through reducing the patent tax maintenance cost (Petrash, 1996).

Alternatively, the processes-based perspective adopts a more pragmatic view towards KM. KM is defined as practices to get knowledge to the right people at the right time in the right presentation at the right cost (C.W. Holsapple & Joshi, 2002). Thus, researchers attempt to identify knowledge manipulation activities to address the problem of how knowledge is used, transferred, created, and shared to create value (Markus, 2001). Given the fact that the bottom-line contribution of KM, such as cost-saving and better decision making, is closely related with knowledge manipulation process, it is not surprising that many KM programs adopt this approach to measure value creation for the justification of the investment, especially for IT projects. KM programs are mapped into various knowledge manipulation activities that they facilitate, and value creation are measured accordingly (Shin, 2004).

However, it has been observed that knowledge manipulation processes are hard to predict. Looking at the work practices of Xerox technical representatives, Orr found that much of the knowledge transferring was carried out through the personal social network of the technicians cultivated around the water cooler machine, rather than through the formally designed knowledge sharing channels (Orr, 1996). This does not suggest the failure of the knowledge sharing processes; instead it suggests that simply tracking those processes may lead to bias.

Davenport and Prusak describe knowledge holders in an organizational context as "buyer" and "seller" and thus argue for a market-based perspective (Davenport & Prusak, 1998; Grover & Davenport, 2001). Price mechanism is used to explain various aspects for knowledge sharing and especially for knowledge hoarding. Consequently, the main task of KM is minimizing barriers to make the market work efficiently. Management practices create value by reducing the transaction costs for people transferring their knowledge (Grover & Davenport, 2001). Internal transaction costs include those associated with searching, storing, distributing, and applying knowledge.

The three perspectives discussed above provide a starting point for analyzing the benefits of KM programs. Attempts have been made to unify these perspectives to create an integrated framework. Shin (2004) has studied how the factors in each perspective affect each other and revealed how the implementation of KM program can lead to both positive and negative effects. In current research, the factors that influence KM are divided into two categories, namely enablers and processes (C. W. Holsapple & Joshi, 2000; Hooff, Vijvers, & Ridder, 2003; Lee & Choi, 2003).

KM enablers are defined as "organizational mechanisms for fostering knowledge consistently" (Ichijo, Krogh, & Nonaka, 1998). The implementation of KM programs is significantly affected by "soft" factors such as company culture and organizational structure, which explains why similar KM approaches and practices can have very different impacts in different organizations (Hansen, Nohria, & Tierney, 1999; Malhotra, 2002). Openness, respect, autonomy, communication climate, clarity, commitment, simulation, feedback, and time pressure are taken as preconditions for the success of KM programs (Hooff, Vijvers, & Ridder, 2003). Companies should adopt the practices that are consistent with their business strategy in order to succeed (Hansen, Nohria, & Tierney, 1999; Malhotra, 2002).

As shown in Figure 1, this enabler-process-performance framework forms a comprehensive value chain for KM program value creation. Resource-based theory helps identify the key KM enablers that provide the foundation for managing knowledge. Process-based theory helps locate the key KM processes.

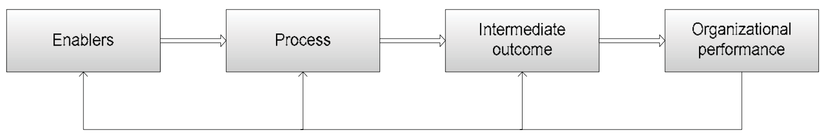


Figure 1. The enabler-process-performance framework.

4. From a Knowledge Sharing Perspective

Behind the reviewed approaches lies the attempt for translating the value creation of KM program into the operational performance improvement, thus placing KM in a supportive role. Empirical studies of various KM initiatives tend to describe the value creation in either monetary form, based on rough estimation of the value of each knowledge manipulation process, or anecdote story of how business has benefited by certain knowledge sharing activity facilitated by management practice.

Yet, the link between operational performance and knowledge manipulation activities remains vague. KM accounts for only one successful factor for operational improvement, while other factors, such as customer relationship management, marketing, quality control, and human resource management, also play important roles. The contribution by KM to the business bottom-line is difficult to track.

The contribution of KM can go far beyond knowledge reusing. More and more people have realized that organizational culture towards knowledge sharing, also known as *information behavior*, embodies an important part of organizational asset (Lev, 2001; Marchand, Kettinger,

& Rollins, 2002). Because proactive information behavior is socially complex and hard to imitate, it is seen as the origin of sustainable competitive edge.

To better define this, we view the value creation from the knowledge sharing perspective, taking it as the primary objective of KM programs. Therefore, each dimension of the framework, as shown in Figure 2 directly contributes to the organizational capability of knowledge sharing. The relationships between each dimension are also discussed to show how they affect each other in knowledge sharing. We will also demonstrate how the value is created and how it is linked to operational performance.

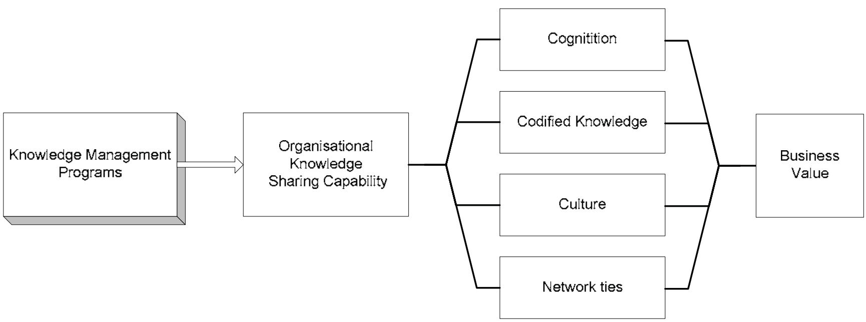


Figure 2. An integrated framework for measuring KM program value creation.

5. Cognitive Dimension

It is well recognized that shared context, also known as "redundant knowledge", plays an important role in knowledge sharing and provides a common ground necessary for any communication (Biosot, 1998; Nonaka & Takeuchi, 1995). Nohapiet and Ghoshal (1998) suggested two main constructs for the cognitive dimension, namely shared language and shared narratives.

5.1. *Shared Language*

A basic infrastructure for knowledge sharing is created by shared language. In the organizational context, however, shared language often goes beyond the linguistic dimension, composed of acronyms, jargons, etc. Almost every business unit has its own set of such codes, thus defining and making sense of their work. Arguably, the taxonomy shared within the organization can also be identified as a form of shared language in the form of taxonomy in certain knowledge repository, directory of skill sets, or even how the functional departments are divided. The shared taxonomy provides the structure under which knowledge is organized and stored. It also provides guidance for the seeking of knowledge.

On the other hand, the shared language dimension not only provides the common context for communication, it also places obstacles for cross-boundary knowledge sharing. It is observed that dealing with this issue can be a task both important and tricky for promoting knowledge sharing, especially in multi-national organizations (Davenport & Prusak, 1998).

5.2. *Shared Narratives*

Shared narratives are built upon shared language and codes. More importantly, they embody the norms, values, attitudes, and shared vision in a rich set of meanings. In the organizational context, it is observed that shared narratives include organizational myths, hero stories, failure stories, war stories, and projection of the future (Cohen & Prusak, 2001).

The cognitive common ground obviously facilitates the combination of knowledge. More importantly, the shared narrative helps keep a consistent understanding of the organization and influence people's identity and behavior. Many of the tacit features of an organization, such as working culture, norms and attitudes, are hard to be documented and communicated, but are embodied in shared narrative. Cohen and Prusak (2001) have demonstrated how an organizational myth represents the organizational culture.

Various management practices have been carried out to build up or enhance the common ground for cognitive dimension and may not have been identified under the name of "knowledge management". To overcome the obstacle placed by different manners of language between workers in North Sea oilrigs and the executives in London, BP had to employ consultants for translation. (Davenport & Prusak, 1998).

As for KM programs, the cognitive common ground acts not only as the foundation for knowledge sharing, but also is reshaped and enhanced continuously. Vocabularies and codes that are used among one party are likely to be exposed to and learned by others during knowledge sharing. An example is the development of taxonomies within the shared knowledge repositories. In order to maintain a concise and well-categorized structure of the repository, adequate understanding of the classification system is necessary. Thus, the established mechanism represents the common understanding of various tasks that are carried out by participants of the repository.

6. Codified Knowledge Dimension

The dichotomy of tacit and explicit knowledge, which is first proposed by Polanyi (1962), is widely cited. *Tacit knowledge* refers to the mental models and experiences that either is difficult or cannot be documented, while *explicit knowledge* refers to the formal models, rules and procedures that can be recorded. Although tacit knowledge is receiving more attention in management research, due to the fact that it is often harder to be captured and shared, explicit knowledge enjoys the intrinsic virtue of being recorded and indexed (Jordan & Jones, 1997). The fast-development of IT has demonstrated enormous power in sharing, retrieving and publishing of codified knowledge. The ultimate objective for codifying and archiving knowledge is to achieve an accumulative and collective organizational memory which could be used to support present activities (Ackerman, 1998; Stein & Zwass, 1995). As for the evaluating of how well codified knowledge is managed in the organization, we further divide codified knowledge into formal knowledge and know-how.

6.1. *Formal Knowledge*

Formal knowledge refers to the formal documents eg. books, reports, and memos, that are produced and used in daily work. For some knowledge-intense work, such as consulting and law industry, these documents are the main products of work. The typical implementation that facilitates the sharing of formal knowledge is the knowledge repository where such document are stored and indexed for future retrieval.

The main obstacle is not the storing, but rather the retrieving of such knowledge. Knowledge repositories of this type tend to be very large, and a keyword based search may yield thousands of results. People often lack the will and energy to sift through piles of documents to find relevant knowledge. It is not surprising to see that many knowledge repositories are left largely unused.

6.2. *Know-how*

Know-how represents the other dimension of explicit knowledge. While formal knowledge documents the outcome of work, know-how concerns the ability to assimilate and use knowledge to get work accomplished. This can be embodied as tricks of operating certain machinery, tips for using particular software tools or recommendations of information resource for certain topics. Know-how is deeply rooted in people's experiences.

When captured and stored properly, know-how can have immense power on organizational operation. Answer Garden exemplifies such a success story (Ackerman, 1998). Software users post problems encountered in their daily work, which are subsequently directed to experts. Experts can upload the problem and answers to the system database if they consider the problem as common. The database grows "organically" as new questions arise. Users can browse the information either by subject or through key word search.

7. Culture Dimension

Recently, the emphasis of KM has shifted from system solutions to people, where organizational culture plays a vital role. Empirical studies have shown that organizational culture, which acts as collective mindset, varies with different societies, industries and departments (Chatman &

Jehn, 1994; Sparrow, 1998). To date there has been no consensus of factors that compose organizational culture in the KM context.

We believe the direct impact of organizational culture on KM is the willingness for knowledge sharing activities. This includes two dimensions, namely knowledge contributing and knowledge transferring. While we acknowledge there are various other factors influencing organizational culture, we argue that it is through the impact on these measures that organizational culture influences knowledge sharing. Thus, the two measures are qualified for indicators of organizational culture in the context of KM.

8. Knowledge Contributing

Knowledge sharing involves the contributing of knowledge from the knower as well as the searching of knowledge from the seeker. From the knower's perspective, the willingness to share determines how much he/she would contribute when noticed a knowledge request. Knowledge contributing takes time and energy, moreover, in a highly competitive working environment where knowledge is seen as power, it is rational for people to be reluctant to share knowledge--defined as knowledge hoarding. Conversely, the sharing of knowledge may build up the knower's reputation, identifying him/her as knowledgeable and a team player, which in turn is helpful for his/her career. In a real working context, however, people rarely engage in such a carefully executed loss-gain analysis before making decisions of whether or not actively to share their knowledge. It is organizational culture that governs people's decision making, as illustrated in Table 1.

9. Knowledge Transferring

The challenges for knowledge transferring are not as obvious as for knowledge contributing. It seems unquestionable that people would be active in seeking and applying knowledge. However, empirical studies have suggested in the real world this is not always the case. Davenport and Prusak (1998) described a case at Mobil Oil in which considerable

Table 1. Implication of organizational culture on knowledge sharing.

Factors	Implication on knowledge sharing
Obligation	Willingness to share knowledge as returning of favor.
Openness	Reduce the "psychological cost" for seeking for knowledge. Failure stories are seen as experience gained rather than evidence of incompetence.
Communication Climate	Encouraging the seeking for knowledge.
Tolerance for mistakes	Reduce the "psychological cost" for both sharing knowledge and seeking for knowledge. Experience from failures can be open for discussion.

effort had to be taken to make widely-adopt a new technology--one that the value was immediate and indisputable.

When seeking knowledge, the resources people choose may not be the best in terms of efficiency and accuracy. It has been observed that engineers prefer resources with the lowest "psychological cost", worrying that seeking information may suggest their incompetence (Ackerman, 1998). Furthermore, the reluctance in knowledge transfer is influenced by people's intrinsic resistant to change. The adoption of knowledge often leads to some form of change, which has intrinsic risk. Thus, the trust of the knowledge resource and the organizational tolerance for mistakes substantially influence people's decisions for transferring knowledge.

The organizational culture not only has significant influence on knowledge sharing, as we have discussed, but also is relevant to other important factors of organizational performance. Cohen and Prusak (2001) have shown how employee satisfaction suffers from a lack of trust in the organization. Consequently, organizational culture presents as an important form of organizational assets, known as social capital. Up to now, organizational culture is generally recognized as “preconditions” for knowledge sharing. Research has focused on how KM program should be designed to acclimatize itself to the organizational culture. We argue that organizational culture is also constantly changed and reshaped over time. Particularly, the implementation of KM program contributes to the transformation. For example, it has been observed that trust, which is highly valued by managers, can be built in virtual teams through repeated rounds of interaction (Buchel & Raub, 2002). Larger and more long-term returns can be expected from the changing of user behavior, which in turn builds the inimitable and sustainable competitive edge.

10. Network Ties Dimension

As knowledge is intrinsically tacit, sticky, local, and contextual, face-to-face conversation and other types of direct interaction between people remain the best ways in many occasions for sharing knowledge (Davenport & Prusak, 1998). Consequently, the social network ties that one holds determine the knowledge resource to which he may have access. It is increasingly recognized that the collective social network ties established within the organization embody an important form of organizational resource.

Various management practices can be used to create and enhance network ties. Face-to-face meetings and presentations provide a venue in which employees meet each other. IT is also used to ease the process for locating and contacting people, and fostering community of practice. For example, people-finder systems are used to locate experts who hold certain knowledge (Becerra-Fernandez, 2004). Online discussion forums help create communities of practices, eliminating the obstacles placed by geographical dispersal and limited resource (Lesser & Storck, 2001; Wenger & Snyder, 2000).

11. Compare with Conventional Framework

From the discussion above we have demonstrated how the knowledge sharing capability can be seen as the vehicle through which the business value is achieved. On one hand, the organizational knowledge sharing capability, which we define as being composed of four dimensions, are promoted by KM programs. On the other hand, this capability also has a direct link with achieving tangible business value.

The implications of different perspectives towards the value proposition are shown in Figure 3. For both the process-based and resource-based perspective, the link to business performance is relatively static. The channels through which KM programs are seen to bring value are constrained to pre-defined knowledge manipulation processes or organizational resources. In contrast, the market-based view and the knowledge sharing perspective provide an increased dynamic link to business performance. This dynamic nature endows the approach with the capability of accommodating various management practices and strategy.

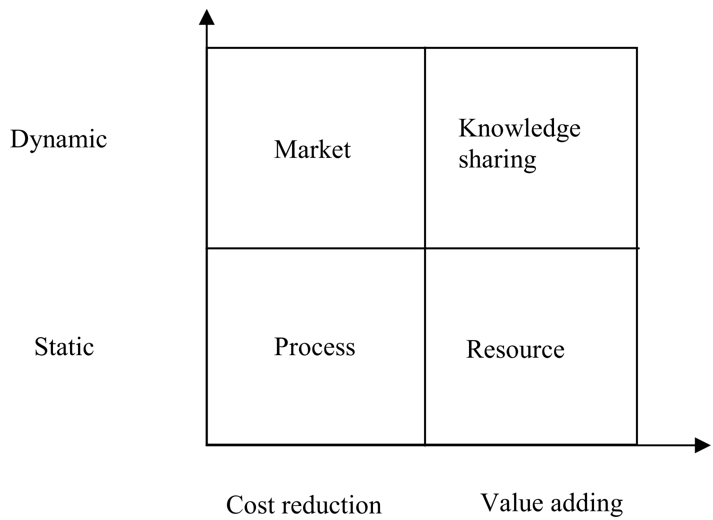


Figure 3. Perspectives towards KM.

The market-based and process-based perspectives view value creation from cost reduction and is often seen as the improvement of internal business process. Whereas, the resource-based and knowledge sharing perspectives are more concerned with the added value from KM programs, which then broadens the scope of measurement.

12. Conclusions and Discussions

KM research is exploring ways for demonstrating and tracking the value creation of KM program. In this paper we have discussed the limitations of conventional perspectives towards KM, namely the resource-based perspective, process-based perspective and market-based perspective, and corresponding value-added proposition. A new framework based on the organizational knowledge sharing perspective is then proposed.

Our main argument is that operational performance is not the appropriate measure for the value creation of KM programs. Alternatively, we take the knowledge sharing capability as the main value proposition. Firstly, from this perspective, a new framework is proposed to demonstrate the value-creation. Using empirical research data, it is demonstrated how organizational knowledge sharing capability is promoted through KM programs, as well as how it contributes to creating tangible business value. Secondly, the inter-relations between the four dimensions of value proposition are also discussed.

Because the discussion is mainly based on previous theoretical and empirical research, further validation of the framework might be needed through implementation. We are developing a methodology based on the framework to provide a practical and procedure-based approach for measuring the business value-creation of KM program. By implementing the methodology in the organizational environment, the link between the framework to business value, especially for the Cognition Dimension and the Culture Dimension, would be further clarified. Moreover, the reflection of the methodology would give a more valuable insight on the subject.

References

1. Ackerman, M. S. (1998). Augmenting organizational memory: a field study of answer garden. *ACM Transactions on Office Information Systems*, 16(3), 203-224.
2. Alavi, M., & Leidner, D. E. (2001). Review: Knowledge management and knolwedge management systems: conceptual foundations and research issues. *MIS Quarterl*y, 25(1), 107-136.
3. Arthur, W. B. (1999). New economics for a knowledge economy: The law of increasing returns. In R. Ruggles & D. Holtshouse (Eds.), *The Knowledge Advantage* (pp. 195-212): Capstone.

4. Becerra-Fernandez, I. (2004). The role of artificial intelligence technologies in the implementation of people-finder knowledge management systems. *Knowledge-Based Systems, 13*(5), 315-320.
5. Biosot, M. H. (1998). *Knowledge assets: Securing competitive advantage in the information economy*. Oxford: Oxford University Press.
6. Brown, J. S., & Duguid, P. (2000). *The social life of information*. Boston: Harvard Business School Press.
7. Buchel, B., & Raub, S. (2002). Building knowledge-creating value networks. *European Management Journa*l, 20(6), 587.
8. Chatman, J. A., & Jehn, K. A. (1994). Assessing the relationship between industry characteristics and organizational culture. *Academy of Management Journ*al, 37(3).
9. Cohen, D., & Prusak, L. (2001). *In good company: How social capital makes organizations work*. Boston: Harvard Business School Press.
10. Davenport, T. H., & Prusak, L. (1998). *Working knowledge: How organizations manage what they know*. Boston: Harvard Business School Press.
11. du Plessis, M. (2005). Drivers of knowledge management in the corporate environment. *International Journal of Information Management, 25*(3), 193-202.
12. Grover, V., & Davenport, T. H. (2001). General perspectives on knowledge management: Fostering a research agenda. *Journal of Management Information Systems, 18*(1), 5-21.
13. Gupta, O., Pike, S., & Roos, G. (2002). *Evaluating intellectual capital and measure knowledge management effectiveness.* Paper presented at the PMA Performance Measurement Conference, Boston.
14. Hansen, M. T., Nohria, N., & Tierney, T. (1999). What's your strategy for managing knowledge? *Harvard Business Review, 77*(2), 106-116.
15. Holsapple, C. W., & Joshi, K. D. (2000). An investigation of factors that influence the management of knowledge in organizations. *The Journal of Strategic Information Systems, 9*(2-3), 235.
16. Holsapple, C. W., & Joshi, K. D. (2002). Knowledge manipulation activities: Result of a Delphi study. *Information & Management, 39*(6), 477-490.
17. Hooff, B. v. d., Vijvers, J., & Ridder, J. d. (2003). Foundations and applications of a knowledge management scan. *European Management Journal, 21*(2), 237-246.
18. Ichijo, K., Krogh, G., & Nonaka, I. (1998). Knowledge enablers. In G. Krogh, J. Roos & D. Kleine (Eds.), *Knowing in Firms*. Sage Publications.
19. Jordan, J., & Jones, P. (1997). Assessing your company's knowledge management style. *Long Range Planning, 30*(3).
20. KPMG. (2000). *Knowledge management research report 2000*. KPMG Consulting.
21. Lee, H., & Choi, B. (2003). Knowledge management enablers, processes and organizational performance: an integrative view and empirical examination. *Journal of Management Information Systems, 20*(1), 179-228.
22. Lesser, E. L., & Storck, J. (2001). Communities of practice and organizational performance. *IBM Systems Journal, 40*(4), 831-841.
23. Lev, B. (2001). *Intangibles: Management, measurement, and reporting*. Brookings Institution Press.

24. Malhotra, Y. (2002). Why knowledge management systems fail? Enablers and constraints of knowledge management in human enterprises. *Handbook on Knowledge Management*.
25. Marchand, D. A., Kettinger, W. J., & Rollins, J. D. (2002). *Information orientation : the link to business performance*. Oxford: Oxford University Press.
26. Markus, M. L. (2001). Toward a theory of knowledge reuse: Types of knowledge reuse situations and factors in reuse success. *Journal of Managment Information Systems, 18*(1), 484-505.
27. Nahapiet, J., & Ghoshal, S. (1998). Social capital, intellectual capital, and the organizational advantage. *Academy of Management Review, 23*(2), 242-266.
28. Nonaka, & Takeuchi. (1995). *The knowledge creating company*. Oxford: Oxford University Press.
29. Orr, J. (1996). *Talking about machines: An ethnography of a modern job*. Ithaca, NY: IRL Press.
30. Petrash, G. (1996). Dow's journey to a knowledge value management culture. *European Management Journal, 14*(4), 365-373.
31. Polanyi, M. (1962). *Personal knowledge*. Chicago, IL: University of Chicago Press.
32. PricewaterhouseCoopers. (1999). KM business value: lessons learned from early adopters. *Global Enterprise Advisor, 10*, 7-8.
33. Shin, M. (2004). A framework for evaluating economics for knowledge management systems. *Information & Management, 42*, 179-196.
34. Sparrow, J. (1998). *Knowledge in organizations: Access to thinking and work*. Sage Publications.
35. Stein, E. W., & Zwass, V. (1995). Actualizing organizational memory with information systems. *Information Systems Research, 6*(2), 85-117.
36. Swan, J., Newell, S., & Robertson, M. (2000, //). *Limits of IT-driven knowledge management initiatives for interactive innovation processes: Towards a community-based approach.* Paper presented at the roceedings of the 33rd Hawaii International Conference on System Sciences. Maui, HI.
37. Wenger, E. C., & Snyder, W. M. (2000). Communities of practice: The organizational frontier. *Harvard Business Review*, 139-145.
38. Wilson, T. D. (2002). The nonsense of “knowledge management”. *Information Research, 8*(1).

Chapter 8

SOCIAL NETWORK ANALYSIS OF FOUR DEPARTMENTS IN THE NATIONAL UNIVERSITY OF SINGAPORE

Chu Keong Lee
Wee Kim Wee School of Communication and Information
Nanyang Technological University
ascklee@ntu.edu.sg

Jee Foon Wee
Nanyang Technological University
W040013@ntu.edu.sg

Social network analysis was performed by examining the co-authorship patterns of authors from four departments at the National University of Singapore whose work was published in 2004. A tie between two authors was established when they co-authored a paper. From the results, inferences were made regarding the social cohesiveness and the level of social capital in networks. It is argued that social network analysis can be used for three purposes: (1) to identify research networks and the inner and outer circle of an author's professional acquaintances; (2) to map the structural dimension of the social capital; and (3) to identify emergent communities of practice. Lastly, avenues for further research are proposed.

1. Introduction

There remains a tremendous interest in the analysis of the publication output from an academic department and in identifying the most prolific authors. Also of interest is the study of the extent of collaboration among authors by using co-authorship of published papers as a basis for review, as well as noting the most cited articles. These inquiries have traditionally been performed using the techniques of bibliometrics, citation analyses, and scientometrics. While these provide a good indication of the amount of material being published, in which journal the articles are being published, and the impact of the publications on the scientific community, these techniques tell us nothing about how the authors are linked in a network. To be sure, the collaborative index (Lawani, 1980), degree of collaboration (Subramanyam, 1983), and collaborative coefficient (Ajiferuke, Burell & Tague, 1988) gives a numerical indication of the extent of collaboration, however, a graph is often required to demonstrate the network of collaborations. This situation is not unlike using Zipf's Law to understand the statistical regularities of a language. Zipf's Law alone does not provide deep enough insight into the organization of a particular language, rather it must be supplemented by a description of the co-occurrence of words using a graph to achieve a deeper understanding (Cancho & Solé, 2001).

Social network analysis (SNA) is a set of techniques developed for analyzing organizations that focus on the relationships between people, the patterns of those relationships, and its implications. SNA borrows a great number of ideas and concepts from graph theory. Dekker (2001) listed four main purpose of conducting SNA. Firstly, SNA can be used to visualize the relationships between people or between groups by using a diagram called a sociogram, which is a special form of a graph. With this, it is widely recognized that while organizational charts show who works where and who reports to whom, this seldom reflects how work gets done. To reflect how work, which is frequently unseen, is achieved in the organization there is a need to highlight whom works with whom and who knows what, and this can be done using sociograms. Secondly, SNA allows an investigation into the factors that influence the relationships, which may be attributed to any one of a plethora of reasons (e.g., homophily). Thirdly, SNA allows the researcher to identify critical nodes, e.g., the bottlenecks, isolates, and cosmopolites. Finally, recommendations to improve the communication within the network can be made on the basis of the computed parameters such as centrality.

2. Literature Review

The techniques of SNA are derived from graph theory, and they have been used to study the characteristics of diverse types of networks. Five applications will be reviewed here, namely, analyses of covert networks,

networks at a military headquarters, the transmission of information, networks of interlocking directorates, and the structure of newsgroups.

Krebs (2002) mapped the terrorist network that was involved in the September 11, 2001, attacks in order to gain an insight on the structure of covert networks. To do this, he used publicly available information from Web sites of major newspapers, principally, those of *The New York Times, The Washington Post, The Wall Street Journal, The Los Angeles Times,* and *The Sydney Morning Herald.* From these Web sites, he collected information about the nodes (i.e., the terrorists, comprising 19 hijackers and their accomplices) and the links connecting them (i.e., the relationships between the terrorists). Krebs believed that as trust was a critical element in the case of covert networks, it was important to reflect the level of trust between the actors in such networks. He achieved this by weighting the links according to the strength of the tie. The terrorists who lived or attended classes together were assumed to have the strongest ties. Those traveling or participating in meetings together were assumed to have the ties of medium strength. Those who had only a single transaction were assigned the weakest ties. Krebs analyzed two networks. The first comprised of only the 19 hijackers, and the second included both hijackers and their accomplices. The network comprised of only the hijackers was sparse. However, where there were links, they were strong, comprising mostly of those that formed between old classmates, kin, or people that had lived together over a number of years. The links in this network had a high average path length of 4.75. On the other hand, the network that included the accomplices was comprised of mainly weak ties that were transitorily formed to coordinate tasks and to report progress. These ties were dormant after the tasks had been accomplished. The weak ties provided by the accomplices reduced the average path length of the network to 2.79. Not surprising, Krebs identified missing data as the biggest problem in the analysis of covert networks and stressed that to obtain a complete picture of covert networks, four networks needed to be mapped: (1) trust; (2) task; (3) money and resources; and (4) strategy and goals. Lastly, Krebs stated that the SNA on terrorist networks is more suitable for prosecution rather than prevention.

Dekker (2001) applied SNA to a military headquarters to achieve two ends. Firstly, to make recommendations about improving work practices, and secondly, to suggest new physical placement of staff within a building to improve information flow. He, too, stressed the importance of using the numerical strength of a score to indicate the strength of the relationship. Dekker used the following values, based on the frequency of interaction, as a proxy for the strength of relationships:

1.0	=	three or more times a day
0.8	=	once per day
0.6	=	three or more times per week
0.4	=	once per week
0.2	=	once per fortnight
0.0	=	less than once per fortnight

To perform the SNA analysis, CAVALIER (ChAnge VisuALIsation for the EnteRprise), a Java-based suite of tools, was developed. It is comprised of four packages, namely, electronic questionnaire generation, data extraction, network editing and visualization, and network analysis.

SNA techniques have also been used to investigate the spread of information. An example of this is Lee's (1969) investigation into the transmission of information about illegal activities. Specifically, she investigated how women located an abortionist who was willing to perform an illegal abortion, running afoul of the law in the process. As this is an illegal activity, such information is not available through conventional means (e.g. through the Yellow Pages). Most seeking illegal abortions ask for leads from their most trusted friends whom they think had some experience with an abortionist in the past. Lee distinguished between the three types of information about abortions: (1) information about abortion as an alternative to unwanted pregnancies and factual information about abortions; (2) gossip about the abortion experience of specific individuals; and (3) information about how to reach an abortionist willing to perform illegal abortions. Concentrating on (3), Lee plotted out the search process diagram, a special form of the sociogram, to visually depict the acquaintance network. Granovetter (1973) studied the relationship between people who changed their jobs (job changes) and people who provided the information that facilitated the job change. He found that 55.6% of the job changers obtained their information from weak ties, i.e., those with whom they had an occasional contact (more than once a year but less than twice a week). Ties are essentially channels through which ideas, influence, and information flows, and the reason why the majority of the job changers obtained the information from people whom they met occasionally is because these people moved in circles different from that of the job changer and will thus be exposed to information different from that which the job changer received from his strong ties. He concluded that from a microscopic perspective, weak ties are crucial in making possible mobility opportunity, and from a macroscopic one, weak ties promote social cohesion.

Alexander (2003) investigated the phenomenon of multiple directorships--which led to corporate interlocks--in the 250 largest companies in Australia in the years 1976 and 1996. The companies comprised the top 200 non-financial companies (as determined by revenue) and the 50 largest financial companies (as determined by assets). Constructing two networks--the inter-corporate (where the nodes are companies) and the interpersonal (where the nodes are company directors)--he noted an increase in the incidence of multiple directorship holding. Although the number of board seats increased from 1,896 to 1,932, the number of directors decreased from 1,586 to 1,508. Therefore, in 1996, fewer people held a greater number of board seats than in 1976. This means that there is an increased concentration of power holding in the population of the top 250 companies in 1996. The number of companies in the main component of the inter-corporate network increased from 155 (62.2%) to 198 (79.2%) and involved more directors (from 992 to 1,163). However, the number of seats an interlocked company director held remained at 2.7. This meant that while the phenomenon has *spread*, there is no *intensification* of it. It is on the basis of this type of research that Web sites like theyrule.net are constructed.

Muncer, Loader, Burrows, Pleace, and Nettleton (2000) used SNA to analyze the structure of two newsgroups giving social support. They found that the virtual social support for a newsgroup dealing with issues related to depression was very different from those dealing with diabetes.

The diabetes group was more concerned with information about the treatment of diabetes while the depression group was moreso with esteem and social companionship. These are needs that are met in different ways. Information needs are more episodic, as once the information is found there is no need for the services of the newsgroup until another information need arises. This is fundamentally different from a need for esteem and social companionship, which are more ongoing and long term in nature and also tend to result in the formation of friendship cliques. The question remains: what is the impact on the form of the newsgroup and structure of the interactions given the different goals? They found that the depression newsgroup had a denser, more vibrant, and more socially cohesive network because of its emphasis on social companionship. The diabetes newsgroup was found to be more diffuse, as it is information driven, and participation rests on whether or not one has information to contribute. Importantly, they highlighted the three difficulties of analyzing the network structure of newsgroups: (1) the directionality of newsgroup postings, unlike email, is unclear; (2) the network does not take into account the other forms of contact (e.g. email) which may be significant precisely because of the support function of the newsgroups; and (3) "negative ties" like those caused by disagreements and arguments are not reflected in the analysis. Using degree centrality, they identified the moderators (hub), liaisons (cutpoints), and cliques in the newsgroups.

Newman (2001) constructed scientific co-authorship networks using bibliometric data from four databases of scientific papers, namely, the Physics E-print Archive, Medline, SPIRES, and NCSTRL. Data from a five-year window period (January 1, 1995 to December 31, 1999) were used to construct large sociograms with the number of nodes ranging from 11,994 (NCSTRL) to 1,520,251 (Medline). The size of the giant component, in percentage terms, ranged from 80 to 90% of the nodes in the network. Newman found the average geodesic distance between pairs of authors to be small (4.0 to 9.7), that the networks displayed "small world" characteristics (high clustering coefficient and small path length), and that the graphs display a significant funneling effect (where a few ties are responsible for the majority of the geodesic paths passing through a node). Lastly, Newman suggested estimating the strength of collaborative ties by accounting for two intuitive:

- for a paper with many co-authors, the co-authors are on the average likely to know each other less well than the co-authors of a paper with few authors, e.g., the authors of a ten-authored paper know each other less well than two authors of a co-authored paper, and the strength of ties between the ten authors should be correspondingly reduced
- authors who have written many papers together will know one another better than those who have written few papers together, and the strength of ties between frequent co-authors should be correspondingly increased

In this paper, we construct the co-authorship networks of authors of four entities (schools or departments) in the National University of Singapore.

3. Background

The National University of Singapore (NUS) is one of three universities in Singapore, and one of the finest universities in the world. In 2004, NUS was rated 18th in The Times Higher Education Supplement global ranking of universities. It was one of only three Asian universities to be rated among the top 20 that year[1]. NUS is a comprehensive university which has 13 faculties with an enrolment of more than 23,000 undergraduate and 8,000 graduate students. It prides itself in being ranked tenth in the world in terms of the publication output of its engineering faculty (National University of Singapore, 2005).

4. Methodology

In SNA, the actors and ties can be defined in different ways depending on the research objective. In this paper, the actors are authors of scientific papers, and two authors are connected by a tie if they have

[1] The other two were the University of Tokyo and Peking University.

co-authored one or more papers together. The resulting network developed from such data is called an *affiliation network*, a network based on common membership. Such networks are typically constructed from very reliable and verifiable data (as co-authorship of a scientific paper can be determined to a high degree of certainty), tends to be large as it is unnecessary to conduct time-consuming interviews and surveys, and is free from the subjective nature of concepts like friendship and trust (Newman, 2001).

The data for the analysis were obtained from the Web of Science®. Two fields were used to retrieve the bibliographic records of the publications. Firstly, the "address" field was used to narrow the search to the specific school or department. For example, publications from the NUS's Department of Mathematics were retrieved by using the search phrase "`natl univ singapore SAME math`". The "`SAME`" operator (Borgatti and Molina 2003) highlighted the fact that missing data can be very damaging to SNA studies, and Krebs (2002) stated that centrality measures are especially sensitive to minor changes in network connectivity. In this study, care was taken to ensure that the data was complete by including all the variant names that represented the same department in the search. For example, the publications from NUS's Business School were retrieved by using the statement:

```
(natl univ singapore SAME sch business) OR
(natl univ singapore SAME dept decis sci) OR
(natl univ singapore SAME nus business school)
OR (natl univ singapore SAME dept business
policy) OR (natl univ singapore SAME dept
finance) OR (natl univ singapore SAME org) OR
(natl univ singapore SAME mkt)
```

Secondly, the date field was used to further limit the results to the year 2004. The data was then cleaned, and a consistent format used to represent author names, using the underscore in place of the space. For example, "Ranganathan, C" became Ranganathan_C. After cleaning, they were converted into the DL file format to enable it to be imported into UCINET for analysis. Excel was used to generate the "papers per author" and "authors per paper" figures. The sociograms were generated using NetDraw and Pajek, and KeyPlayer was used to identify the key scientists whose removal would cripple the network.

5. Results and Discussion

The results of the analyses are summarized in Table 1. The networks vary from 76 to 274 nodes in size, making density comparisons across the four networks difficult. The average number of nodes per component varies from 3.18 in the Business School to 20.9 in the Chemical Engineering Department. This indicates that the business school network is fragmentary and consists of many small components while the chemical engineering network is made up of networks of significantly larger size. The fragmentary nature of the business school network is further illustrated by the high percentage of isolates (7.9%) it contains. Isolates indicate that a particular author has not co-authored in 2004. In the chemical engineering department, the percentage of isolates is a low 0.7%.

The degree centrality is an indication of the network activity of a particular actor. The maximum degree centrality is unsurprisingly in the chemical engineering sociogram, and the minimum in the business sociogram. Buchanan (2002) divided networks into two different types, egalitarian and aristocratic: egalitarian networks, namely, those whose nodes that have a fairly even distribution of links, and aristocratic networks, namely, those that have a few nodes that are very densely connected and many nodes that are sparsely connected. The population standard deviation of the degree centrality can be used to identify these network types. The chemical engineering and dentistry networks are clearly aristocratic sociograms, having a population standard deviation of 9.6 and 7.9 respectively. The business and mathematics sociograms are egalitarian, as they have a low standard deviation of 1.1 and 1.9 respectively.

The chemical engineering and dentistry networks are also characterized by a high number of authors per paper (3.54 and 3.75, respectively). This means that there is much collaborative research work being performed in these departments. The number of authors per paper

Table 1. Summary of results for the four departments.

Department	Business	Chemical Engineering	Mathematics	Dentistry
No. of papers	41	164	83	59
No. of authors	76	274	137	120
Papers per author	0.54	0.60	0.61	0.49
Authors per paper	2.24	3.54	1.94	3.75
No. of components	22	13	27	9
No. of isolates	6 (7.9%)	2 (0.7%)	5 (3.6%)	0 (0.0%)
Ave no. of nodes in a component	3.18	20.9	4.89	13.3
Degree centrality	0–6 (s.d. 1.1)	0–97 (s.d. 9.6)	1–10 (s.d. 1.9)	1–72 (s.d. 7.9)
Main (Giant) Component				
No. of nodes (inclusivity)	7 (9.2%)	189 (69.0%)	19 (13.9%)	79 (65.8%)
Density	47.62%	3.98%	16.4%	9.09%
Average Geodesic Distance	1.62	4.87	2.71	2.46
Maximum Geodesic Distance	3	11	5	4
Clustering Coefficient	0.882	2.093	0.969	1.583

is low for both the business and mathematics networks (2.24 and 1.94, respectively). In the chemical engineering network, KeyPlayer identified Kawi_S., Liu_Y., and Zhang_Y. as being the scientists whose removal will fragment the network.

Figures 1 and 2 are the sociograms of the business and chemical engineering networks respectively. The fragmentary nature of the business network, and the aristocratic nature of the chemical engineering network are immediately apparent. What other things can such sociograms tell us? Clique analysis can help to identify research networks, which have been defined by Mulkay (1977) as amorphous social groupings of researchers that are in a state of constant flux. In his research, Becher (1989) found that for many disciplines, scientists have their own inner and outer circles of professional acquaintances. The size of the inner circle depends on the gregariousness of the researcher, but

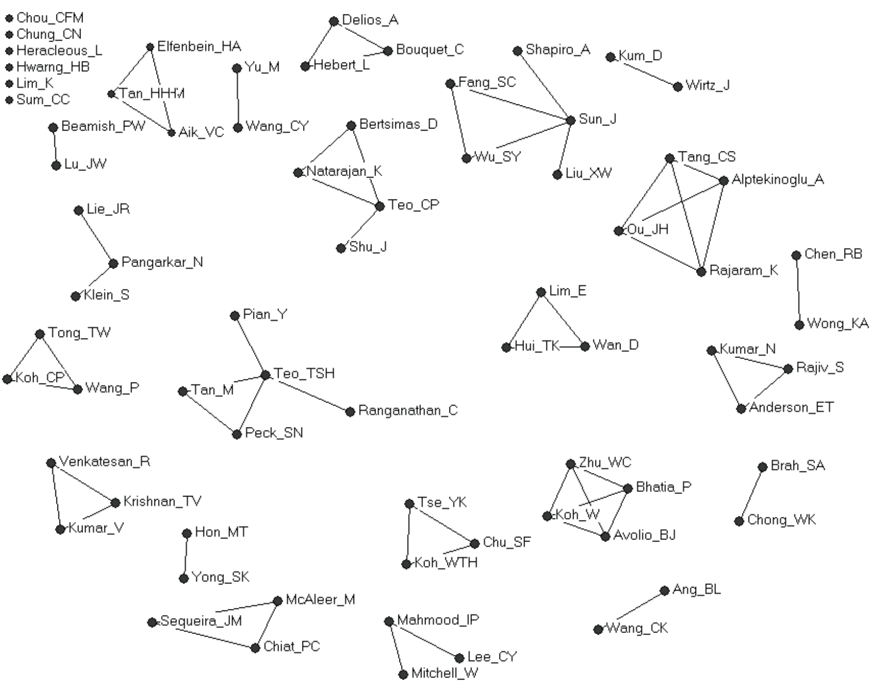


Figure 1. Sociogram of NUS business school (based on the publications in 2004).

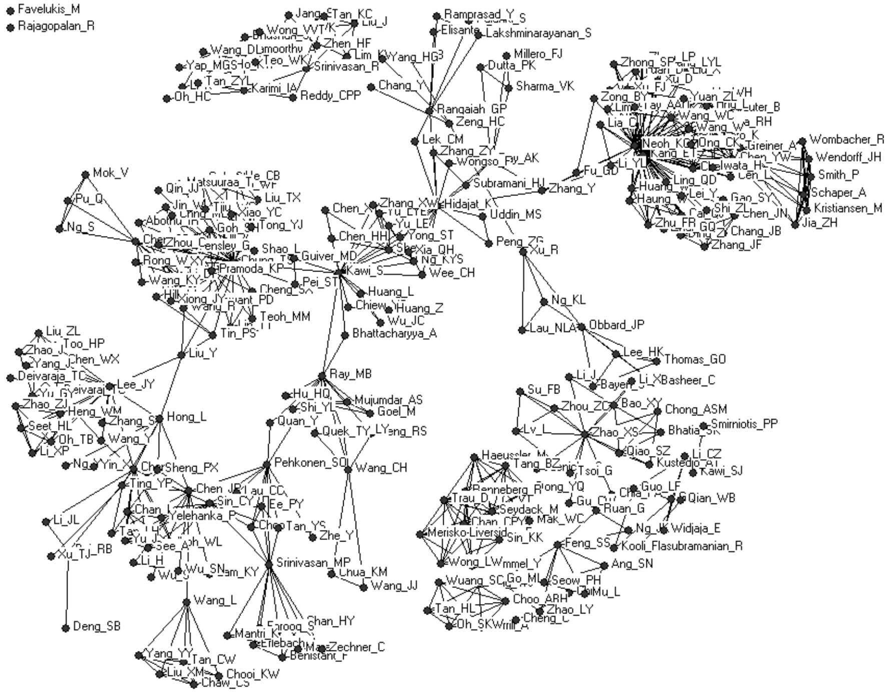


Figure 2. Sociogram of NUS chemical engineering department (based on the publications in 2004).

tends to be small, with six to 20 members. On the other hand, the outer circle can number between 100 to 400, depending on the specialism. (Scientists working in hard knowledge fields are known to have large outer circles while scientists working in soft knowledge fields have smaller ones.) A use can be found for the sociogram here as well. A sociogram can be used, at the level of the individual researcher, to identify his inner and outer circles of professional acquaintance. We would like to propose that distance be used to identify the inner and outer circles of scientists. Nodes at a distance of 1 from the scientist in question, e.g., Neoh_KG in the NUS's Chemical Engineering Department, can be considered her inner circle, while those located at a distance of 2 from her can be considered her outer circle. For organizations interested in starting communities of practice, clique analysis can also be used to identify emergent communities of practice.

In addition to this, measures of density and degree centrality can be used to indicate the level of social cohesion. Low values of density and degree centrality indicate a low level of social cohesion and vice versa. What binds the network together and enables its members to work effectively together is the level of social capital within the network. Social capital comprises three dimensions, namely, the structural, content, and relational dimensions. The structural dimension has to do with the access to other actors. This is precisely what the sociogram illustrates. Thus, the sociogram makes the structural dimension of social capital visible.

6. Implications for Management and Future Research

There are several implications for management. Previous studies (e.g. Harande, 2001; Pao, 1982) have established that highly productive or prolific authors are also highly collaborative. This does not bode well for the three schools that have isolates. Further investigation can be made to ascertain the reasons for isolation. Could this be due to the specialized nature of the research that they do? Could this be due to the lack of sparring partners in their department? In a major UNESCO study, "The International Comparative Study on the Organization and Performance of Research Units" (Stankiewicz, 1979), it was found that the optimum size for research groups was three to four, as this was when the output of research papers per scientist was at its maximum. Sparing partners are also necessary for "redundancy", one of the five conditions that facilitate "learning by intrusion" (Nonaka, 1994). Redundancy is important as it encourages frequent dialogue and communication, resulting in the creation of a "common cognitive ground" among academics (Nonaka, 1994). This facilitates the transfer of tacit knowledge, as over time, individuals acquire the ability to sense what each other are struggling to articulate. This is especially important in the concept creation process, where individuals use metaphors rooted in tacit knowledge. Redundancy enables individuals to invade each other's functional boundaries to offer advice or to provide new information from different perspectives, speeding up the knowledge creation process. The writing of papers is an instance of knowledge creation. Is the lack of redundancy the reason why these academics are isolates? The business school network is fragmentary, with only 3.18 nodes per component. Fragmentary networks are low in social cohesion. Could the reason be a high level of turnover in the NUS Business School?

Gladstein and Caldwell (1985) stressed that in exchange relationships, boundary management is a critical predictor of team performance. He developed a set of boundary roles that are critical in new product development. In a sense, the authorship of research papers involves many exchange relationships, and the authorship process can indeed be regarded as a type of new product development. Therefore, the set of boundary roles they proposed would also be useful in the context of authorships. Gladstein and Caldwell characterized boundary exchanges in two ways-- according to the direction of flow, whether into or out of the group, and according to whether the initiator was a group member or an outsider (Figure 3). The *scout* brings information or resources (I/R) needed by the group across the boundary. The *ambassador* carries I/R that the group chooses to transmit to others. The *sentry* acts as a filter, controlling the I/R that external agents want to transmit into the group. The *guard* monitors what I/R external agents request from the group, determines what the group will release, and responds to their requests appropriately. In future research, it will be useful to identify the roles that bridge nodes in the networks play. The critical roles that bridge nodes have to play to ensure successful co-authorship of research papers is an interesting problem to investigate.

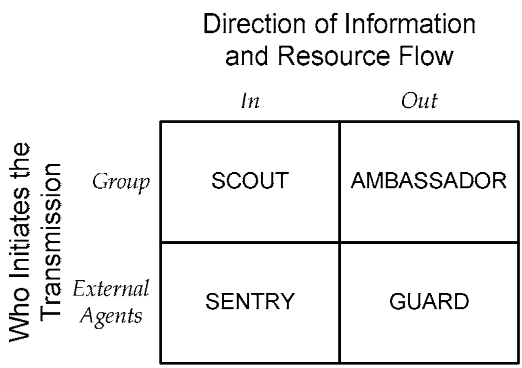


Figure 3. Boundary roles in groups.

7. Conclusions

In this paper, the research papers of four departments at the National University of Singapore were analyzed using social network analysis. Ties occur between the co-authors of a paper. From the results, some inferences were made regarding the social cohesiveness and the level of social capital in the networks. It was argued that social network analysis can be used for three purposes: (1) to identify research networks and inner and outer circle of an author's professional acquaintances; (2) to map the structural dimension of the social capital; and (3) to identify emergent communities of practice. Finally, avenues for further research were proposed.

References

1. Ajiferuke, I., Burell, Q. & Tague, J. (1988). Collaborative coefficient: a single measure of the degree of collaboration in research. *Scientometrics, 14*(5–6), 421–433.
2. Alexander M. (2003). Boardroom networks among Australian company directors, 1976 and 1996: The impact of investor capitalism. *Journal of Sociology, 39*(3), 231–251.
3. Becher, T. (1989). *Academic tribes and theirt: Intellectual enquiry and the cultures of disciplines*. Bristol, PA: SRHE and Open University Press.
4. Buchanan, M. (2002). *Nexus: Small worlds and the groundbreaking science of networks*. New York: W.W. Norton.
5. Borgatti, S.P., & Molina, J.L. (2003). Ethical and strategic issues in organizational social network analysis. *Journal of Applied Behavioural Science, 39*(3), 337–349.
6. Cancho, R.F.i., Solé, R.V. (2001). The small world of human languages. *Proceedings of The Royal Society of London. Series B (Biological Sciences), 268*(1482),2261−2265.
7. Dekker, A. (2001). *Visualising social networks using CAVALIER*. Retrieved March 25, 2006, from http://crpit.com/confpapers/CRPITV9Dekker.pdf.
8. Gladstein, D., & Caldwell, D. (1985). Boundary management in new product teams. *Academy of Management Proceedings, 45*, 161−165.
9. Granovetter, M.S. (1973). The strength of weak ties. *American Journal of Sociology, 78*(6), 1360–1380.
10. Harande, Y.I. (2001). Author productivity and collaboration: An investigation of the relationship using the literature of technology. *Libri, 51*, 124–127.
11. Krebs, V.E. (2002). Uncloaking Terrorist Networks. *First Monday, 7*(4). Retrieved March 25, 2006, from http://firstmonday.org/issues/issue7_4/krebs/index.html.
12. Lawani, S.M. (1980). *Quality, collaboration, and citations in cancer research: A bibliometric study*. Unpublished doctoral dissertation, Florida State University.
13. Muncer, S., Loader, B., Burrows, R., Pleace, N., & Nettleton, S. (2000). Form and structure of newsgroups giving support: A network approach. *Cyberpsychology and Bebavior, 3*(6), 1017−1029.

14. National University of Singapore. (2005). *Standing tall in worldwide rankings*. Retrieved March 25, 2006, from http://www.nus.edu.sg/annualreport/2005/14ranking.htm.
15. Newman, M.E.J. (2001). Who is the best connected scientist? A study of scientific co-authorship networks. *Physical Review, 64*(016131; 016132).
16. Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization Science, 5*(1), 14–37.
17. Pao, M.L. (1982). Collaboration in computational musicology. *Journal of the American Society for Information Science, 33*(1), 38–43.
18. Stankiewicz, R. (1979). The size and age of Swedish academic research groups and their scientific performance. In F.M. Andrews (Ed.), *Scientific Productivity: The Effectiveness of Research Groups in Six Countries* (pp. 191–222). Paris: Unesco.
19. Subramanyam, K. (1983). Bibliometric studies of research collaboration: A review. *Journal of Information Science, 6*(1), 33−38.

Chapter 9

MANAGING COLLABORATIVE NETWORKS

Bruce Cronin
Department of International Business & Economics
University of Greenwich Business School
Park Row, London SE10 9LS
+44(0)20-8331-9786
b.cronin@greenwich.ac.uk

Collaboration within and between organizations is usually examined in dyadic terms, between pairs of organizations or organizational subunits. But, it is argued here that there are analytical benefits from considering collaboration in the context of the network of relationships in which individuals and organizations are typically embedded. This paper extends the exploration-exploitation framework, conceiving this as the product of regions of strong and weak ties within complex networks. Drawing from a wide review of recent research, a number of insights arising from this network are presented, together with significant challenges for management. Some principles for effective management of collaborative networks are proposed.

1. Introduction

It seems these days that everybody is networking. People and organizations are increasingly discovering the networks they are part of and are seeking to leverage these. In today's world, there is value in connecting. Of course, there has always been value in connecting, as the infamous old boy's club demonstrates. The difference today is that more people have this knowledge and have access to powerful analytical tools to try to optimize their connections.

What are the implications of this network awareness for organizations seeking to enhance collaboration and innovation? Connecting people across boundaries and out of silos has been a key theme in knowledge management. But is connection always beneficial? What sorts of networks are optimal for collaborative and innovative practice? How can networks be optimized to this end?

This paper examines collaborative networks in terms of the differential benefits of strong and weak forms of collaboration, positing that the relationship between these is at the centre of collaborative networks. It is argued that the benefits of strong collaboration lie primarily in the exploitation of economies of scale and scope, while weaker forms of collaboration underpin diversity and exploration. But examining collaboration in network terms provides insights not apparent from an examination restricted to dyadic relationships and highlights some particular managerial challenges.

2. Escaping Silos

Collaboration is at the heart of organizational effectiveness and innovation. The coordination of related activities allows the division of labor and specialization that underpins productivity. This coordinated specialization of, often difficult to transfer tacit knowledge, allows economies of learning and economies of scale, underpinning much of the growth of large firms (Cohen, 1995). Yet the danger of over-concentration of scale has long been recognized. Large vertically integrated centralized bureaucracies have been slow to respond to changes in technology, costs and competition and to recognize the potential of radically new ideas. Major concentrations of activity are prone to turn into insular silos.

The silo is a metaphor for unreflective practice. People carry on with their day-to-day activities with little reflection, at best Argyris and Schon's (1978) "single loop learning", reviewing activities within their existing mind-set. Valuable learning comes from the second loop, critical review of the existing mind-set itself. Senge (1990) argues that a key condition to creating a learning organization is a systems perspective, looking beyond the immediate and seeing one's part in the whole.

But, as Stacey (2002) asks, is this anything more than a moral exhortation to be better people regardless of the social situation? Not only do people and groups react defensively to the questioning of existing mental modes, for fear of losing control (Argyris, 1990), but vested interests tend to emerge in and block the collegial situations necessary for effective critique (Stacey, 2002; Ferlie, Fitzgerald, Wood, and Hawkins, 2005).

Gupta and Govindarajan (2000) argue that the flow of knowledge in or out of an organizational group depends on the value of the knowledge, source motivation, recipient motivation, recipient absorptive capacity, and richness of transmission channels. Unique knowledge provides an organization with distinct capabilities and so is potentially valuable. But the value held at its source is generally evaluated by subsidiary rather than senior managers, and this often highly influenced by their personal networks rather than organizational goals *per se* (Arvidsson, 2000).

There are normally greater benefits to a subunit in hoarding valuable knowledge and brokering this in exchange for resources than sharing it. Locally-oriented resource and operational priorities normally crowd out knowledge transfer to the rest of the organization (Forsgren, 2000; Moore and Birkinshaw, 1998). Fear of appropriation and rivalries with other groups also limit knowledge sharing (Wilmott, 2000; Alvesson and Karreman, 2001; Kamoche and Mueller, 1998).

Recipients of externally-generated information are often weakly motivated to make use of it. Such information is often undervalued due to causal ambiguity, perceptions of the reliability of the source, rivalries with other groups about relative contributions and egoistic biases, generating a "not-invented-here" bias (Szulanski, 1996).

Perhaps the key factor influence the transfer of knowledge between organizations is the recipient's absorptive capacity (Cohen and Levinthal, 1990). This is the ability to recognize the value of knowledge, to assimilate it and apply it. This may be limited by incompatible practices, particularly where multiple parties and technical differences are involved (Cohen, 1995; Kogut and Zander, 1992). Absorptive capacity may also be limited by "cognitive distance" (Nooteboom, 2004) encompassing spatial, cultural and institutional asymmetries (Delios and Beamish, 1999; Gupta and Govindarajan, 2000; Kogut and Singh, 1988).

Absorptive capacity is in turn influenced by the richness of transmission channels, classically the distinction between narrow channels provided by information technologies and broad channels provided by face-to-face interaction. Spatial distance, for example, can be mitigated by information technologies where knowledge is separable, that is, explicit and codifiable, but this is not the case for tacit knowledge, which includes most of the interpretation and application of codified knowledge (Delios and Beamish, 1999; Wenger, 1998).

Organized systematic coordination across boundaries is widely recognized as a precondition for transcending the preoccupations of the immediate. Hence the prominence of boundary spanning activities as conferences, job rotation, cross-boundary teams, taskforces an steering groups, knowledge sharing practices, and standardized processes across organizational units (Evans, Pucik, and Barsoux, 2002). With growing awareness of the existence and impact of networks, these traditional boundary spanning approaches are being supplemented by recruitment of well connected personnel and active networking, the purposeful establishment of connections.

We are immediately confronted with a dilemma, however. The cultivation of these boundary spanning connections needs to be purposeful if they are to serve an integrative purpose, that is, compatible with organizational goals. But the pursuit of organizational goals is subject to the defensive behaviors and vested interests that make double loop learning problematic; even when new channels are open, it is difficult to leave the silos. The solution seems to be in authentic collaboration. Boundary-spanning is much more effective if this activity is authentic and self-driven rather than imposed. Despite the great efforts that have gone into the design of knowledge sharing and accumulation technologies, knowledge sharing generally occurs on the basis of informal, personal social networks (Birkinshaw, 2001; Cross, Parker, Prusak, and Borgatti, 2001; Häkansson, 1990). Yet for self-driven connection and collaboration to meet organizational needs, a degree of normative alignment is needed across the connections.

Normative alignment is classically facilitated by face-to-face relationships and various forms of collocation, socialization to develop shared language, assumptions, attitudes, values, leadership development, and development of a global mindset - “the ability to cope with conflict and contradiction" (Evans *et al.*, 2002, p. 308). Employing an explicit networking model, (Cross *et al.*, 2001) see effective collaboration arising from an organizational culture where people know what others know, know how to access this knowledge, and engage with each other safely.

Again, however, there is something of a moral exhortation in these various prescriptions for normative alignment that seems removed from the organizational *realpolitik* of defensive and vested behaviors. This is especially difficult in multi-professional collaborations or networks because of the existent institutionalized professional communities, institutional power and the intricacies of translation and coalition-building (Contu and Willmott, 2003; Ferlie *et al.*, 2005). Further, collegial, pluralistic collaborative relationships tend to atrophy into centralized elite direction in the face of increased competition and financial imperatives (Robertson and Swan, 2004). Yet these amount to demands for more detailed management of collaborative interaction. As Greiner and Schein (1988) note, passive compliance and vested interests are less likely to hold sway in the face of active leadership.

3. Exploration – Exploitation

From a pervasive network perspective, silos and cross-boundary activities are simply different types of networks, or different regions in a sea of interaction. Silos are areas of dense multiple interactions, "strong ties", while boundary-spanning involves sparsely connected non-redundant "weak ties" (see Figure 1).[1] While the benefits of weak ties in terms of extended vision and flexibility have been eulogized in the notion of the "virtual firm" or "network organization", it is clear that successful innovation involves a combination of weak and strong ties.

In general, weak ties are a source of novelty and creativity but strong ties are needed to turn creativity into marketable innovation. Yet, weak dispersed ties provide limited transmission channels, normally developing only fragile and temporary "swift trust" (Jarvenpaa and

[1] Burt (1992) argues that it is the uniqueness of ties that underpins their sociological impact rather than their strength. Here, strength is used as a synonym for the redundancy of ties.

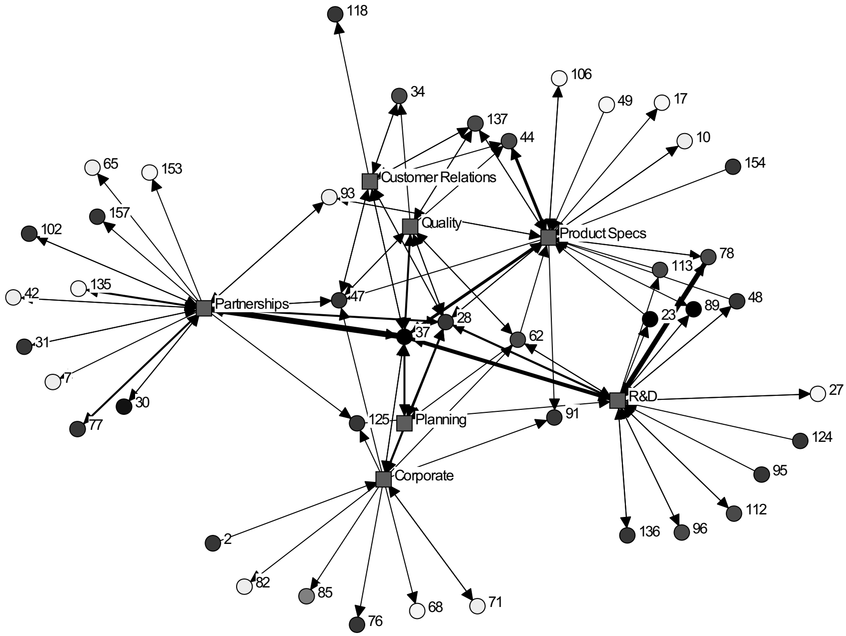


Figure 1. Example of clustering within a network of information exchange.

Leidner, 1999). Explorative performance, by contrast, is enhanced by extensive external ties and exploitation by dense concentration of internal ties (Hansen, Podolny, and Pfeffer, 2003). However, there is evidence of performance enhancement when experience of a one technology (scale economies) is combined with knowledge spillovers from other technologies (scope economies), experience producing greater absorptive capacity (Macher and Boerner, 2006).

But the particular combination of ties within and outside is significant as this provides differing access to diverse information. The benefits of a network of weak ties are dependent on the capabilities of the particular network members, so different networks provide different value (Adler and Kwon, 2002; Galaskiewicz and Zaheer, 1999; Gulati and Gargiulo, 1999; Podolny and Stuart, 1995). This is particularly evident in the presence of market imperfections, with firms establishing complicated international networks of subsidiaries to access particular markets and deal with particular governmental requirements and competitor actions (Nooteboom, 2004). Central and brokering positions within any network are of great significance; firms in intermediary positions between otherwise disconnected parts of a network are able to benefit by timely brokering of information across these parts (Burt, 1992; Uzzi, 1996). But complementarities between new knowledge and existing capabilities are also critical, particularly absorptive capacity (Dosi, 1988). Dyer and Hatch (2006), for example, found that Toyota gained better performance from its US suppliers than its competitors did from the same suppliers because of the relative configuration the supplier-purchaser networks. This arises from particular inter-organizational routines that provide long-term stability of supply and demand and knowledge sharing.

Mechanisms for optimizing collaboration for creativity along weak ties include positioning, establishing generative space, and employing knowledge brokers. Positioning involves locating skilled people with great tacit knowledge where uncertainties are greatest, such as marketers where consumer tastes are changing or technologists near centers of technological development (Afuah, 2003). This may be formalized in product development projects with suppliers or customers or in acquisitions of firms with new technology (Birkinshaw, 2001). Generative space, as celebratedly argued by Nonaka and Konno (1998), involves the creation of shared and safe location, time, context and knowledge to facilitate emerging relationships. Knowledge brokers are accidental or specially commissioned or figures, such as consultants or researchers, who by spanning different industries or technologies bring together strands of know-how (Hargadon, 1998).

In terms of strong ties, it has long been recognized that collaboration is facilitated by informal, decentralized and well-resourced organization frequent communication, proximity, clear goals and operational rules and influential project champions (Burns and Stalker, 1961; Gupta and Govindarajan, 1991). Mechanisms for optimising collaboration for the exploitation of new ideas, making use of strong ties include embeddedness, management by objectives, strategic alliances, centers of excellence, and communities of practice.

Close ties in themselves, that is, deep embeddedness, facilitate normative alignment (Ahuja, 2000; Dyer and Nobeoka, 2000; Gulati, 1995). Tacit and explicit knowledge is shared more readily along established relationships (Uzzi, 1996). This may be enhanced by management by objectives, combining empowerment, task setting and accountability to utilize knowledge (Dodgson, 1993; Grant, 1997; Zarraga and Bonache, 2003) and leveraging this approach beyond organizational borders with strategic alliances.

Because of the tendency for formal objectives to militate against authentic collaboration, generating defensive behaviors and rivalries, there has been increasing attention to the establishment of strongly embedded generative spaces, such as centers of excellence and communities of practice. Centers of Excellence are small groups with strategically valuable knowledge, tasked to develop this on the cutting edge and to make this known throughout the organization. These often act as a hub in an informal network (Moore and Birkinshaw, 1998; Andersson, Forsgren, and Holm, 2002; Frost, 2001). However, as Adenfelt and Lagerström (2006) point out, knowledge transfer may be more limited in a centre of excellence than in a purpose-built team because of process incompatibilities, that is, limitations in absorptive capacity. On the other hand, purpose-built teams normally lack experience working together, so take time to develop and are focused on perhaps overly-narrow objectives. Communities of Practice combine purpose-built teams with centers of excellence. Experts enrich their own expertise through collective reflective practice; collaboration around shared jobs or tasks, brokering relations with external experts, supported by IT, workshops, training, exchange and rewards (Wenger, 1998).

The critical condition for collaboration and innovation is likely to be dynamic leadership of divergent and convergent processes. Lindqvist, Söllvell, and Zander's (2000) "hourglass" model of varying degrees of control embraces this idea, identifying distinct managerial approaches for three stages of innovation. In this model, the first, initiation, stage demands control for divergence, encouraging diversity, initiative, entrepreneurialism, global perspective, loose structure, open communication, collaboration, slack, and tolerance for mistakes. This is supported by a culture of trust and reciprocity, participatory human resource management practices, long-term employment relationships, shared vision and goals, equitable sharing of rewards; socialization to shared norms, feedback, coaching and relationship building (Miles, Snow, and Miles, 2000). The second, development, stage is convergent, involving the selection of promising innovations, channeling resources to develop these (Weick and Westley, 1996). This is facilitated by local, tight control, clear goals and responsibilities, deadlines, and intense face-to-face communication of detailed tacit knowledge. The final, commercialization, stage is a return to divergence, involving global outreach, local trials in a wide variety of settings, conferences, and wide circulation of information (Miles *et al.*, 2000; Van de Ven, 1999).

Again, as long been recognized (Schumpeter, 1934), there is a major challenge in determining the balance between exploration, the search for novelty, and exploitation, the implementation of new developments. An overemphasis on exploration diverts resources and undermines exploitative capacity while an overemphasis on exploitation creates silos and limits renewal. But the value of either is difficult to determine as returns vary in terms of certainty, timing and impact. Organizations tend to focus on exploitation more than exploration because the returns from improved capabilities are more certain, more quickly realized and are more pervasive across the organization (March, 1991). March argues that the value of exploration is greatest when new findings cause the greatest variance in exploitative performance, allowing pockets of high performance to develop. Thus multiple exploratory projects with low single returns are likely to have a greater impact than a single coordinated project with a large return, provided communication is effective. And, given a requisite degree of complementarily, the linking of distinct competences is likely to be particularly innovative (Foss and Pedersen, 2004; Nohria and Ghoshal, 1997); higher levels of innovation have been found in firms that maintain a number of autonomous subsidiaries (Christensen, 2000; Chandy and Tellis, 2000).

4. The Network Perspective

Collaboration is generally examined in dyadic terms, the relationship between two collaborating parties, and this tends to create a focus on individual psychological states. A network perspective widens the focus of collaboration to a systemic level and provides new insights about the significance of network position, relative resources, network configuration, and multiplicity of relationships.

The position of an individual or an organizational unit in a network of informational interaction affects its access to information and ability to act. Significant positional features are connectedness to the rest of the network, centrality, and betweenness, or the ability to broker information between otherwise disconnected parts of the network (Burt, 1992).

Relative resources weights network position by the organizational importance of those most connected to or mediating. A limitation of the "virtual enterprise", or the firm as a portfolio of relationships without "bricks and mortar", is that some particular relationships are critical to the flow of value in a productive system and scale economies are normally concentrated here. Toyota's assembly plants, for example, are such sites in the value production network of suppliers, producers and retailers. It is difficult to conceive of a virtual enterprise that does not interact with such powerful sites in a productive system. Thus competing organizations with similar topographical positions in terms of network structure may be in widely different competitive positions because of the relative resources available in the immediate neighborhood.

Further, as already briefly discussed, it is not simply the position of a single individual or organizational unit within a network that influences the nature of collaboration but the configuration of the network as a whole determines the flow of information and the information available at any point. A notable feature of network configuration is that it is very sensitive to small relational changes, with changes to a few links having a great effect on network connectedness (See Figure 2). Further, a certain degree of redundancy in weak ties seems likely to be optimal, as demonstrated in the design of the internet. Too few connections are vulnerable to loss and to hostage situations; too many weak ties limits the dense connections necessary for exploitation. Because it is costly to maintain ties, bridges across "structural holes", or gaps between otherwise disconnected or relatively disconnect parts of a network, are most efficient (Burt, 1992).

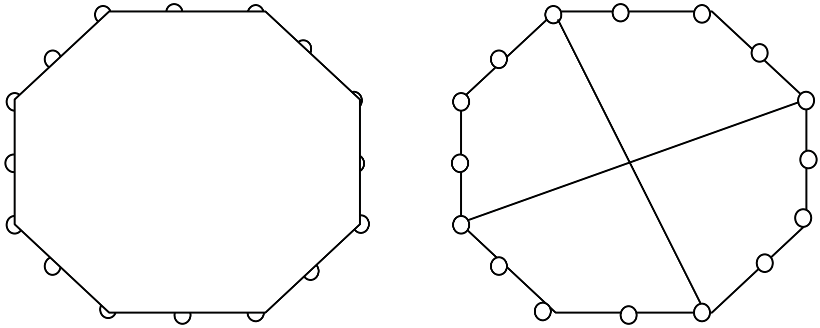


Figure 2. Sensitivity to small changes in connectivity.

Finally, while a social network mapping is essentially one-dimensional, social relationships are invariably multidimensional "bundles" of interactions (Burt and Schøtt, 1985). Analytical distinctions are commonly made between affinity, friendship, advice, and trust networks. Organizational collaboration normally encompasses multiple interacting networks beyond the organization, embracing suppliers, consumers, competitors and assorted stakeholders on many levels. Yet, in one of the few studies to examine this in detail, Cross *et al.* (2001) found significant interaction among analytically distinct relationships involved in the provision of advice; where a contact provided one function, they also tended to provide lower-order functions.

Network functionality, however, depends critically on the normative and purposive alignment discussed earlier. Because network outcomes emerge from the interaction of independent parties, outcomes can easily diverge from initially collectively agreed purposes (Rhodes, 1996). And because networks are dynamic, network configurations once contributing to core capabilities can become rigidities, external economies becoming external diseconomies (Leonard-Barton, 1992).

Network processes can delay feedback from stakeholders or clients and can isolate individuals or groups who could potentially utilize particular knowledge and can reinforce the insularity of silos (Rivera and Rogers, 2006; Cohen and Levinthal, 1990). Networks can become overloaded when these are used managerially to try to deliver new organizational imperatives (Miles and Snow, 1992). And networks often facilitate isomorphic responses (Westphal, Gulati, and Shortell, 1997).

5. Challenges in Managing Collaborative Networks

Recognizing the network context then adds a major layer of complication to attempts to manage collaboration. Managerial challenges include the diversity of potential and actual partners, the need to preserve flexibility, the complexity of network dynamics, the need for technical engagement, the maintenance of political support for a network approach, and assessing performance.

Prioritizing and sustaining valuable partnerships among the myriad of potential and changing connections is a formidable task, even given the difficulty of balancing the needs of exploration and exploitation. Lechner and Dowling (2003) suggest strong ties are likely to be more important at early phases of a venture and partnerships and weaker ties later. But there are also likely to be general biases towards established relationships, impeding the potential gains from diversity and flexible response to changing conditions. Continuous tension can be expected between direction and autonomy, and among differing goals (Agranoff and McGuire, 1999; Cohen and Rogers, 1992).

This is exacerbated by the complexity of network dynamics, the sensitivity of network effects to small changes in connectivity and the active reconfiguration of relationships by participants entails. In particular, network benefits spread rapidly once a critical mass of participants is reached. In a telephone network, for example, the benefits to any individual user rise exponentially with the ability to connect to a greater proportion of a population (Katz and Shapiro, 1985). For the same reason, network dysfunction is likely to spread rapidly. Because network connectivity provides benefits faster than simply the number of users, networks generate particularly rapid economies of scale; larger networks reduce unit costs as producers can spread fixed costs across a greater volume and reap learning economies. But where there are competing networks, this effect will not occur until one network gains a critically larger share of potential users, the “tipping point” where the network benefits outweigh the risks of purchasing an obsolete product or service. This produces a "winner takes all" situation that makes rivalry between networks intense (McGee, Thomas, Wilson, and Sammut-Bonnici, 2005). Yet, because of the active reconfiguration of multiple networks by strategizing participants, it is only possible to determine the critical point of connectivity after the fact.

While this demands the maintenance of a continuous strategic sensitivity, there is a continuous counteracting demand as effective engagement in networks requires a higher degree of technical involvement than in traditional managerial roles. This involves not only active involvement in group and inter-organizational processes but also a degree of inter- and trans-disciplinary knowledge (Agranoff and McGuire, 1999). In addition, there is a strong political dimension to active management of networking. This involves gaining and maintaining support for the use of resources outside the formal boundaries of the organization, consensual negotiation among external parties, systematic communication among a wide range of interests, and maintaining normative and purposive alignment (Agranoff and McGuire, 1999; Dougherty and Hardy, 1996). Further, while external collaboration is normally seen as risk-reducing (Podolny, 1994), the potential for rapid accumulation of network diseconomies demands sensitivity to potential new vulnerabilities (McLoughlin, Koch, and Dickson, 2001). Finally, great performance ambiguity appears in network contexts as the network externalities arise from the interaction of many varying small effects. This makes it very difficult to assess the contribution of particular individuals or organizational units (Agranoff and McGuire, 1999).

These challenges suggest a series of principles for managing collaborative networks. First, active engagement in the technical work of network building is necessary. Secondly, initiatives need to be carefully timed and targeted towards key nodes in networks, and must use appropriate agents and communications for those targets. Thirdly, dedicated brokers are needed to put together coalitions, while champions are needed to maintain them. Fourthly complex dynamics must be anticipated in networks with rapidly escalating positive and negative effects; this demands a management style of probing, sensing and responding rather than the linear approach of analyzing and directing. Finally, these considerations suggest the need for a distinctive set of management skills, beyond mastery of group process, including task integration, joint strategy, operations and finance, multiparty negotiation and contracts, and facilitation of complex dynamics (Agranoff and McGuire, 1999; Cigler, 1999; Kurtz and Snowden, 2003; Rivera and Rogers, 2006).

6. Conclusion

A network perspective, then, provides a radically different context in which to consider collaboration than that of dyadic relationships. While the qualities of immediate relationships remain important, these are contingent on the broader structure of interrelationships. A partnership is valuable in relation to other partnerships and the relationships of third parties. Network mapping appears to hold the promise of visualizing this sea of interrelationships.

Clearly, an important distinction for the management of collaborative networks is that between strong and weak ties. Regions of strong and weak ties underpin the exploitative and exploratory aspects of innovation respectively. Maintaining the balance between these different relationships is a major challenge, with strong biases in organizations to limit exploration, requiring high levels of detailed managerial intervention to counteract. While the exploration-exploitation dichotomy may be generally optimized with independent but complimentary projects, the network perspective offers the potential to marshal specific configurations of regions of strong and weak ties.

The network perspective also offers considerable analytical power as to the value of different sets of relationships. An organization's configuration of ties is not simply positional in terms of the centrality of the organization to the network as a whole but is also relational to the network ties of others, particularly competitors. And some ties are more valuable than others because of the particular resources available to different closely related groups. At the same time, networks are analytically complex; they are highly sensitive to small changes, over determined by relationships on different dimensions, subject to "winner-takes-all" phase-changes and prone entropically towards dysfunction.

This presents a formidable challenge for the management of collaboration. The scale of potential direct and indirect partnerships and the complexity of these extended arrangements demands a sophisticated strategic oversight coupled with technical and political skills in building, maintaining and assessing a wide set of relationships. This involves a much more detailed managerial intervention, yet simultaneously, the fluidity of networks militates against directives; much of the value of networks comes from the emergence of flexible authentic relationships.

The tensions surfaced by the network perspective between exploration and exploitation and between mandated and authentic interaction permeate global management today (see Evans *et al.,* 2002). Above all, it seems the precepts of complex management to use targeted interventions to probe, sense and respond offer the closer purview necessary for effective channeling of these fluid processes.

References

1. Adenfelt, M., and Lagerström, K. (2006). Knowledge development and sharing in multinational corporations: The case of a centre of excellence and a transnational team. *International Business Review, 14*(4), 381-400.
2. Adler, P. S., and Kwon, S.-W. (2002). Social capital: Prospects for a new concept. *Academy of Management Review, 27*(1), 17-40.
3. Afuah, A. (2003). *Innovation management: Strategies, implementation, and profits*. Oxford: Oxford University Press.
4. Ahuja, G. (2000). Collaboration networks, structural holes, and innovation: A longitudinal study. *Administrative Science Quarterly, 45*(1), 425-455.
5. Agranoff, R., and McGuire, M. (1999). Managing in network settings. *Policy Studies Review, 16*(1), 18-41.
6. Alvesson, M., and Karreman, D. (2001). Odd couple: Making sense of the curious concept of knowledge management. *Journal of Management Studies, 38*(7), 995-1018.
7. Andersson, U., Forsgren, M., and Holm, U. (2002). The strategic impact of external networks: subsidiary performance and competence development in the multinational corporation. *Strategic Management Journal*, 23(11), 979-996.
8. Argyris, C. (1990). *Overcoming organizational defenses: Facilitating organizational learning*. Boston: Allyn and Bacon.
9. Argyris, C., and Schon, D. A. (1978). *Organizational learning: A theory of action perspective*. Reading, MA: Addison-Wesley.
10. Arvidsson. (2000). Knowledge management in the multinational enterprise. In J. Birkinshaw and P. Hagström (Eds.), *The flexible firm: Capability management in network organizations*. Oxford: Oxford University Press.
11. Birkinshaw, J. (2001). Why is knowledge management so difficult? *Business Strategy Review, 12*(1), 11-18.
12. Burns, T., and Stalker, G. M. (1961). *The management of innovation*. London: Tavistock.

13. Burt, R. S. (1992). *Structural holes: The social structure of competition*. Cambridge, MA: Harvard University Press.
14. Burt, R. S., and Schøtt, T. (1985). Relation contents in multiple networks. *Social Science Research, 14*(4), 287-308.
15. Chandy, R. K., and Tellis, G. J. (2000). The incumbent's curse? Incumbency, size, and radical product innovation. *Journal of Marketing, 64*(3), 1-17.
16. Christensen, C. M. (2000). *The innovator's dilemma*. Boston: Harvard Business School Press.
17. Cigler, B. A. (1999). Pre-conditions for the emergence of multicommunity collaborative organizations. *Policy Studies Review, 16*(1), 86-102.
18. Cohen, J., and Rogers, J. (1992). Secondary associations and democratic governance. *Politics and Society, 20*(4), 393-472.
19. Cohen, M. D., and Levinthal, D. A. (1990). Absorptive capacity: a new perspective on learning innovation. *Administrative Science Quarterly, 35*(1), 128-152.
20. Cohen, W. (1995). Empirical studies of innovative activity. In P. Stoneman (Ed.), *Handbook of the economics of innovation and technological change* (pp. 182-264). Blackwell: Oxford.
21. Contu, A., and Willmott, H. (2003). Re-embedding situatedness: The importance of power relations in learning theory. *Organization Science, 14*(3), 283-296.
22. Cross, R., Parker, A., Prusak, L., and Borgatti, S. P. (2001). Knowing what we know: Supporting knowledge creation and sharing in social networks. *Organizational Dynamics, 30*(2), 100-120.
23. Delios, A., and Beamish, P. W. (1999). Ownership strategy of Japanese firms: transactional, institutional, and experience influences. *Strategic Management Journal, 20*(10), 915–933.
24. Dodgson, M. (1993). Learning, trust and technological collaboration. *Human Relations, 46*(1), 77-95.
25. Dosi, G. (1988). Sources, procedures, and microeconomic effects of innovation. *Journal of Economic Literature, 26*(3), 1120-1171.
26. Dougherty, D., and Hardy, C. (1996). Sustained product innovation in large mature organisations: Overcoming innovation to organisation problems. *Academy of Management Journal, 39*(5), 1120-1153.
27. Dyer, J. H., and Hatch, N. W. (2006). Relation-specific capabilities and barriers to knowledge transfers: creating advantage through network relationships. *Strategic Management Journal, 27*(8), 701-719.
28. Dyer, J. H., and Nobeoka, K. (2000). Creating and managing a high-performance knowledge-sharing network: The Toyota case. *Strategic Management Journal, 21*(3), 345-367.
29. Evans, P., Pucik, V., and Barsoux, J.-L. (2002). *The global challenge: Frameworks for international human resource management*. Boston: McGraw-Hill.
30. Ferlie, E., Fitzgerald, L., Wood, M., and Hawkins, C. (2005). The nonspread of innovations: The mediating role of professionals. *Academy of Management Journal, 48*(1), 117-134.

31. Forsgren, M., Johanson, J., and Sharma, D. D. (2000). Development of MNC centres of excellence. In U. Holm and T. Pedersen (Eds.), *The emergence and impact of MNC centres of excellence*. London: MacMillan.
32. Foss, N. J., and Pedersen, T. (2004). Organizing knowledge processes in the multinational corporation: An introduction. *Journal of International Business Studies, 35*(5), 340–349.
33. Frost, T. S. (2001). The geographic source of foreign subsidiaries innovation. *Strategic Management Journal, 22*(2), 101-123.
34. Galaskiewicz, J., and Zaheer, A. (1999). Networks of competitive advantage. S. Andrews and D. Knoke (eds.), *Research in the Sociology of Organizations* (Vol. 16, pp. 237-261). New York: Elsevier/JAI.
35. Grant, R. M. (1997). The knowledge-based view of the firm: Implications for management practice. *Long Range Planning, 30*(3), 450–454.
36. Greiner, L. E., and Schein, V. E. (1988). *Power and organization development: Mobilizing power to implement change*. Reading, MA: Addison-Wesley.
37. Gulati, R. (1995). Social structure and alliance formation patterns: A longitudinal analysis. *Administrative Science Quarterly, 40*(4), 619-652.
38. Gulati, R., and Gargiulo, M. (1999). Where do interorganizational networks come from? *American Journal of Sociology, 104*(5), 1439-1493.
39. Gupta, A. K., and Govindarajan, V. (1991). Knowledge flows and the structure of control within multinational firms. *Academy of Management Review, 16*(4), 768-792.
40. Gupta, A. K., and Govindarajan, V. (2000). Knowledge flows within multinational corporations. *Strategic Management Journal, 21*(4), 473-496.
41. Häkansson, H. (1990). Technological collaboration in industrial networks. *European Management Journal, 8*(3), 371-379.
42. Hansen, M. T., Podolny, J. M., and Pfeffer, J. (2003). So many ties, so little time: A task contingency perspective on corporate social capital in organizations. In S. M. Gabbay and R. T. A. J. Leenders (Eds.), *Research in the Sociology of Organizations* (Vol. 18, pp. 21–57). New York: Elsevier/JAI.
43. Hargadon, A. B. (1998). Firms as knowledge brokers: Lessons in pursuing continuous innovation. *California Management Review, 40*(3), 209-227.
44. Jarvenpaa, S. L., and Leidner, D. E. (1999). Communication and trust in global virtual teams. *Organization Science, 10*(6), 791–815.
45. Kamoche, K., and Mueller, F. (1998). Human resource management and the appropriation learning perspective. *Human Relations, 51*(8), 1033–1060.
46. Katz, M., and Shapiro, C. (1985). Network externalities, competition and compatibility. *American Economic Review, 75*(3), 424-440.
47. Kogut, B., and Singh, H. (1988). The effect of national culture on the choice of entry mode. *Journal of International Business Studies, 19*(3), 411–432.
48. Kogut, B., and Zander, U. (1992). Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science, 3*(3), 383-397.
49. Kurtz, C. F., and Snowden, D. J. (2003). The new dynamics of strategy: sense-making in a complex and complicated world. *IBM Systems Journal, 42*(3), 462-483.

50. Lechner, C., and Dowling, M. (2003). Firm networks: external relationships as sources for the growth and competitiveness of entrepreneurial firms. *Entrepreneurship and Regional Development, 15*(1), 1–26.
51. Leonard-Barton, D. (1992). Core capabilities and core rigidities: a paradox in managing new product development. (Summer Special Issue), *Strategic Management Journal, 13,* 111-125.
52. Lindqvist, M., Söllvell, O., and Zander, I. (2000). Technological advantage in the international firm: Local and global perspectives on the innovation process. *Management International Review, 40*(1), 95-126.
53. Macher, J. T., and Boerner, C. S. (2006). Experience and scale and scope economies: trade-offs and performance in development. *Strategic Management Journal, 27*(9), 845-865.
54. March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science, 2*(1), 71-87.
55. McGee, J., Thomas, H., Wilson, D., and Sammut-Bonnici, T. (2005). Strategy in the new economy. In J. McGee, H. Thomas, and D. Wilson (Eds.), *Strategy: analysis and practice* (pp. 457-498). London: McGraw-Hill.
56. McLoughlin, I., Koch, C., and Dickson, K. (2001). What's this "tosh"?: Innovation networks and new product development as a political process. *International Journal of Innovation Management, 5*(3), 275-298.
57. Miles, R. E., and Snow, C. C. (1992). Causes of failure in network organizations. *California Management Review, 34*(4), 53-72.
58. Miles, R. E., Snow , C., and Miles, G. (2000). TheFuture.org. *Long Range Planning, 33*(3), 300-332.
59. Moore, K., and Birkinshaw, J. (1998). Managing knowledge in global service firms: Centers of excellence. *Academy of Management Executive, 12*(4), 81-92.
60. Nohria, N., and Ghoshal, S. (1997). *The differentiated network: Organizing multinational corporations for value-creation*. San Francisco: Jossey-Bass.
61. Nonaka, I., and Konno, N. (1998). The concept of 'ba': Building a foundation for knowledge creation. *California Management Review, 40*(3), 40-54.
62. Nooteboom, B. (2004). *Inter-firm collaboration, learning and networks: An integrated approach*. London and New York: Routledge.
63. Podolny, J. (1994). Market uncertainty and the social character of economic exchange. *Administrative Science Quarterly, 39*(3), 458-483.
64. Podolny, J. M., and Stuart, T. E. (1995). A role-based ecology of technological change. *American Journal of Sociology, 100*(5), 1224-1260.
65. Rhodes, R. (1996). The new governance: governing without government. *Political Studies, 44*(4), 652-667.
66. Rivera, M. A., and Rogers, E. M. (2006). Innovation diffusion, network features, and cultural communication variables. *Problems and Perspectives in Management, 1*(2), 126-135.
67. Robertson, M., and Swan, J. (2004). Going public: The emergence and effects of soft bureaucracy within a knowledge-intensive firm. *Organization, 11*(1), 123-148.
68. Schumpeter, J. A. (1934). *The theory of economic development*. Cambridge, MA: Harvard University Press.

69. Senge, P. (1990). *The fifth discipline: the art and practice of the learning organisation.* London: Century.
70. Stacey, R. D. (2002). *Strategic management and operational dynamics: The challenge of complexity*. (4th ed.). Harlow: Prentice Hall.
71. Szulanski, G. (1996). Exploring internal stickiness: Impediments to the transfer of best practice within the firm. (Special Issue), *Strategic Management Journal, 17,* 27-43.
72. Uzzi, B. (1996). The sources and consequences of embeddedness for the economic performance of organizations: The network effect. *American Sociological Review, 61*(4), 674-698.
73. Van de Ven, A. H., Polley, D. E., Garud, R., and Venhataraman, S. (1999). *The innovation journey*. New York: Oxford University Press.
74. Weick, K. E., and Westley, F. (1996). Organizational learning: Affirming an oxymoron. In S. R. Clegg, C. Hardy, and W. R. Nord (Eds.), *Handbook of organization studies* (pp. 440-458). London and Thousand Oaks, CA: Sage.
75. Wenger, E. (1998). *Communities of practice: Learning, meaning and identity*. Cambridge: Cambridge University Press.
76. Westphal, J. D., Gulati, R., and Shortell, S. M. (1997). Customization or conformity? An institutional and network perspective on the content and consequences of TQM adoption. *Administrative Science Quarterly, 42*(2), 366-394.
77. Wilmott, H. (2000). From knowledge to learning. In C. Pritchard, R. Hull, H. Chumer, and H. Willmott (Eds.), *Managing knowledge: Critical investigations of work and learning* (pp. 216-222). London: Macmillan.
78. Zarraga, C., and Bonache, J. (2003). Assessing the team environment for knowledge sharing: An empirical analysis. *International Journal of Human Resource Management, 14*(7), 1227–1245.

This page intentionally left blank

Chapter 10

CONTEXT-AWARE AND ONTOLOGY-DRIVEN KNOWLEDGE SHARING IN P2P COMMUNITIES

Philip O'Brien
Faculty of Computer Science
Dalhousie University, 6050 University Ave
Halifax, Nova Scotia B3H 1W5, Canada
pobrien@cs.dal.ca

Syed Sibte Raza Abidi
Faculty of Computer Science, Dalhousie University
6050 University Ave
Halifax, Nova Scotia B3H 1W5, Canada

The knowledge management portfolio includes knowledge sharing as a means to connect knowledge to knowledge and knowledge to actors to support decision making, problem solving, viewpoint resolution, conflict negotiation, education and innovation. A knowledge sharing activity comprises two elements—(i) the content of the knowledge being shared and (ii) the *context* within which the knowledge is being shared. The context in which knowledge is sought and shared amongst peers is of significant importance in establishing the relevance and applicability of the knowledge content. In this paper we present a context-aware, ontology-driven knowledge-sharing framework that leverages ontology to both describe the knowledge sharing actors and the knowledge being shared. We model knowledge sharing in a peer-to-peer (P2P) network. Our P2P knowledge sharing framework comprises: (a) a domain ontology that is used to semantically model each peer; each peer is described as an instantiation of the ontology, (b) a weighted structural graph-based approach to establish affinity between peers and their contexts for the purpose of sharing relevant, needed knowledge resources, and (c) a task-feature relevance matrix to model the domain tasks influencing contextual affinity determination.

1. Introduction: Community Knowledge Sharing

The knowledge management agenda promotes (a) the *conversion* of information to knowledge and knowledge to actions leading to tangible outcomes; and (b) the *connecting* of knowledge to knowledge actors in order to ensure the sustainability of a knowledge-centric environment (O'Leary, 1998). Knowledge sharing, a profound and integral aspect of the knowledge management portfolio, stipulates the identification and dissemination of both tacit and explicit knowledge resources between like-minded individuals. Advances in information and communication technology have led to the design of innovative knowledge sharing environments and programs, whereby geographically dispersed individuals are virtually accessible to meet, collaborate, create, and share knowledge (Indulska et al., 2003; Strang & Linnhoff-Popien, 2004). We argue that, knowledge sharing is not just an activity, but in itself is a knowledge resource that needs to be properly harnessed and exploited.

Knowledge sharing can best be characterized as the behavior by which a community of practice—i.e. a group of individuals who share a common interest, need, or enterprise—engages in the sharing of its knowledge resources, insights, and experiences for a defined objective. Typically, knowledge-sharing activities involve the explication and sharing of tacit and experiential knowledge in terms of narrative and perspective sharing (Boland & Tenkasi, 1995), shared practices and activities (Cook & Brown, 1999), problem-solving experiences and evidence sharing (Curran-Smith et al., 2005), and exchanges of best practices (Szulanski, 1996). The objective of knowledge sharing spans from organizational learning to collaborative problem solving to peer-support to capacity building. However, in practice, the dynamics of knowledge sharing is quite complex and involves an active interplay between various operational determinants such as culture, community, incentives, medium, facilitation, ubiquity, and most importantly, trust (Abidi, 2006).

Promoting an effective knowledge-sharing environment is both intuitively attractive and functionally desirable. We argue that to design effective knowledge sharing environments it is not sufficient to address just the operational determinants of knowledge sharing. Rather, it is important to systematically model the prevailing *context* of the participating knowledge sharing actors—the contextual determinant of knowledge sharing will illustrate the knowledge actor's roles, responses, motivations, attitudes, actions, skills sets, and contributions to the knowledge sharing exercise, thus leading to more effective and focused knowledge sharing through identified and validated donors. We posit that a combined characterization of both the operational and contextual determinants of a knowledge sharing exercise may ultimately enable the knowledge actors to successfully share knowledge more effectively and also attribute a degree of trust in the knowledge being shared.

The emergence of dedicated communities of practice or special interest research groups has led to the creation of a large cumulative explicit knowledge capital—e.g. reports, articles, case-studies, etc.—that can be shared between the participating knowledge actors to enhance peer collaboration and peer support. We know that, in a research community the individual researchers hold within their localized knowledge repositories (i.e. computer memory) a vast number of research publications pertaining to their research interests and expertise. More importantly, these personal knowledge repositories withhold knowledge that is systematically validated by an actor—i.e. the actor has reviewed a paper and decided to store it or even use it in his or her own research work. The research question therefore is, how to design a knowledge sharing framework that facilitates the sharing of the personal knowledge artifacts of a group of actors; this is to be achieved in a pro-active manner such that a knowledge seeking actor is dynamically linked with other knowledge sharing actors based on the commonality of both their current research interests and the tasks that they are currently performing. Our objective is to effectuate *context-aware knowledge sharing*—i.e. a group of knowledge actors dynamically constellate to share knowledge that is directly pertinent to their immediate context.

In this paper, we present a context-aware knowledge sharing framework that facilitates the proactive exchange of knowledge artifacts—i.e. research papers—within a community of researchers actively engaged and inter-connected through a Peer-to-Peer (P2P) network. Knowledge sharing is guided by the active *context* of the knowledge actors or *peer*. We present a novel knowledge sharing context representation that extends beyond the typical simple feature-based profiling of a peer, whereby *context* is defined as a dynamic characterization of a peer in terms of (a) the peer's profile which entails the peer's research interests, activities, background, and demographic details; and (b) the peer's current *tasks* that mitigate the need for knowledge sharing, for instance whether the peer's current knowledge task is writing a research paper, preparing a critical review, or developing an educational module. Our context aware knowledge-sharing framework features the following elements:

(i) A dynamic *peer profile* that temporally models the ever-changing interests of the peer in order to present an accurate account of the peer's research interests and expertise. We argue that a peer's research interest varies with time and with involvement in different activities. Based on the peer's recent activities, our peer profile models the temporal variation in the peer's research interests by automatically modulating the significance associated with the peer's research interests. We use a domain ontology to generate a peer profile.

(ii) A detailed characterization of the various knowledge tasks that a peer may be involved with at the point of knowledge sharing. We argue that different knowledge tasks demand both a different knowledge sharing behavior and a different type of knowledge. Hence, for ensuring focused knowledge sharing it is important to identify the peer's current knowledge tasks and then accordingly direct the right knowledge to him or her. We have developed a unique time-sensitive *task-feature relevance matrix* that identifies the various tasks a peer is currently involved with and uses this information in the formulation of the peer's current context.

(iii) An *ontology-based context matching* method for determining the affinity between peers to ensure focused knowledge sharing between contextually similar peers. Our method uses weighted graph isomorphism, which compares the structure of two graphs, each representing the context of a peer.

(iv) *Active desk-top agent* that keeps track of all the applications currently running on the peer's computer in order to determine the current tasks being performed by a peer. For instance, if MS Word is the active window then the reading and/or writing of a document is implied as the peer's current task in the task-feature relevance matrix. Up-to-date information on the tasks being currently performed by a peer leads to the generation of his or her current context.

Operationally, our context-aware knowledge-sharing framework proactively determines the active context of the peers, establishes affinity between the peers based on their contexts, and subsequently facilities the sharing of relevant knowledge artifacts between contextually close peers.

The paper is organized as follows: in Section 2 we present a novel context representation for knowledge sharing. Section 3 details how our context model can be operationalized and leveraged in a knowledge sharing application. We detail the application's architecture and workflow in Section 4 and conclude with related work and future research in Section 5.

2. Eliciting Context for Knowledge Sharing

Efficient knowledge sharing is predicated upon identifying the right content within a specific context; the context is the encapsulation of the dynamic adaptive environment. Context, for our purposes, consists of: the orientation of the content author, elicited elements of a knowledge entity, the description of the active task—i.e. what, when, where and why a task is being performed and its corresponding knowledge needs (Kwan & Balasubrimanian, 2003). Context in knowledge management has not been sufficiently investigated (Ahmadjian, 2004). We argue that for effective knowledge sharing purposes the underlying constituents and characteristics of context need to be fully understood and formally defined to establish the scope and functional validity of any knowledge sharing solution. We have reviewed diverse approaches for defining context (O'Brien & Abidi, 2006), leading to our specification of context for knowledge sharing purposes.

In our work, *context* comprises two components: (1) an ontologically defined *peer profile.* The peer profile characterizes a peer in terms of relatively static elements such as research interest; some of the profile elements may vary albeit over a long time period; and (2) a description of the *knowledge task* mitigating the need to either share or source knowledge. The knowledge tasks correspond to the dynamic activities of a user that demand relevant knowledge, for instance writing a paper, reviewing a paper, preparing teaching material, and so on. This aspect of the context is quite dynamic as the user may engage with different knowledge tasks over a short time period.

2.1. *Peer Profile*

A peer profile entails salient attributes and behaviors of a peer that are relevant towards a knowledge sharing activity. The peer profile comprises attributes such as demographics, interests, expertise, knowledge seeking behavior and patterns, etc. In our work, the research interest aspects of the peer profile are derived from a domain ontology that characterizes salient domain concepts. We use the ACM Computing Classification System (CCS) for (a) representing the set of potential research interests and (b) indexing the knowledge resources (along the ontology-based concepts) that the peer may have in his or her knowledge repository. Figure 1 is a peer profile instantiation of an ontology representing the ACM CCS for use in a knowledge sharing system based in an academic setting.

```
..
<User rdf:ID="Philip">
  <hasUserID>Philip</hasUserID>
  <hasResearchInterests>#ArtificialIntelligence</hasResearchInterests>

<hasResearchInterests>#ObjectOrientedProgramming</hasResearchInterests>
</User>
<Article rdf:ID="Doc882885">
  <hasSubject rdf:resource="#ObjectOrientedProgramming" />
  <hasTitle>"A Study of Three Alternative Workstation-Server Architectures for
     Object-Oriented Database Systems"</hasTitle>
  <hasAuthor rdf:resource="#D DeWitt" />
  <hasAuthor rdf:resource="#P Futtersack" />
  <hasURL>http:##www.cs.wisc.edu#~fischer#ftp#pub#tech-
reports#ncstrl.uwmadison#CS-
      TR-90-907#CS-TR-90-907.ps.Z</hasURL>
</Article>

...
```

Figure 1. Academic research peer profile excerpt showing a User instance with its associated research interests, and an Article instance with its research topic and list of authors.

As shown in Figure 1, a peer profile is encoded as an RDF file, according to the semantics of the ontology. The peer-profile is operationalized using the Jena reasoning engine to collect statements about a peer interest and his or her knowledge artifacts for use in knowledge sharing activities.

2.2. *Knowledge Sharing Tasks*

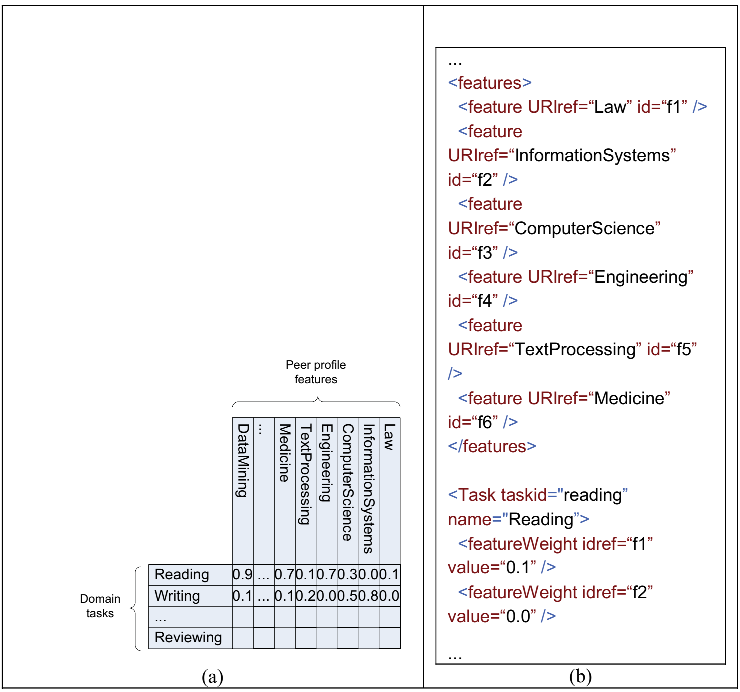
Knowledge sharing is pursued in response to the knowledge demands imposed by the *task* being performed by an actor. In our work, the spectrum of tasks that demand knowledge sharing may involve: (i) *Knowledge composition* in terms of an artifact, such as writing a research paper, literature review, technical/business report, operational guideline, educational material, etc; (ii) *Knowledge collection* in order to objectively respond to an issue; to gather evidence to critique or to validate a viewpoint/theory; to develop an understanding about a topic; to compare or analyze different viewpoints around a topic; and (iii) *Knowledge dissemination* to instruct or educate others; to inform others about research findings, viewpoints and outcomes; and to respond to a request from some actor for a knowledge artifact.

We argue that the nature of knowledge sharing for each of the abovementioned tasks is quite different—firstly the role, responsibilities, and expectations of the actors is different for each task; secondly the knowledge content being shared is different for each task; and thirdly the manner in which knowledge is consumed or discharged is different for each task. Therefore, the active knowledge sharing context for an actor needs to be characterized in terms of (i) the set of tasks in which the actor may be involved at any given point in time; and (ii) the knowledge content being shared—i.e. the subject/topics of the knowledge item and the modality of the knowledge item.

The knowledge demands for each of the abovementioned tasks significantly vary, albeit this variation is not accounted for and hence not incorporated in a traditional knowledge sharing system. This potentially leads to non-focused knowledge sharing—i.e. the user is required to explore the entire space of knowledge resources and establish the veracity of the knowledge with respect to his or her current task—which exerts an undue cognitive overload on the user. We argue that knowledge sharing, particularly in a P2P network, should be proactively adapted to the peer's current task, interest, disposition, etc.

2.3. *Task Feature Relevance Matrix*

We use the above characterization of knowledge sharing tasks to formulate a peer's context. Context is the coalescence of the *tasks* and the *features* of the environment in which the knowledge sharing activity takes place. Each task-feature pair is quantified by a *weight* that indicates the temporal and situational relevance of the profile feature to the peer's knowledge task. In our work, we have developed a *Task-Feature Relevance Matrix* (TFRM) that represents the active context of a peer, as shown in Figure 2.



Domain tasks	DataMining	...	Medicine	TextProcessing	Engineering	ComputerScience	InformationSystems	Law
Reading	0.9	...	0.7	0.1	0.7	0.3	0.0	0.1
Writing	0.1	...	0.1	0.2	0.0	0.5	0.8	0.0
...								
Reviewing								


```
...
<features>
  <feature URIref="Law" id="f1" />
  <feature
URIref="InformationSystems"
id="f2" />
  <feature
URIref="ComputerScience"
id="f3" />
  <feature URIref="Engineering"
id="f4" />
  <feature
URIref="TextProcessing" id="f5"
/>
  <feature URIref="Medicine"
id="f6" />
</features>

<Task taskid="reading"
name="Reading">
  <featureWeight idref="f1"
value="0.1" />
  <featureWeight idref="f2"
value="0.0" />
...
```


Figure 2. (a) Task-feature relevance matrix illustrating the active *context* of a peer. (b) matrix encoding using XML.

Contextual adaptation is achieved by varying the weights of each task-feature pair. The more relevant a feature to a particular task, the higher its corresponding value in the TFRM for that task vector—a row in Figure 2(a). The weights of the features that recede into disuse gradually decay, whereas active features have their weights increased. The range of the weight is [0.0, 1.0], and weights are mutually independent.

The matrix is implemented in an XML file. Based on the ontology used, features are declared which are uniform resource identifiers (URI) to the ontology file. These are acquired via the `rdf:ID` attribute of an ontology—each concept, property and relation bears this attribute. For each task encoded in the application, we define a block heading— Task in Figure 2(b)—with the task identifier followed by a series of feature references and their weights. This nesting and feature reference provides the intersection mechanism of domain features and domain tasks.

As the TFRM encoding is built upon the domain ontology, a unidirectional relationship exists between the TFRM and the ontology. As such, changes to the ontology conceptualization are reflected in the TFRM, and the TFRM acts as an overlay of the ontology. The overlay structure of the TFRM is shown in Figure 3 depicting the references from the matrix to the peer ontology.

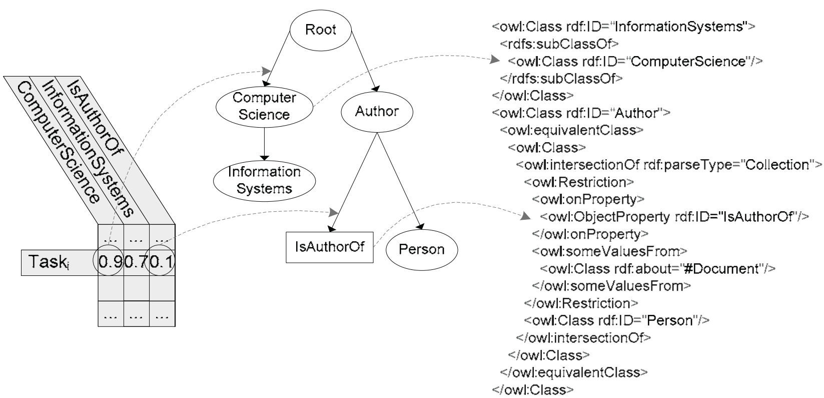


Figure 3. Task-feature matrix cross-referencing with the semantic object model of an RDF encoding for a peer.

Preliminary work indicates representing context as a weighted task-feature relevance matrix provides (a) fast computation of context affinity between knowledge actors, (b) accurate and flexible access to dynamic context in a network such as distributed knowledge resources, and (c) easy operationalization of the context feature relevancies as primitive data types.

3. Knowledge Sharing Operationalization

Having described the make-up of context in the previous section, in this section we describe the main elements of our context-aware knowledge-sharing framework. Our system is based on the Edutella P2P infrastructure (Nejdl et al. 2002). Edutella provides services for: (1) *querying* RDF metadata which describes the network and its peers, (2) *replication* for data persistence and workload balancing, (3) vocabulary *mapping* for interoperability, subsequently easing heterogeneity of network peers, and (4) *annotation* for describing material located within the Edutella network. Edutella is itself based on JXTA which provides facilities for peer discovery, peer group formation and registration, communication pipelines, content indexing, and file exchange. Upon the Edutella framework, we introduce our knowledge sharing application that comprises the following main components.

3.1. *Peer Affinity via Context Matching*

We use the TFRM to ascertain the affinity between peers of a knowledge network. The TFRM is an overlay of relevance values on top of the peer profile. A peer profile is based on an ontology, which is a web, or graph, of concepts. We can therefore combine the two to yield a weighted (relevance) semantic graph. Taking two peers to be represented contextually by these weighted graphs, we apply a weighted graph isomorphism technique for semantic objects (Oldakowski, 2005) to calculate their similarity.

A peer's profile features are arranged in descending order according to its relevance values for its current task. In this order, we traverse the graph of the other peer and multiply each relevance value by a scaling factor. The scaling allows features of less importance, but high relevance to hold some consideration in determining context without dominating more important features. Similarly, features of high importance but low relevance are scaled down to force less similar peers to the bottom of the list or resultant peers. As an example, suppose one peer (P1) has as its two most relevant feature weights P1:$\{f_1 \to 0.8, f_2 \to 0.7\}$ and the peer we are comparing to has P2:$\{f_1 \to 0.1, f_2 \to 0.84\}$ for these features. A scaling factor will place more importance on f_1 than on f_2. However, since the second most relevant feature (f_2) of P2 has a high value, it is given the chance to have an impact on the calculated similarity, albeit it to a lesser degree than for f_1. One of our scaling factors (S) uses an exponential scale to place significant emphasis on the most relevant features. For example, when we have S:$\{f_1 \to 0.8, f_2 \to 0.4, f_3 \to 0.2, \ldots\}$, then the partial calculated similarity between P1 and P2 is:

$$sim = (0.1 \times 0.8) + (0.7 \times 0.4) \times \ldots$$

The scaling values allow f_2 to bear influence on the result without skewing it such that a false affinity is calculated. If other peers—i.e. P3, P4, etc.—have values for f_1 higher than 0.1, these peers will be determined to be more contextually similar than P2. Algorithms 1 and 2 summarize how exponential scaling and peer affinity matching is performed, respectively.

Algorithm 1 EXPONENTIAL-SCALING

Input: The *size* of the peer profile feature set

Output: A list of scaling values, *w*, of length *size*, in descending order

sum := 0
for *i* in range [1, *size*] **do**
    sum := *sum* + i^2
end for
for *i* in range [0, *size*-1] **do**
    w[*i*] := $(size\text{-}i)^2$ / *sum*
end for

Algorithm 2 PEER-AFFINITY-MATCH Context matching between peers of a knowledge sharing network

Input: Two peer profile task-feature vectors: P1, the peer for which contextual similarity is being calculated, and P2, a remotely located peer

Output: The context-based affinity between P1 and P2

1. *order* := order-features(*P1*)
2. *scale* := EXPONENTIAL-SCALING(sizeof(*P1*))
3. *sIndex* := 0
4. *sim* := 0.0
5. **foreach** *o* in *order* **do**
6.   *sim* := *sim* + *P2*[*o*] x *scale*[*sIndex*]
7.   *sIndex* := *sIndex* + 1
8. **end for**

3.2. *Context Monitoring*

An active peer manifests a dynamic context; as the peer switches between knowledge tasks and manipulates different knowledge resources—e.g. reads research papers, views the web, exchanges email conversations, etc.—his or her context is continuously changing. These contextual changes are recorded in the TFRM and are tracked by context monitoring agents.

Our system has two environments: a desktop workspace and a mobile workspace. On the desktop, opening and closing applications indicates a change in task. Using a predefined set of applications installed on the test-bed network allows our system to be notified when an application is opened or closed, as well as any periphery file handles and interface usage. At such an event, the system inspects the list of currently opened applications and automatically infers the user's task based on pre-defined associations between actions and corresponding tasks.

The contextual change is determined by inspecting how the user is manipulating its environment with the opened applications. By parsing open files, URLs visited, or the content of email opened, for example, our system collects indicators of the type of work the peer is performing. The domain ontology is consulted when an application's state changes and the system determines what concepts are actually being exercised by the peer through the active tasks, which in turn leads to which profile features (research interests) are active at a given time. With the task and relevant features determined, we can cross-reference the TFRM with the user ontology and update the necessary values of the matrix—e.g. when a knowledge actor closes a document on “scrum software development”, a negative value is added to any entries of the matrix that are semantically related to that topic. This set of related concepts and their initial values is determined by a pre-classification bootstrapping process performed on each knowledge resource prior to its admittance into the knowledge-sharing network. The degree of the change is determined by: (a) the frequency of access and update of a task-feature pair, (b) values of other features in proportion to the one being changed, and (c) the absolute value of the feature value being changed—i.e. higher values are decremented slightly more than lower values.

3.2.1. *Semantic Propagation of Contextual Relevance Change*

To maintain semantic and logical consistency within the ontology and the TFRM, the complete set of concepts related to a context change is updated using cascading propagation of context. Increased relevance of a concept inherently implies increased relevance of its parent concepts, albeit to a lesser degree. Similarly, concepts that have a relationship but which are not direct ancestors in the ontology will also witness an increased relevance to a task. An example of context relevance propagation is shown in Figure 4. In our implementation, detailed in following sections, we propagate changes to a depth of three links from the source of a change.

Organizational hierarchies, for example, are a prime instance of where relevance propagation and context-awareness are prevalent. Consider an education institution such as a university. Graduate students work with professors who may run, or belong to, a specialized research lab. This lab in turn is founded in a department or faculty which is ultimately a component of the university. The knowledge of these people may be similar to their peers and less similar to their associates in

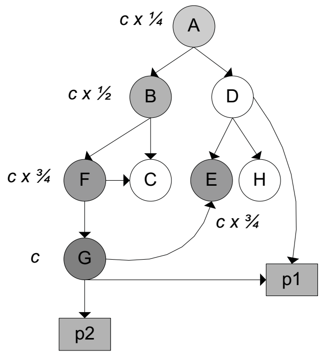


Figure 4. Sample context propagation. Changes in the relevance of feature G (by some amount *c*) will propagate up to F, B and A to a lesser degree, as well as to semantically related concepts and properties: E, p1, & p2.

other labs, departments, or universities. Knowledge sharing between peers mandates an increase in the relevance of common interests and goals of the knowledge-sharing actors. When knowledge flows between actors of similar labs, there is an innate rise in relevance of the actors of these labs. Furthermore, there is a heightened affinity between the departments and universities within which the peers reside. While the extent of the latter may be minor or negligible, it can and should be witnessed. The knowledge resource shared will have a set of features associated with it, derived from a domain ontology. These features are used to determine the related features that will observe the propagation of context change in other labs, departments, and universities.

3.2.2. *Relevance Decay*

In the earlier discussion, we argued a peer's research interests may vary over time depending on their longitudinal involvement with different research topics during various knowledge tasks. We model this modulation of the peer's research interests by updating the weights associated with the research interest features within the peer's profile. Note that the weights are bound in the range [0,1].

We employ two weight update functions that are periodically applied (say, at an interval of 1 day):

1. *Weight decay function*: This function models a peer's current lack of interest in a particular research topic (note that the research topic is part of the peer's profile). This is achieved by minimizing the weight of those research interest features that have not been involved in any tasks over a specific period. The weight decay is carried in an exponential manner—i.e. we begin by decrementing the feature's weigh in a smaller quantum over the short-term, and then with longer periods of inactivity we increase the weight decay quantum. We use the following intervals: (a) from latest update to 1 week we decay at a rate of 0.015, (b) from 1 week to 6 weeks we decay by 0.03 and (c) when a feature is neglected for more than 6 weeks its relevance has fallen sufficiently in the second interval, so it decays by a rate of 0.2.
2. *Weight excitation function*: This function models a peer's current increased interest in a particular research topic. The research interest involved in a current task experiences a weight increase.

3.3. *Profile Monitoring*

Throughout the knowledge sharing initiative, resources are retained and stored by peers. Monitoring a peer's profile for these changes involves documenting new knowledge resources added to the peer's knowledge repository, categorizing these according to the peer ontology, and adding them to the list of retained documents that may later be shared.

The peer profile encoding employs a mechanism to explicitly state the interests of a peer, the set of resources held by that peer, and the domain features associated with each resource. A sample encoding with live information from our application is shown in Figure 3. When affinity between peers is high, knowledge sharing takes place. The peer profile is used to determine which resources may be recommended by consulting the types of features with high importance and high relevance in the context and selecting the resources from the profile that have those features as attributes.

4. Knowledge Sharing Workflow and Control Flow

In this section, we present the operational workflow for context-aware knowledge sharing. Figure 5 shows the workflow process of the application from observing workspace context changes to ordering of affinity matching results to a user. When a context change occurs in the workspace, a change notification is sent to the context monitor ❶. The context monitor submits the change for the appropriate features to the TFRM ❷ for the current task, inputted from the task feed—a real-time agent for determining the task of the user. The context monitor notifies all interested agents of the change including the query agent ❸, propagation agent, and decay agent. While the others were described above, the query agent is tasked with interfacing with the profile search engine ❹ to acquire a list of peers from the network with their similarity to this peer. The profile search engine performs its work in two steps for each peer of the network: it consults the Edutella network interface to acquire the peer's task-feature relevance vector ❺a and submits this vector to a matching engine ❺b to calculate a similarity value. With an ordered list containing each peer and its similarity to this peer (for knowledge sharing purposes), the profile search engine returns the list to the query agent ❻ which posts the list to subscribed user interface components ❼. Subscribed interfaces may be desktop applications, mobile interfaces, or remote networked services. The user interface allows the human user to inspect the results and request transfer of knowledge resources from peers for as many resources as desired ❽+.

4.1. *Prototype*

We have developed a prototype implementation to test the operational aspects of the context matrix and elicit further requirements and possible improvements.

The prototype is designed to be resident in a peer's workspace and continually monitors the peer's context for changes that mandate an update of the list of relevant peers in a knowledge sharing community. A desktop agent is resident during execution of the application and

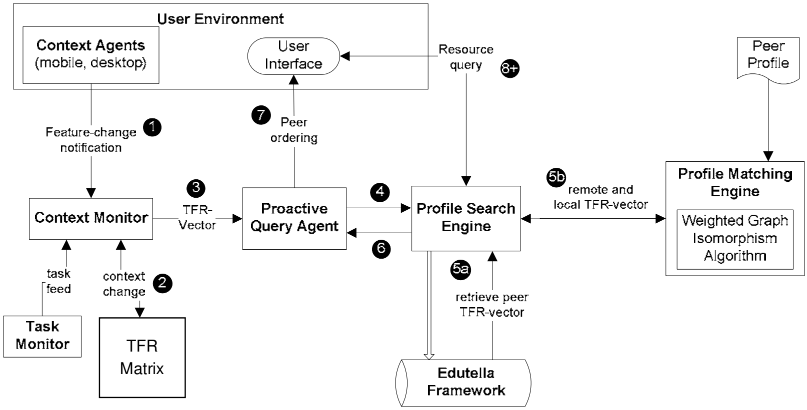


Figure 5. Dataflow model for context-aware peer knowledge sharing.

monitors the applications and file handles a user is using. When changes in context occur, relevance values in the user's TFRM are updated and affinity of the resident peer to all remote peers is recomputed. The results of matching against each remote peer are ordered according to similarity and the list is presented to all subscribed user interfaces. A user's peers are listed in descending order according to their contextual similarity, as shown in Figure 6. The peer alias and its similarity value in the interval [0.0,1.0] is given with a nested set of resources recommended by the system based on each peers current task and relevant interests.

Selecting a resource produces a metadata report window, Figure 7, with filename, keywords, optional abstract, and the sharer's comments if available. Herein lies the entry point of sharing knowledge via user confirmation of the knowledge resource's validity and relevance to the user.

5. Concluding Remarks

To date, much research effort has been devoted to identifying the relevant content within a knowledge network but without any consideration to the role and impact of the context in which the knowledge was sought and in which it is intended to be used. We have

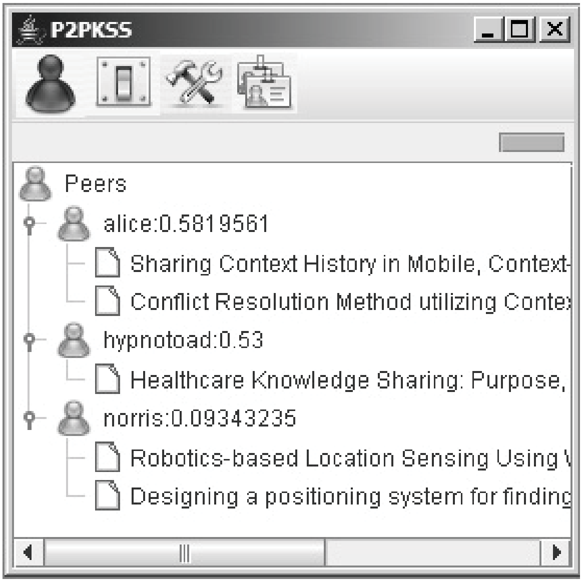


Figure 6. Application screenshot with peer list and contained resource list.

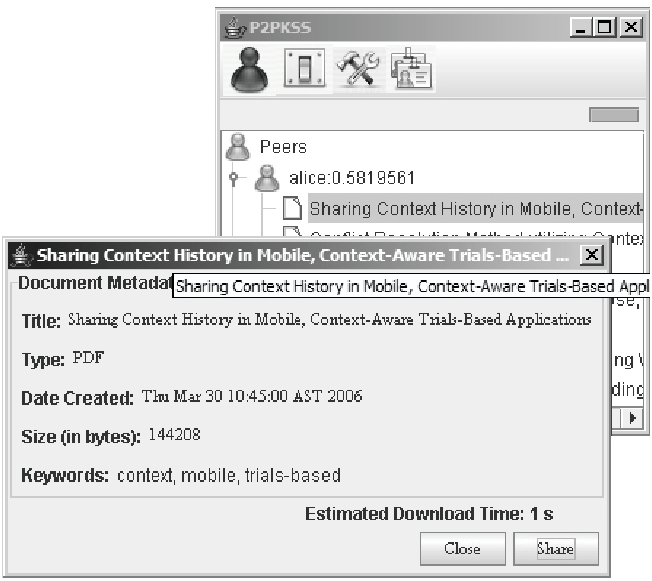


Figure 7. Selected resource can be viewed with resource metadata and can be shared.

introduced a novel representation model for context in knowledge sharing. It approaches context from a task-oriented, multifaceted ontology standpoint, utilizing domain semantics to systematically calculate peer affinity for the purpose of sharing knowledge.

Our prototype system has shown the model to be accurate and significant by way of preliminary investigations. Further research and application aims to expound the prototype in a second-phase implementation over a P2P knowledge network including desktop and mobile clients. We are currently completing implementation with a user feedback evaluation of its performance to follow. Evaluation of the desktop monitoring, task elicitation, and decay and excitation functions will be performed to gauge their effectiveness in modeling a user's knowledge sharing context behavior.

References

1. Abecker, A., Bernardi, A., Hinkelmann, K., Kühn, O., & Sintek, M. (2000). Context-aware, proactive delivery of task-specific information: the knowmore project. *Information Systems Frontiers, 2(3-4),* 253-276.
2. Abidi, S. S. R. (2006). Healthcare knowledge sharing: purpose, practices and prospects. In R. K. Bali & A. Dwivedi (Eds.), *Healthcare knowledge management: Issues, advances and successes*, Springer-Verlag. (in press).
3. Ahmadjian, C. L. (2004). Inter-organization knowledge creation: knowledge and networks. In H. Takeuchi & I. Nonaka (Eds.), *Hitotsubashi on knowledge management* (pp. 227-245). Singapore: John Wiley & Sons (Asia) Pte Ltd.
4. Boland, D., & Tenkasi, R.V. (1995). Perspective making and perspective taking in communities of knowing. *Organization Science, 6(4),* 350-372.
5. Cook, S. D., & Brown, J.S. (1999). Bridging epistemologies: the generative distance between organizational knowledge and organizational knowing. *Organization Science, 10(4),* 381-400.
6. Curran-Smith, J., Abidi, S. S. R., & Forgeron, P. (2005). Towards a collaborative training environment for children's pain management: leveraging an online discussion forum. *Health Informatics Journal, 11(1),* 31-43.
7. Indulska, J., Robinsona, R., Rakotonirainy, A., & Henricksen, K. (2003). Experiences in using cc/pp in context-aware systems. In LNCS 2574: In *Proceedings of the 4th International Conference on Mobile Data Management* (pp. 247-261).
8. Kwan, M. M., & Balasubrimanian, P. (2003). KnowledgeScope: managing knowledge in context. *Decision Support Systems, 35(4),* 467-486.
9. Nejdl, W., Wolf, B., Qu, C., Decker, S., Sintek, M., Naeve, A., Nilsson, M., Palmer, M., & Risch, T. (2002). Edutella: a p2p networking infrastructure based on rdf. In *Proceedings of the 11th World Wide Web Conference.* Honolulu, HI: ACM Press.

10. O'Brien, P., & Abidi, S. S. R. (2006). Contextual knowledge sharing in a p2p network: *Proceedings of the 8th IEEE International Conference on Engineering of Intelligent Systems,* Islamabad, Pakistan.
11. O'Leary, D. E. (1998). Guest editor's introduction: knowledge management systems: converting and connecting. *IEEE Intelligent Systems, 13(3),* 30-33.
12. Schmidt, A., Beigl, M., & Gellersen, H. (1999). There is more to context than location. *Computers & Graphics Journal, 23(6),* 893-902.
13. Strang, T., & Linnhoff-Popien, C. (2004). A context modeling survey. In *1st International Workshop on Advanced Context Modeling, Reasoning and Management* (UbiComp-2004).
14. Szulanski, G. (1996), Exploring internal stickiness: implementation to the transfer of best practice within the firm. *Strategic Management Journal, 17,* 27-44.
15. Oldakowski, R., & Bizar, C. (2005). SemMF: *A framework for calculating semantic similarity of objects represented as rdf graphs.* Poster at 4th International Semantic Web Conference, PID-58. Galway: Ireland.

Chapter 11

APPROACHING SCIENTIFIC KNOWLEDGE MANAGEMENT FROM ACTIONS COORDINATION, COMPUTER-SUPPORTED COLLABORATION, AND CONCEPT MAPPING

Germana M. da Nóbrega
Mestrado em Gestão do Conhecimento e da Tecnologia da Informação –MGCTI
Universidade Católica de Brasília – UCB
SGAN 916, Módulo B – Asa Norte, Brasília – DF (Brazil). CEP 70.790-160
gmnobrega@pos.ucb.br

Eduardo J. R. de Castro
Mestrado em Gestão do Conhecimento e da Tecnologia da Informação –MGCTI
Universidade Católica de Brasília – UCB
SGAN 916, Módulo B – Asa Norte, Brasília – DF (Brazil). CEP 70.790-160
ejrcastro@gmail.com

Edilson Ferneda
Mestrado em Gestão do Conhecimento e da Tecnologia da Informação –MGCTI
Universidade Católica de Brasília – UCB
SGAN 916, Módulo B – Asa Norte, Brasília – DF (Brazil). CEP 70.790-160
eferneda,glucena@pos.ucb.br

Gentil J. de Lucena-Filho
Mestrado em Gestão do Conhecimento e da Tecnologia da Informação –MGCTI
Universidade Católica de Brasília – UCB
SGAN 916, Módulo B – Asa Norte, Brasília – DF (Brazil). CEP 70.790160

The literature on computer-supported scientific collaboration substantiates that the community often focuses on supporting tools to assist in the execution of research projects. We are currently concerned with the design of a system that enables scientists to collaboratively structure a research project. This system is reliant upon both a widely known actions coordination model and a three-level workspace model. In addition, the academic work to which the proposed system is accounting is first examined with respect to higher regulating levels constrain work at the organizational and regulating agencies level. Such a perspective allows us to foresee further development toward actual scientific knowledge management.

1. Introduction

In this paper we focus on scientific knowledge by considering it within the tri-dimensional space engendered by three complementary axes, namely, *what*, *where*, and *how*, in an effort to establish a context in which to consider management. The *what* dimension allows us to characterize knowledge as scientific. We recall Valdés-Pérez's characterization of discovery in science (VALDÉS-PÉREZ, 1999) as the generation of *novel*, *interesting*, *plausible*, and *intelligible* knowledge about the objects of study as the initial criteria for thinking on scientific knowledge.

In order to establish a *where* dimension, we consider higher-education institutions and research centres as typical knowledge organizations, since generating (scientific) knowledge appears among their main goals. Taking the idea a bit further, we consider such academic work, aside from the scientific community regulation itself, as constrained by organizational interests and objectives drawn from the market place and/or social engagements. These are, in turn, constrained by (often federal) regulating agencies, as depicted in Figure 1. From such a context, we extract the academic level as the one we are currently addressing, i.e., a level where researchers have governance, even if constrained as we suggest above.

The third dimension characterizing our context with regard to scientific knowledge is *how*. To start, we look at scientific knowledge within its lifecycle, ranging from background research through publication. Our current work is considering such a lifecycle for the *development* of a research project. In order for a project to succeed, we suggest that a preceding and crucial phase similar attention, namely *elaborating* a research project. Being so, we address this important preceding phase by means of both collaboration among scientists and computational systems.

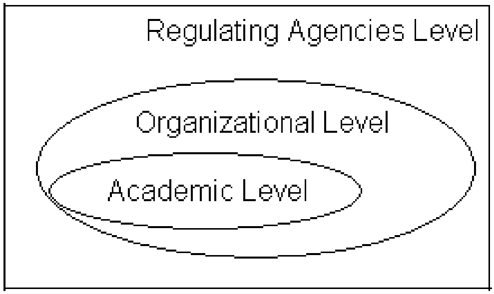


Figure 1. Academic work constrained by organizational and regulating levels.

In this paper, we present a portion of our current work concerning the direction of providing computer support for scientific collaboration during project preparation. In §2, we bring an answer to the question "what have computers been doing for scientific research?" by progressively characterizing the field known as *computer-supported scientific collaboration* (CSSC). In §3, we introduce conceptual grounds we borrow from the literature in order to build up the intended supporting system. In §4, the resulting CSSC system is described as an instantiation of those conceptual grounds. Finally, in §5, we present our concluding remarks and give focus to both ongoing and further work.

2. Characterizing Computer-Supported Scientific Collaboration

In conjunction with human scientists, computers currently play an important role in managing scientific knowledge. In this section, we develop this statement by identifying a number of computer systems. We firstly highlight how the field of Artificial Intelligence has moved one of its foci from *automated* to *human-computer* discovery systems. Then, we present more specific systems designed to support collaboration within *groups* of scientists.

2.1. *From Discovery Systems to Systems for Aiding Discovery*

The research on computational modelling of (scientific) discovery process has strongly contributed to the automation of such processes. As a starting point, Newell (1962) proposes the consideration of a discovery system as a problem-solving system guided by heuristic search methods. AM (LENAT, 1983), EURISKO (LENAT, 1983), BACON (LANGLEY, 1981), GLAUBER (LANGLEY, 1987), and BOOLE (LEDESMA, 1997) figure among the classical discovery systems acknowledged for the satisfactory results they exhibit on simulating historical discoveries from science.

In spite of this, as stated by Simon, Valdés-Pérez and Sleeman (1997), efforts are needed to guide the development of systems capable of solving problems in *cooperation* with domain experts. According to the authors, such systems would play an important role in discovery processes since when dealing with complex domains, there is little hope to successfully embody an achieved theory into the system, as this should constantly evolve. As a consequence, a system--in which design has accorded minor attention to the interaction with the system's user (or user's group)--will possibly fail when accounting knowledge evolution in its working domain. Thus, rather than designing discovery systems, we offer an alternative in terms of systems *aiding* discovery when one is interested in the system's capability to account for continuous evolution of its user's domain knowledge.

In that direction, Jong and Rip (1997) discuss the "computer-aided discovery environments" as a future designing perspective for integrating a variety of tools available for (a group of) scientists. As largely known, the scientific rationale includes a number of phases or processes (scientific knowledge life-cycle): scientists identify a problem then find an adequate representation for it; then they collect data by observation or experiment; and finally they find regularities and generalizations describing the data. As a consequence of the development of increasingly powerful AI tools, new possibilities are foreseen for discovery environments to support such phases or processes. For Jong and Rip, a set of integrated computational tools composes a computer-aided discovery environment. The authors highlight, however, the fact that such an integration should account for a "socio-technical system" view. The expression is borrowed from organization theory (TRIST, 1981) and, within the context of computer-aided discovery environments, it would refer to the way by which the available tools are brought together in order to be effective in practice. Jong and Rip propose a number of guidelines for the development of future computer-aided discovery environments.

Yet relevant recent work from the Computational Scientific Discovery (CSD) community point out to the benefits of the computer-aided discovery approach. Such benefits appear concretely through the results obtained by a number of systems capable of supporting the accomplishment of *totally original* discoveries. For instance, Pat Langley (2000) recalls AI research within CSD and its recent application to the discovery of new scientific knowledge. As evidence of the advantages of such human or machine cooperation, Langley reports seven examples of new (computer-aided) discoveries that have appeared in related scientific literature. He highlights the role played by humans in each case. Going further, Langley suggests five phases for computational scientific discovery in which human intervention may influence system's behaviour. The author explicitly recommends the computer-aided discovery approach, rather than criticizing human intervention, as often done in past AI. Following those guidelines, a more recent development, by Sanchéz and Langley (2003), is PROMETHEUS, a system providing interactive and iterative scientific model construction in a formalism allowing users to organize quantitative models into explicit processes.

2.2. *Computer-Supported Scientific Collaboration*

Beginning with open issues raised by scientific communities and eventually motivated by a formal call from funding agencies, scientists often organize in groups in order to elaborate a research project. Within our current work, we see such events as the beginning of scientific knowledge lifecycle. Considering the possibilities that may arise from distinct viewpoints, multi- or inter-disciplinary teams have been increasingly adopted to deal with complex problems requiring scientific approaches. However, in spite of the nature of a scientific team, elaborating and developing a research project usually requires some form of collaboration between individuals through the project lifecycle.

Moreover, considering Scientific Collaboration as enhanced by Information and Communication Technologies, we highlight in what follows (§2.2.1), some open questions addressed by the the concerned community and (§2.2.2) several works attesting some of the community's achievements in terms of supporting computational systems.

2.2.1. *Open Issues from the CSSC Community*

Let us raise some of the questions addressed by the CSCW 2000 Workshop on Lifecycle Support for Collaborative Science[1]:

- Do existing communication, coordination, and information-sharing tools provide the appropriate infrastructure for supporting scientific work throughout its lifecycle? If not, what new types of tools are needed?
- What are the differing collaborative support requirements during distinct phases of the scientific lifecycle?
- How do collaboration needs vary across scientific specialties? Is it possible to build general tools that service a variety of scientific communities?
- What are the special collaborative needs of cross-disciplinary teams?
- What are the challenges of supporting remote scientific fieldwork?
- What methodologies are most appropriate for studying scientific teams?
- How strong are the social barriers to information sharing in scientific communities? How do we build systems that respect the current social norms regarding information accessibility?
- How is e-commerce impacting the work practice of science, and how might e-business be integrated into collaborative project management?

Another discussion forum more recently organized concerning the subject of Scientific Collaboration in light of CSCW was the Workshop on Computer Supported Scientific Collaboration[2] held in conjunction

[1] http://sciencedesk.arc.nasa.gov/cscw2000/.
[2] http://ecscw2003.oulu.fi/ecscw_2003__ws10.htm.

with the European Conference on CSCW (ECSCW 2003). The workshop call also points to the interest in interdisciplinary collaboration, validating that scientific work involves orchestrating the manipulation of considerable amounts of variables and data which are well-structured or not. Bearing in mind such a perspective of heterogeneity in scientific work, a number of current complex problems and multi- and inter-disciplinary collaboration are concerned with certain challenges and necessitate a privileged research field for those interested in computer-supported collaboration.

2.2.2. *On Supporting Tools*

The ScienceDesk Project[3] carried out by NASA Ames Research Centre "conducts computer science research and development aimed at providing collaborative infrastructure support for distributed teams of scientists." The project includes among its research areas the one of Scientific Knowledge Management, aiming at capture, preservation, and traceability of scientific knowledge. Yet, the research area referred as to "Collaboratories" focus on tools for synchronous and asynchronous collaborative scientific teamwork.

A resulting tool from the project is SemanticOrganizer[4], "a specialized knowledge management tool designed to enhance the information storage, organization, and access capabilities of distributed NASA science teams". The tool provides a common electronic repository in which scientists can store and share information about projects.

Another tool from the ScienceDesk Project is InvestigationOrganizer[5], "a Web-based collaborative information management and modelling tool designed to support mishap investigation teams". The tool provides a centralized information repository that can be used by a distributed team of investigators to store digital products relevant to an ongoing mishap investigation. Also, "users

---

[3] http://sciencedesk.arc.nasa.gov/.

[4] http://sciencedesk.arc.nasa.gov/so/.

[5] http://io.arc.nasa.gov/.

can create and view overarching analysis models that identify causal factors or hypothesized event sequences leading up to a mishap".

3. On Actions Coordination and Asynchronous Collaboration

Literature review has revealed to us that the main focus of CSSC research regarding tools is supporting the execution of research projects. In a complementary approach, our current goal is to address research project elaboration on the grounds that it would increase the chances of execution to succeed.

In this section, we introduce the main pillars grounding the design of our system for supporting research project elaboration, namely (§3.1) a model of actions coordination exploited for organizing in (commitment-driven) phases such elaboration and (§3.2) a model for organizing the process into a three-level workspace. Within the section, these models are partially presented as they are originally proposed. In §4 we recall the description detailing how the models are tied together to compose the proposed system.

3.1. *Actions Coordination Cycle for Establishing Engagements*

Flores *et al.* (1988) propose an *actions coordination cycle* (Figure 2) as a model of actions coordination between "customers" and "providers", i.e, respectively, those owning a demand and those owning an offer. According to the authors, ensuring the customer's satisfaction requires two basic phases: *(i)* establishing a *commitment* and *(ii) accomplishing* it. The phase in which the commitment is attained includes two sub-phases, namely, *(i.a) context setting* and *(i.b) negotiation*. The phase in which the commitment is accomplished includes *(ii.a) execution* and *(ii.b) assessment*.

The actions coordination may take place under two possible workflow directions: one triggered by the customer's request, engendering a *request cycle*, and another triggered by the provider's offer, engendering an *offer cycle*. Within the request cycle, the action flow begins with the context setting for the customer request and finishes

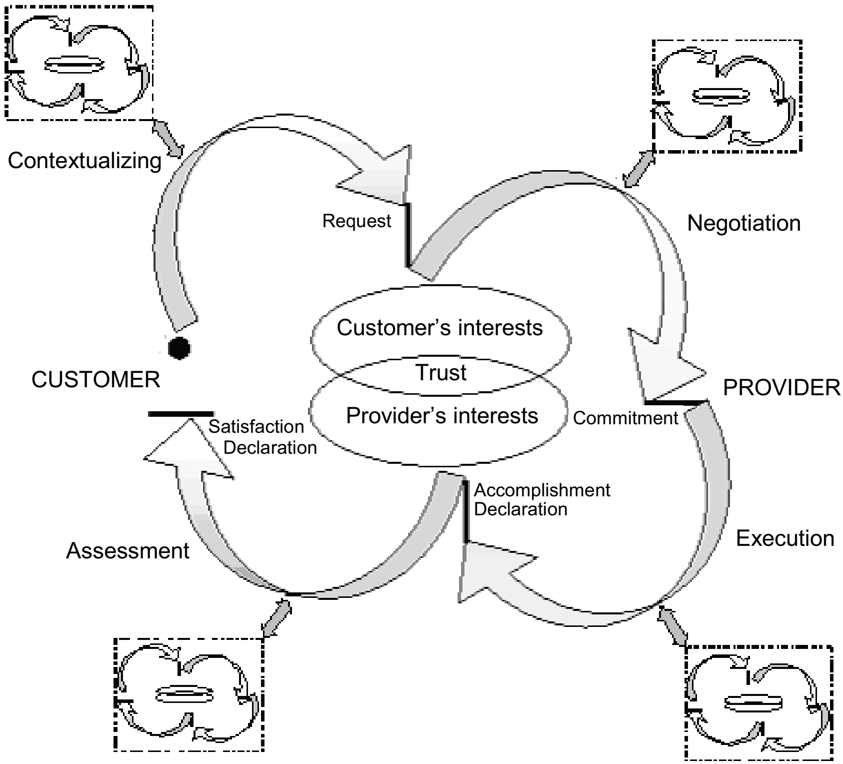


Figure 2. A simplefied view of Actions Coordination Cycle (Flores *et al.*, 1988).

at the starting point. Within the offer cycle, the action flow begins with the context setting for the provider, but, the request cycle, it ends up with a declaration of satisfaction or insatisfaction by the customer. Let us focus on the context setting phase.

As one might expect, the context setting phase not only prepares a context for the subsequent phase, i.e., negotiation, but also, and mainly, for all other subsequent phases. In other words, the quality and richness of communication between the parties in the context setting phase is crucial for the effectiveness of the whole interaction. According to Flores and his team, a clear and well-elaborated context may strongly contribute to establishing a trust-based relationship between the parties, and as a consequence, the possibility of satisfaction at the end is increased.

The first step of the context-setting phase is *formulating* the request. This begins with the inspection of customers' concerns. Once such concerns are identified, the next step is to *articulate* the request, starting from a clear identification of the involved parties--who is asking, what is being asked, whom is the request being addressed to--ending with a clear specification of the satisfaction conditions underlying the request.

In addition to the main workflow, the *actions coordination cycle* from Flores and colleagues foresees the possibility of recurrent cycles during each phase, as complex customer-provider engagements are seldom linearly deployed.

3.2. *A Three-Level Workspace Model for Promoting Collaboration*

Figure 3 depicts the AC-Hybrid "architecture", where "A"" stands for Asynchronous and "C" for Collaborative (CASTRO *et al.*, 2004). We recall those features further below. In Figure 3, Client A and Client B are each representing (abstractly speaking) a participant's machine. Server is representing a machine controlled by the group's coordinator. In general, a model is the result of representing the organization of one's ideas. Three kinds of models are distinguished, namely Individual Model, Global Model, and Collaborative Model. In handling an Individual Model, one has the opportunity to organize his or her ideas in a private manner, before he or she feels able to propose them to the group. Such a Model lives in a Client, and is viewed and controlled by the participant owning it.

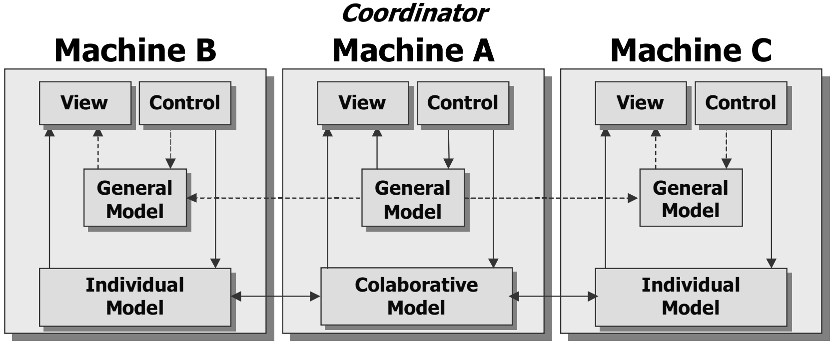


Figure 3. AC-Hybrid "Architecture" (CASTRO et al., 2004).

The Global Model should represent group consensus at a given moment. On the one hand, it should be stable so as to be usefully exploitable for the group's further elaborations. In other words, it would serve as a current group memory available anytime to be inspected by group members. On the other hand, a Global Model is supposed to continuously evolve in order to capture the group's cognitive progress. Similar to Individual Models, Global Models may also be versioned so as to keep track of group evolution. These two features assigned to Global Models--stability and predisposition to evolve--have suggested the need for an additional model justifying thus the "AC" part of the AC-Hybrid "architecture" or the so-called Collaborative Model.

The Collaborative Model plays the role of an intermediate model candidate in order to replace the Global one. It arises as a suggestion from a group member whose aim is modifying the Global Model. Such a suggestion would then be submitted to the group's analysis, and thus, trigger a debate. The environment in which the debate takes place, namely the Collaborative Environment, borrows the main structure from a forum. While the idea of asynchronous communication seems convenient herein to hold a discussion, the objective is, however, to provide a group with a mean to reach a consensus--a kind of drawn-up conclusion about one's suggestion. The idea of using models rather than free-speech which is usually considered within forum tools appears here to respond to that objective, thanks to a model's underlying structure.

4. A CSSC Environment for Supporting the Elaboration of Research Projects

Figure 4 depicts an actions coordination cycle, particularly a request cycle, beginning with a request from a funding agency--an offer calling (for research project proposals). A document describing the request may correspond to the contextualization within this request cycle. The negotiation phase would result in a contracted offer, corresponding to the execution of a project proposal. The execution phase represents the project implementation and finishes with the declaration of project conclusion by the contracted researcher(s). The product from this phase

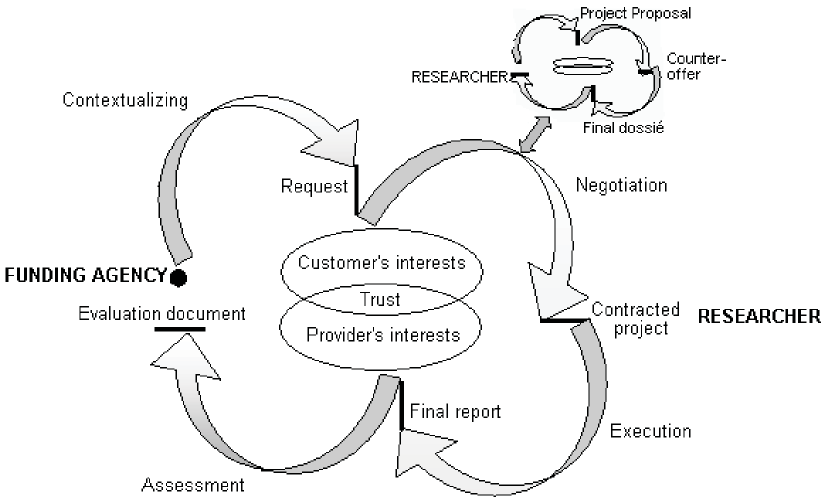


Figure 4. Main actions coordination cycle involving a funding agency and a researcher.

may be regarded as the final reports sent to the contracting funding agency. Next, the Assessment phase is corresponding to the evaluation carried out on the final reports which were received by the agency. The phase (as well as the entire scientific lifecycle) ends with the agency declaring its satisfaction or dissatisfaction of an evaluation document.

As suggested in §3, the actions coordination cycle foresees recurrent micro-cycles in order to account for the non-linearity of the process, thus allowing the ability to detail one phase in terms of a new cycle. As we intend to focus on the elaboration of a research project, the afore described cycle suggests that this would take place within the Negotiation phase, i.e., from the moment an offer calling is detected by a researcher, up to the moment the project that he or she has proposed is contracted by the funding agency. Let us then deploy the micro-cycle illustrated in Figure 4 recurring from the Negotiation phase.

The new actions coordination cycle begins with a researcher submitting a project proposal to the funding agency. The proposal document is then the product from the contextualizing phase. Next, the negotiation phase is accounting for the evaluation of that proposal by the funding agency, and the phase finishes with a response in terms of financial acceptance. In case modifications are required in the originally submitted proposal, the execution phase represents the process of modifying the proposal, as well as the arrangements to get the proposal actually contracted. The result from this phase is, then, the set of all documents required for the contract, including the proposal eventually modified. The following phase is the assessment one, representing the evaluation by the funding agency on the correctness and completeness of all required documents. This phase, along with the whole micro-cycle, are finished with commencement of the communication of the contract (and the liberation of the resources).

Within the aforementioned micro-cycle, the proposal elaboration takes place in the contextualizing phase. We reach the point where the AC-Hybrid "architecture" (§3.2) intervenes: we assume that such a proposal is collaboratively elaborated by a group of researchers, starting from a call from a colleague (who may possibly becoming the project coordinator). The group work is conduced in the three-level workspace suggested by AC-Hybrid. In Figure 5 we illustrate the progress of a hypothetical project proposal within the Collaborative Environment. The method we suggest is Concept Mapping, briefly presented below.

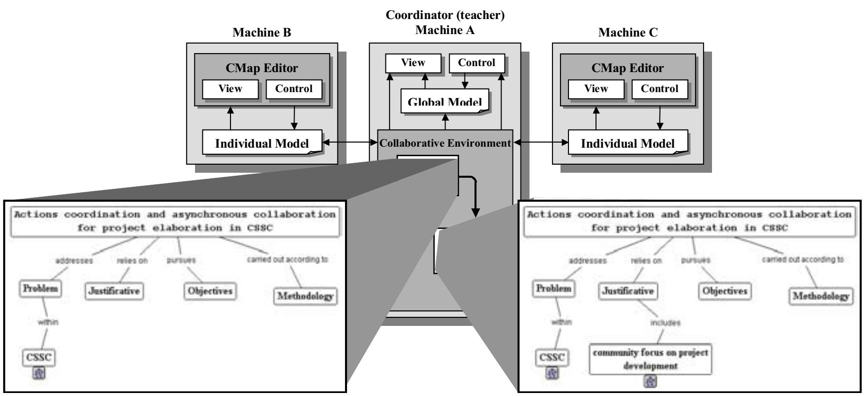


Figure 5. An AC-Hybrid instantiation for scientific project proposal construction.

According to Novak and Gowin (1984), Concept Maps (CMaps) are diagrams representing organized knowledge. They may be seen as graphical representations of one's knowledge, understood from concepts and relationships among them. We exploit Concept Mapping for the design of our system since they have been widely used to explain science in educational context, including collaboration (CAÑAS *et al.*, 2001). In addition to the opportunity to study concepts from a particular discipline, it offers an interesting means for developing soft skills, such as abstraction skill and others, underlying scientific reasoning. Such use context consider Cmaps' suitabilty for one to structure his or her ideas. This encourages us to highlight our system's feature on providing scientific work under a structuralist approach when one considers the project elaboration phase.

5. Conclusion

In this paper we introduce an overall context for scientific knowledge by characterizing knowledge as scientific, then considering a three-level scale for constraining scientific work and, finally, highlighting our concern with research project elaboration, supported by computational systems. We propose, then, the pillars we exploite toward an environment provide project elaboration within groups of researchers, namely, an actions coordination model and a three-level workspace model for promoting collaboration. While the system's current design is accounting to the academic level and not explicitly related to the organizational and regulating levels, our mid-term research agenda foresee to increase the environment, such as to help researchers in elaborating projects aligned to the goals established by those higher entities.

Acknowledgments

The work reported in this paper is under development with the financial support of *Conselho Nacional de Desenvolvimento Científico e Tecnológico* – CNPq, a brazilian governmental entity promoting the scientific and technological development, under the grant CT-INFO/MCT/CNPq no. 031/2004 Proc. no. 506911/2004-7.

References

1. Castro, E.J.R. de, Nóbrega, G.M. da, Ferneda, E., Cerri, S., & Lima, F. (2004). Towards interaction modelling of asynchronous collaborative model-based learning. *Proceedings of the ITS'2004--The 2nd International Workshop on Designing Computational Models of Collaborative Learning Interaction* (pp. 71-76), Maceió, Brazil.
2. Flores, F., Graves, M., Hartfield, B., & Winograd, T. (1988). Computers systems and the design of organizational interaction. *ACM Transactions on Office Information Systems, 6(2),* 153-172.
3. Jong, H. de, & Rip, A. (1997). The computer revolution in science: steps towards the realization of computer-supported discovery environments. *Artificial Intelligence*. Elsevier.
4. Langley, P. (2000). The computational support of scientific discovery. *International Journal of Human-Computer Studies, 53,* 393-410.
5. Langley, P., Bradshaw, G. L., & Simon, H. A. (1981). Bacon.5: The discovery of conservation laws. *Proceedings of the Seventh International Joint Conference on Artificial Intelligence* (pp. 121-126). Vancouver, British Columbia: Morgan Kaufmann.
6. Langley, P., Bradshaw, G.L., & Simon, H.A. (1981). Bacon5: The discovery of conservation laws. *Proceedings of the IJCAI-81.*
7. Langley, P., Simon, H.A., Bradshaw, G.L., & Zytkow, J.M. (1987). *Scientific discovery: Computational explorations of the creative process*, Cambridge, MA: MIT Press.
8. Ledesma, L., Pérez, A., Borrajo, D., & Laita, L.M. (1997). A computational approach to George Boole's discovery of mathematical logic. *Artificial Inteligence, 91,* 281-307.
9. Lenat, D.B. (1983). The role of heuristics in learning by discovery: Three case studies. In J. G. Carbonell, R. S. Michalski and T. M. Mitchell (Eds.), *Machine Learning An Artificial Intelligence Approach*. Palo Alto, CA: Morgan Kaufmann.
10. Lenat, D.B. (1983). Theory formation by heuristic search--The nature of heuristics II: background and examples. *Artificial Intelligence, 21(1-2).*
11. Newell, A., Shaw, J. C., & Simon, H.A. (1962). The process of creative thinking. In H. E. Gruber *et al.* (Eds*.), Contemporary Approaches To Creative Thinking*. New York: Atherton.
12. Sánchez, J.N.;, & Langley, P. (2003). An interactive environment for scientific model construction. *Proceedings of the Second International Conference on Knowledge Capture* (pp. 138-145), Sanibel Island, FL: ACM Press.
13. Simon, H.A., Valdés-Pérez, R.E, & Sleeman, D. H. (1997). Scientific discovery and simplicity of method [Special issue on scientific discovery]. *Artificial Intelligence*, *91(2),* 177-181.
14. Trist, E. (1981). The evolution of socio-technical systems. In: Van de Ven A. H. y F. Joyce (Ed.) *Perspectives on Organization Design and Behavior*. New York: John Wiley.
15. Valdés-Pérez, R.E. (1999). Principles of human computer collaboration for knowledge discovery in science. *Artificial Intelligence*, *107(2),* 335–346.
16. Novak, J.D. & Gowin, D.B. (1984). *Learning how to learn*. New York: Cambridge University Press.
17. Cañas, A.J., Ford, K.M., Novak, J.D., Hayes, P., Reichherzer, T. & Niranjan, S. (2001). Online concept maps: enhancing collaborative learning by using technology with concept maps. *The Science Teacher, 68(4),* 49-51.

Chapter 12

COMPARATIVE STUDY BETWEEN HOTELS AND AIRLINES E-COMMERCE SITES IN SINGAPORE

Yun-Ke Chang
School of Communication and Information
Nanyang Technological University, Singapore
65-6790-6866
ykchang@ntu.edu.sg

Miguel Angel Morales Arroyo
School of Communication and Information
Nanyang Technological University, Singapore
65-8123-0652
osornoyk@yahoo.com

The objective of this study is to compare the e-commerce sites for two industries, namely hotels and airlines, in Singapore using a comprehensive set of evaluation criteria designed specifically for e-commerce sites. The instrument used in this research was developed by Van der Merwe and Bekker (2003);); the e-commerce portal sites for this study were selected randomly. Multidimensional Scaling (MDS) was used in the data analysis to examine interval variables. Overall, the study showed that, compared with airlines, hotels received lower values and greater variability with interface, navigation, and content categories.

1. Introduction

Research and literature focusing on e-commerce evaluation is inadequate, in spite of an awareness of the importance of good portal design. The authors of this paper found this especially true for Singapore's e-commerce portals. This paper reports on evaluations of six randomly selected e-commerce portals in Singapore--three hotel sites and three airline sites--in terms of their degree of progress in five different dimensions, namely interface, navigation, content, reliability, and technical infrastructure. All e-commerce portals evaluated contain the domain ".sg". Recommendations based on the author's research are included at the conclusion of this paper.

2. Background

The Internet offers organizations a simple choice with regard to portal design: proactively take action to reorganize its business portal to accommodate the changing needs of on-line customers or reorganize it in response to advances made by organizations who have obtained a competitive advantage. One of the superior values a business can offer is its capacity to provide information, answers, and customer service to existing customers through the Internet, and in so doing, increase customer retention and improve business relationships. As a result, competition between organizations is intense because any competitor is but a click away (Yan, et al, 2003).

E-commerce over the Internet has created significant opportunities for companies in countless business environments, enabling them to go beyond geographical boundaries and locations in order to interact with their customers, members, and other stakeholders. Company portals can either enhance or hamper this important interaction. While well-designed portals can help create loyal clientele and increase profits, poorly designed ones may lead to frustrated consumers and subsequent losses for the enterprise (Cunliffe, 2000).

According to the fifth annual A.T. Kearney-Foreign Policy Globalization index, Singapore was ranked first and recognized as the most global nation. This accomplishment is attributed in large part to its high trade levels, heavy international telephone traffic, and steady stream of international travelers. The report described Singapore as a country with a modern, open economy, and one with little corruption.

According to the CIA's World Factbook, there were 2.75 million online populations in Singapore in 2005, a figure representing approximately 62% of the population. With a comprehensive strategy to transform its economy into a network economy and build an 'intelligent' island, the Singapore government has carried out a series of actions ranging from establishing a large scale information infrastructure and removing hindrance to private sectors' participations, to developing a skilled workforce. However, to fully take advantage of these government initiatives, companies must ultimately assume responsibility to determine the role of and subsequently design their online activities. Customers must then be able to interact with e-commerce Web sites to receive products and/or services provided by these companies.

In spite of the fact that measuring and describing the Information Communication Technology (ICT) development of a country can easily be done with indicators such as the number of online users, the availability of high speed broadband network, and its governmental policy environment, the true revenue of e-commerce comes from the actual number of transactions on which customers' online experience and satisfaction have great impact. Yet, due to the lack of a systematic approach to assess Web sites, it is difficult to know how well online customers are actually served or informed via a company's online portal. Existing assessment models for e-commerce Web sites include those by:

- Schubert and Selz (1999) who described a Web assessment model created by the Competence Center for Electronic Markets;
- Liu et al. (2000) who proposed criteria for the design of e-commerce Web sites derived from a survey of Web masters working for Fortune 1000 companies;
- DeLone and McLean (2004) who proposed six dimensions, namely system quality, information quality, service quality, use, user satisfaction, and net benefits, for the evaluation of e-commerce Web sites;
- Phippen et al. (2004) who considered customer lifecycle analysis and customer behavior analysis in their research in Web analytics;
- Kim and Lee (2003) who conducted research on e-catalogs evaluation;
- Mao et al. (2005) who reported survey data regarding top measures of effectiveness for Web sites in businesses employing User Center Design (UCD); and
- Van der Merwe and Bekker (2003) who proposed a comprehensive set of evaluation criteria for e-commerce Web sites.

The evaluation criteria proposed by Van der Merwe and Bekker (2003), which focus on the customer's purchasing process, was adopted in this study because they are broader than other aforementioned frameworks of which the author's reviewed. This model's evaluation criteria incorporate five distinct categories: interface, navigation, content, reliability, and technical infrastructure.

3. Objectives

The objective of this study is to compare the e-commerce Web sites for two industries, namely hotels and airlines, in Singapore using a comprehensive set of evaluation criteria designed specifically for e-commerce Web sites. The significance of the study is based on two factors: (i) little research has been done in this area, and our literature review has found this especially true for Singapore; and (ii) the results obtained from this research could help businesses enhance their e-commerce Web sites.

4. Methodology

By and large, people believe that in order for a research study to be statistically sound, to identify problems, or to make generalized results, a research study comprised of a large sample size is preferable to a smaller one. However, it is unfeasible to perform a large-scale Web site evaluation if one wants to conduct a comprehensive assessment. Because Nielsen and Landauer (1993) have suggested that utilizing three test users for a medium-large software project could give the maximum cost/benefit ratio, the authors of this paper believe that three randomly selected sites are sufficient to reveal the general features of and problems with the entirety of airline and hotel e-commerce Web sites in Singapore. The entire population of hotels and airlines that offer their services on-line with Singaporean domain (.sg) were identified using a popular search engine. A list of these Web sites was compiled, and three sites from each category were randomly selected for evaluation. The criterion for inclusion in the list was that the Web site be registered in Singapore with .sg domain name. A total of six e-commerce Web sites were evaluated. The rationale for choosing the Web sites specific to Singapore was to enable our study to focus on those activities labelled as interesting to tourists visiting Singapore. The data for this study was collected in February 2006.

In order to collect e-commerce evaluation data and summarize the results, three graduate students were trained in the assessment criteria required by the Van der Merwe and Bekker (2003) instrument. The students collected data under the same conditions for each e-commerce Web site studied, using the same type of Internet connection, computer hardware, and operating system. After data for each site was collected, the evaluators met to identify and resolve disparities. Appendix 1 details the evaluation criteria. MDS analysis was applied to the interval data with values from zero to ten. For nominal variables the values are tabulated and displayed in a graphic.

Instrument. The instrument used in this research took criteria developed by Van der Merwe and Bekker (2003). We adapted their measure to our needs. The instrument has approximately 100 items within five categories in which to assess data. Appendix 1 details the way in which these categories are operationalized.

Study limitations. Some items related to security, system design, software, and databases used by the e-commerce Web sites were unfeasible to evaluate. Likewise, some sub-categories in the instrument are not mutually exclusive, and therefore, a source to validate such was not found.

5. Data Analysis

Multidimensional Scaling (MDS) is a set of non-inferential statistical scaling procedures facilitating the discovery of hidden patterns in empirical data (Kruskal & Wish, 1978; Shepard, Rommey & Nerlove, 1972). Furthermore, MDS attempts to represent in *n*-dimension space the similarities between entities by mapping the distance between them. Distance between entities represents a measure of similarity (Coxon, 1982; Dunn-Rankin et al., 2002).

Multidimensional Scaling (MDS) was used for the analysis of the data. The objective was to create a map using SPSS™ 11.0 for Windows® that helps one to visualize the e-business of an organization and identify those dimensions that explain their differences. The best representation requires finding a solution with minimal stress, and one that is not degenerated.

In this study, a two-dimensional representation offers the best combination of minimal stress value and RSQ, an indicator of variance in relation to the original data. The Shepard diagram shows a good relationship between $f(\delta_{ij})$ and (d_{ij}), and the result of MDS mapping can be seen in Figure 1.

According to Dunn-Rankin et al. (2002) the process for doing MDS is described as follows: a) using a given a set of (m) entities, calculate the matrix of proximity. In this case, Euclidian distance was used; b) a number of dimensions (n) *were* chosen. The m entities are placed in the n-dimension space; c) MDS represents the proximities $f(\delta_{ij})$ between the pair of entities in the n-dimension chosen attempting to maintain the original distances (d_{ij}); d) the objective of MDS is to minimize the difference between $f(\delta_{ij})$ and (d_{ij}). In other words, when $f(\delta_{ij}) = (d_{ij})$, the algorithm represents perfectly the original data in the n-dimension space, which can then be represented $f(\delta_{ij})-(d_{ij}) = 0$. MDS uses the following formula to calculate the compatibility: $(\Sigma\Sigma[f(\delta_{ij})-(d_{ij})]^2/\text{scale factor})^{1/2}$, known as stress. The closer the stress value is to zero, the better the fit (Kruskal & Wish, 1978; Borg & Lingoes, 1987), thus the method approaches the solution in successive approximations.

Minimal stress can be found using the least square regression; however, computational procedures use what is known as "the method of steepest descent" (Kruskal & Wish, 1978, p. 27). The number of dimensions affects the stress value. The higher the number of dimensions, the lower the minimum value of stress. However, complications associated with the number of dimension include the fact that: a) it is not possible to visualize more than three dimensions; b) the solution could degenerate or become trivial--"the points of the configuration are strongly clumped" (Kruskal & Wish, 1978, p. 29); and c) given m entities and n dimension, there is a rule of thumb: if $m > 4n$ the stress is not sensitive to m and n (Kruskal & Wish, 1978).

The results of MDS were obtained and are shown in Figure 1. The difference in value for different variables was calculated for interpretation. For example, to identify the differences between hotel 2 and airline 5, the differences for each pair of values was calculated, isolating and studying, both the highest positive and negative values of the difference.

Furthermore, multiple variables were compared simultaneously to identify differences. For example, hotel 3 (h3) and airline 5 (a5) were compared in order to analyze the differences in a quadrant. The differences of scores of each evaluation criteria within subcategories were compared. The same procedure was repeated on airline 3 (a3) and airline 2 (a2). Once the differences for each quadrant were identified, comparisons between quadrants were performed in order to identify the evaluation criteria that caused those differences. This analysis was performed with the objective to identify the dimension's components.

Those with positive values in dimension 2 were found to be sites with higher scores in evaluation criteria, e.g. the privacy of their users protected; their security systems properly accredited; ease of log-on; perceived benefits for registration; and ease to register on the Web site (Figure 1).

Those with postitive values in dimension 1 were found to be the sites with higher scores in these criteria, e.g. no broken links, ease of site use, adequate amount of advertisement by other companies, alternative methods for ordering and payment, the use of graphics and multimedia, consistent navigation throughout site, ease of help function use, and consistent navigation through the site (Figure 1).

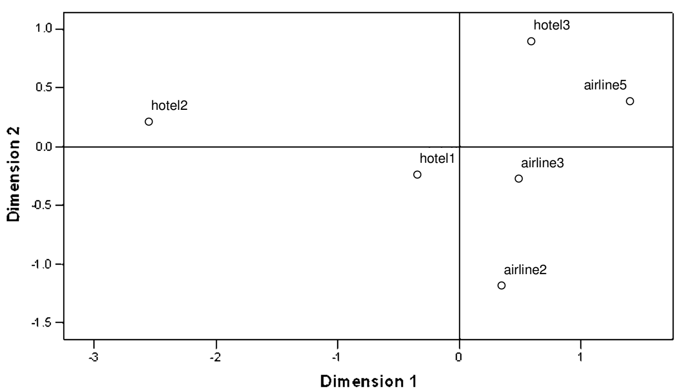


Figure 1. MDS map for e-commerce sites.

In the MDS map, Hotel 2 is plotted far from the rest of the Web sites (See Figure 1). The reasons were identified as follows: the site does not have search engine or help function; the ordering process is not transparent and nor easy; and the site does not have accredited security system or cross-browser capability.

MDS and cluster analysis were performed, and the variables were collected in four groups (a, b, c, and d) based on their proximity in the MDS map and the average per cluster. Figure 2 shows the average for each type of group and for each business. It can be appreciated immediately that hotel 2 has the lowest values for clustered variables "a" and "b". Variables "a" include the criteria in the following categories: technology, reliability, and content. In the technology category, hotel 2 has problems with security such as payment-related issues and lack in the use Hyper-Text Transfer Protocol Secure. In the reliability category, hotel 2 has problems with reply to user emails, hotel's contact information, and price. In the content category, this hotel has problems describing the service provided and the fact that the home page is confusing.

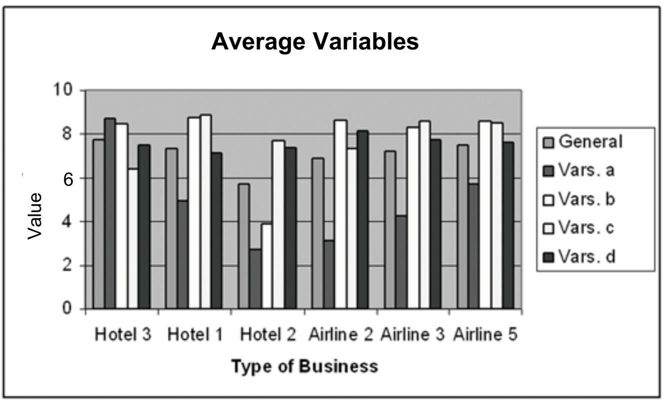


Figure 2. Average values from groups of variables derived from MDS Analysis.

Group "b" includes criteria from the following categories: interface, reliability, navigation, and how frequently the content is updated. Hotel 2 has problems with this group of variables, specifically with the order process including terms and conditions; the way in which the interface is designed; and its search engine.

Group "c" includes criteria from the following categories: technology, interface, navigation, and some issues with information quality. Hotel 3 has the lowest values in this variables analysis. The main technology problems are related with respect to the speed to access the sites. The problems with navigation in this category are related to the logical structure of the site. The problems with interface are the use of color, as well as the use of media when the connection to the server is slow. Also in Hotel 3, there are problems with information quality.

Group "d" includes criteria related to interface, navigation, and content categories. In general, all the businesses studied have relative high values.

Figure 3 shows the average for category as defined in the instrument used in this study. From this figure, we can appreciate that hotels have more problems with reliability than airlines. Variability in the technology category can be appreciated in Figure 3. However, this result is questionable due to the fact that the researcher did not have access to examine that category with detail. From this graphic, we can appreciate that hotels have more variability with interface, navigation, and content categories than airlines.

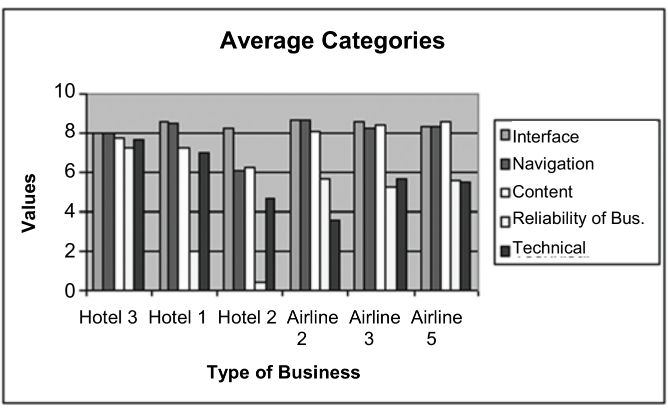


Figure 3. Average by category.

The representation of nominal variables is shown in Figure 4 in which we can appreciate that all businesses show some type of contact information, which is part of the content category. The items that belong to the content information sub-category are email address, telephone and fax numbers, and physical address. In this area, the hotels, with the exception of hotel 1, do better than airlines.

Flexibility and compatibility sub-category belongs to the interface category and includes items like printable version, text-only version, foreign language support, and accommodation for disable users. In this area, none of the business presents well, but overall hotels present a little better than airlines. This is an unexpected result.

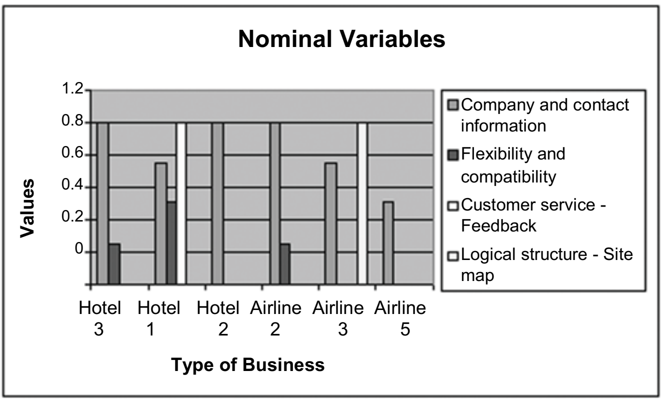


Figure 4. Nominal variables analysis.

The only Web site to offer feedback opportunity was hotel 1. A site map is offered only by airline 3. Feedback form falls under the customer service sub-category, which is part of reliability category. Site map is part of the logical structure sub-category, which is part of navigation category.

6. Conclusions

The results of this study are preliminary. The authors do not know if the instrument used has been validated, and from the results obtained in this study, it appears that the sub-categories are not mutually exclusive. Differences were found between the two business categories, but given the restrictions of this study it is not known if the differences are significant.

It was found that consistent navigation throughout the site is higher among hotels than airlines, and more broken links were found among airlines than hotels. However, airlines studied did better than hotels in the following items: the use of graphics and media, the existence of a help function, the use of a search engine, better management of user information, the use of more advertisement on other sites, and the offer of an increased number of methods for ordering services. In general, the hotels studied have lower values and more variance with interface, navigation, and content categories than airlines. Finally, the authors were expecting that in a multi-cultural society, more foreign language support could have been found when compared to other countries; however, that was not the case.

References

1. Borg, I., & Lingoes, J. C. (1987). *Multidimensional similarity structure analysis*. New York: Springer-Verlag.
2. Coxon, A. P. M. (1982). *The user's Guide to Multidimensional Scaling* (1st Ed.). Exeter, NH: Heinemann Educational Books.
3. Cunliffe, D. (2000). Developing usable web sites: A review and model. *Internet Research: Electronic Networking Applications and Policy*, *10(4)*, 295-307.
4. D'Angelo, J. & Little, S.K. (1998). Successful web pages: What are they and do they exist? *Information Technology and Libraries*, *17(2),* 71-81.
5. DeLone, W. H., & McLean, E. R. (2004). Measuring e-commerce success: Applying the DeLone and McLean information system success model. *International Journal of Electronic Commerce*, *9(1),* 31-47.
6. Dunn-Rankin, P., Knezek, G., Wallace, S., & Zhang, S. (2004). *Scaling Methods* (2nd Ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
7. Kearney, A.T., (2005). Measuring Globalization. *The Fifth annual A.T. Kearney-Foreign Policy Globalization index,* Foreign Policy, 52-56.
8. Kruskal, J. B., & Wish, M. (1978). *Multidimensional Scaling*. Beverly Hills, CA: Sage Publications.
9. Liu, C., Arnett, K.P., & Litecky, C. (2000). Design quality of web sites for electronic commerce: Fortune 1000 webmasters' evaluations. *Electronic Markets*, *10(2),* 120-129.
10. Mao, J., Vrendenburg, K., Smith, P. W., & Carey, T. (2005). The state of user-centered design practice. *Communications of the ACM*, *48(3),* 105-109.

11. Phippen, A., Sheppard, L., & Furnell, S. (2004). A Practical Evaluation of Web Analytics. *Internet Research*, *14(4),* 284-293.
12. Schubert, P., & Selz, D. (1999). Web assessment--measuring the effectiveness of electronic commerce sites going beyond traditional marketing paradigms. *Proceedings of the 32nd Hawaii International Conference on System Sciences, vol 5*. Retrieved October 18, 2005, from http://csdl2.computer.org/comp/proceedings/hicss/1999/0001/05/00015040.PDF.
13. Shepard, R. N., Rommey, A. K., & Nerlove, S. B. (Eds.) (1972). *Multidimensional scaling: Theory and applications in the behavioral sciences*. New York: Seminar Press.
14. Van der Merwe, R., & Bekker, J. (2003). A framework and methodology for evaluating e-commerce web sites. *Internet Research*, *13*(5), 330-341.
15. Yang, X. Ahmed, Z., Ghingold, M., & Boon, G. S., (2003). Consumer preferences for commercial web site design: An Asia-Pacific perspective. *The Journal of Consumer Marketing*, *20*(1); 10-28.

Appendix 1

Definition of Evaluation Criteria

The interface category has the following sub-categories: graphic design principles, graphics and multimedia, style and text; and flexibility and compatibility.

- Graphic design principles sub-category includes: the effective utilization of backgrounds, color, text, and other general graphic design principles.
- Graphics and multimedia sub-category contains: the value of the graphics and multimedia used on the portal.
- Style and text sub-category incorporates: text consistency and relevance, and the adequacy of the style.
- Flexibility and compatibility sub-category involves: the capability of the interface design to cope with exceptions.
- The navigation category has the following sub-categories: logical structure, ease of use, search engine and help function, and navigational faults.
- Logical structure sub-category includes: the organization and menu of the portal-like logical structure of sections and levels.
- Ease of use sub-category contains: ease to navigate and to find the information for which the costumer is searching.

- Search engine and help function sub-category incorporates: the capability of the search engine to find the correct information without difficulty and supply unambiguous representation of the search outcomes.
- Navigational faults sub-category involves: under construction pages, broken links, and similar items.

The content category has the following sub-categories: product or service related information, company and contact information, information quality, and interactivity.

- Product or service related information sub-category includes: the description of products or services is accurately and systematically displayed.
- Company and contact information sub-category contains: company information contact information.
- Information quality sub-category incorporates: current and relevant information.
- Interactivity sub-category involves: the degree the customer may have access to other users, like the creation of a users' community.

The reliability category has the following sub-categories: stored customer profile, order process, after-order to order receipt, and customer service.

Stored customer profile sub-category includes: registering process and the way business uses the stored client information.

- Order process sub-category contains: effectiveness and simplicity of the online buying procedure.
- After-order to order receipt sub-category incorporates: the process from ordering the product or service until the order is completed.
- Customer service sub-category: the way the e-commerce business communicates and facilitates its online customers.

The technical category has the following sub-categories: speed, security, software and database, and system design.

- Speed sub-category includes: different aspects of the loading process.

- Security sub-category contains: the way the company protects clients' privacy on the portal.
- Software and database sub-category incorporates: when possible, flexibility of the software and telecommunication used.
- System design sub-category involves: the efficiency of the site and how well the site integrates the internal and external aspects of the site.

Chapter 13

IMPORTANT COMPETENCES OF STRATEGISTS AND DECISION MAKERS IN THE STRATEGIC KNOWLEDGE MANAGEMENT MODEL

Roberto Campos Da Rocha Miranda
SQN 115, Bloco C, Apto. 301
Asa Norte
70772-030 – BRASÍLIA – DF
+55 61 3347-7732
bob.fields@uol.com.br

This article aims to present a first approach to characterize strategists and decision makers, in the context of the Strategic Knowledge Management (SKM) model. SKM is a process of creation, capture, assimilation and diffusion of the organizational knowledge. It involves knowledge about planning, description, impact, prediction, evaluation and generation of strategies. This knowledge is formed by strategic and no strategic information, as well as the wisdom accumulated by strategists and decision makers in the process of formulate and make strategic decisions. The methodology applied in this study comprises a research of kinds of skills and competences, focusing activities developed by strategists and decision makers. Thus, it was developed a survey with practitioners in formulate strategies and/or take strategic decision to verify which competences and skills are involved with their jobs.Peliminary results point to different tendencies about which competences and skills are most important to strategists and decision makers. They also show that there is a tendency to emphasize intellectual and knowledge competence in strategic formulation and strategic decision process.

1. Introduction

1.1. *Justifying the Research*

Literature treats strategists and decision makers as important players in company scenario, but rarely their competences and skills are deeply evaluated. In other hand, it is hard to define what characterize their action in a company, particularly in a Strategic Knowledge Management

(SKM) context. Some questions can be formulated considering this problem:

i. How strategists and decision makers think about strategy?
ii. How they formulate them?
iii. Which are the cognitive approaches they take in account to decide?
iv. How they create Strategic Knowledge (SK)? And so on.

Obviously, those questions are too much complex to be answered by one research. It demands many studies to be completely solved. The point here is to establish the main features of each actor (strategist and decision maker), in order to give a contribution on elucidating the most important features and competencies strategists and decision makers must have.

1.2. *Context and Research Question*

The framework considered in this study is the Strategic Knowledge Management Model – SKM (Miranda & Costa, 2005), in which strategists and decision makers are the main players.

Another scope to be considered is that not all features and competencies of strategists and decision makers will be treated. The focus is on the ordinary ones, considering kinds, structures, size, and other variables used to differ organizations.

Hence, the research question can be formulated as follows:
Which are the main competencies and skills of strategists and decision makers in the context of SKM?

2. Theoretical Framework

This topic will present major concepts related with SKM, strategists and decision makers, trying to create a framework to describe features and competences involved with both group of players.

Likewise, some aspects related to competencies, abilities and features of the players mentioned before will be described and they will be used to organize information about strategists and decision makers, heading towards a descriptive players model in the context of SKM.

2.1. ***SKM Model***

The Strategic Knowledge Management model – SKM is "a process of creation, capture, assimilation and diffusion of the organizational knowledge. It involves knowledge about planning, description, impact, prediction, evaluation and generation of strategies" (Miranda & Costa, 2005).

The main features of SKM model can be described as follows:

i. SMK models are formed by three integrated models: a conceptual model, a mathematical model and a system model.
ii. Both, strategists and decision makers are considered in the model, comprising novices and experts.
iii. Strategic knowledge is defined as a kind of organizational knowledge that takes in wisdom concerned with planning, describing, predicting, evaluating, generating and using strategies; thus, SK embraces strategic formulation and strategic decision.
iv. Cognition, context, technology, organizational culture and managerial style are the main SKM systemic factors.

2.2. ***Players in SKM***

Players in SKM refer to the essential players concerned to formulate strategies and take strategic decisions in any kind of company group. They can be found at any corporate level. However, those activities are concentrated in high decision business level, which is why this study will focus only on that professional group.

2.2.1. *Strategist*

According to WordNet (Fellbaum, 2005), strategist is "an expert in strategy (especially in warfare)." In fact, many topics related to strategy are from war, considering the needs to gain victory and overcome a rival, it is necessary to have intellectual instruments to combat. It is similar in the business world. Thus, it is important to have, in any corporation, people worried about stakeholders from internal and external environment. Those people are responsible for "giving a north" to the business, by elaborating strategies to support a competitive advantage in the market. In this paper, strategist is seen as an "organizational professional without decision authority, which uses strategic knowledge to formulate strategies" (Miranda & Costa, 2005).

Those professionals can be classified as follows:

i. Observer: a "scanner" professional responsible for seeking any data or information that can impact an organizational process. A collector works as an "antenna," always looking at internal and external corporate environment, looking for tendencies, conflict points Etc., which is important to a strategic decision.
ii. Collector: a professional responsible for organizing and keeping sources, data and information identified by observers. A collector prepares information for the analysts.
iii. Analyst: a forward planner responsible for analyze data and information and evaluating market tends. He uses this information to infer future conclusions about environment and forces that can influence a corporation. Thus, the generated knowledge is used to formulate alternative strategies to create competitive advantage.

Indeed, this is a particularly view of strategists, considering that more than one feature can be found in one person involved in strategic formulation.

2.2.2. *Strategic Decision Maker*

A Strategic Decision Maker is a formal authority at a corporation responsible for selecting the best strategy for a specific situation. Commonly, a board of directors, guided by information about alternative strategies, takes a strategic decision. Considering that, each and every director, president, and CEO in a corporation is a strategic decision maker. It is important to clarify that those people are on the top corporate level and they have authority to determine the way things are going to do and what goals will be achieved.

2.3. *Competence*

Competence is a term, which can be used in many situations, even if it is considering a corporate environment. For example:

i. "The quality of being adequately or well qualified physically and intellectually" (Fellbaum, 2005).
ii. "Competencies represent clusters of skills, abilities and knowledge needed to perform jobs" (Wikipedia, 2005).
iii. "In human resources, competence is a standardized requirement for an individual to properly perform a specific job" (Wikipedia, 2005).
iv. "(1) Competence concerns the actions and behaviors identified by change agents as contributing in their experience to the perceived effectiveness of change implementation. (2) Competences are those behaviors required for satisfactory ('threshold competence') or excellent ('superior competence') performance in a job" (Prenhall, 2002).
v. "The ability to use knowledge, understanding, practical and thinking skills to perform effectively to the national standards required in employment. This is a broad concept which embodies all aspects of competence relevant to an occupational area, and not just those aspects of the various technical and task components, which are readily observable" (Ntatt, 2004).

In this paper, competence will be considered capacity of judgment, attitude, skill, experience and relationship network, which characterize a co-worker in a corporate field and gives him/her a unique position.

The variables considered in this study will comprise the Sveiby's (1997, 2005), Le Bortef's (1995) and Resende's (2003) view of competence and they will be described in the methodology section.

3. Methodology

3.1. *Method and Technique*

A survey with different kinds of strategists and decision makers was the technique applied to develop the most important part of this research. The main objective was to identify the competences and features that most influence activities related to formulating strategies and to making strategic decision.

The method applied can be described as follows:

i. Bibliographical study mostly about strategic knowledge management and competence.
ii. Definition of each variable related to strategist's and decision maker's competence and skills.
iii. "Survey test" with 3 analysts that had same characteristics of answers, trying to find inconsistencies in answer sheet.
iv. Survey with different kind of strategist and decision maker, using Internet resources to collect data.
v. Consolidation and analysis using MS Excel statistical resources.
vi. Results presentation in order to show the principal competence identified.

It is important to emphasize that each participated person was asked to sort competencies and skills in decreasing order of importance, in order to establish the most and the last important competence/skill to both process: strategic decision or strategic formulation.

3.2. *Population Features*

Population considered to this preliminary research was defined by the followed criteria:

i. 20 people, random chosen among strategists, decision makers, novices and experts.
ii. Different kind of organization, considering private and public ones.
iii. Selected people from organization with wide influence in all states of Brazil.

3.3. *Variables*

Variables are clustered by kind of topic analyzed, and each one is defined according to the main research goals. Thus, variables can be described and defined as follows:

i. Demographic Data
    a) Sex, male or female.
    b) Age, indicates the respondent chronological age, varying from 20-years-old to upper, with cluster of five years.

ii. Professional Data
   a) Kind of professional, strategist, strategic decision maker, both or none.
   b) Level of expertise, Novice – less than 10 years performing strategic activities or Expert – more than 10 years performing strategic activities.

iii. Corporation Data, with those data was possible to classify a corporation as a micro, small, medium or big one
   a) Number of Employees, related to number of occupied persons in a corporation.
   b) Annual Income, related to total company balance sheet.
   c) Core Business.

iv. Competence Date
   a) Kinds of Competence (just called competence in analysis topic)
      a. Information Competence, related to the ability to generate, transmit and publish information to coworkers or stakeholders.
      b. Knowledge Competence, related to the ability to add value to information, by selecting, multiplying, associating and applying information.
      c. Technical Competence, related to ability to develop operational competence, which is often acquired by formal education.
      d. Intellectual Competence, related to ability to develop cognitive features as creativity, memory, innovation and reasoning.
      e. Emotional Competence, related to ability to control aggressiveness, excitement, expectations and ambition.
   b) Skills
      a. Attitude, related to the position adopted for a person as politeness, meticulousness etc.
      b. Capacity, related to special abilities as hand making, abstract and special reasoning etc.
      c. Expertise, related to knowledge and skills gained from both/either training and/or experience, i.e., practice wisdom.
      d. Formal Knowledge, related to technical information and knowledge acquired by formal education.
      e. Judiciousness, related to tacit factors that affects any decision and they can be seen as "sense making".
      f. Network Relationship, related to the influence of relationship among relatives, friends, suppliers, etc. that can affect formulation or decision strategic activities.

4. Findings and Discussion

Preliminary conclusions of this on going research are based on sample of survey responses and it will be showed as:

i. Description of the sample.
ii. Evaluation of competences.
iii. Evaluation of Skills.

4.1. *Description of the Sample*

The most important features of the sample considered can be expressed as follows:

i. 90 percent of answers are male and they are from 31 to 55 years old, mostly between 41 and 45 years old.
ii. 45 percent of the organizations are governmental ones, including banks and postal office, all of them with annual incoming over R$60 million (about USD28 million, what characterize big companies in Brazil). A telecommunication corporation and a private school are included among private organizations.
iii. 74 percent of corporations considered have over 250 employees.
iv. Distribution of strategists and decision makers considered in terms of kind of organization can be seen in Table 1.

Table 1. Kind of organization X kind of professional.

		Kind of Professional				
		Strategist	Both(*)	Decision Maker	None	Total
Kind of Organization	Private	1	3	2		6
	Governmental	3	2	2	2(**)	9
	Other	1	2	1		4
	TOTAL	5	7	5	2	19 (***)

(*) Strategist and Decision Maker.
(**) Although these answers are not included as strategist or decision maker, both have experience in strategic matters.
(***) One of respondents did not mark his/her condition in the answer sheet.

v. 65 percent of the sample is composed by people with most than 10 years in activities related to formulating strategies and strategic decision making, as shown in Table 2.

Table 2. Kind of organization X expertise.

Kind of Professional		Expertise: Expert	Novice	Total
	Private	2	5	7
	Governmental	5	4	9
	Other		4	4
	TOTAL	7	13	20

vi. Finally, mostly of novices act both as strategists and decision makers, as shown in Table 3.

Table 3. Kind of professional X expertise.

Kind of Professional		Expertise: Expert	Novice	Total
	Strategist	2	3	5
	Both(*)	2	5	7
	Decision Maker	3	2	5
	None		2	2
	TOTAL	7	12	19

(*) Strategist and Decision Maker.

4.2. *Evaluation of Competences*

In order to identify interesting tendencies related to the most important competencies that interfere in strategic formulation process and strategic decision process, the most frequent answer to each topic competence was considered.

In fact, analysis considered two groups of answers:

i. importance of competence versus kind of professional, and
ii. importance of competence versus expertise.

4.2.1. *Importance of Competence Versus Kind of Professional*

The importance of each competence to strategic formulation process is different as can be seen in Table 4.

Table 4. Importance of competence versus kind of professional, considering strategic formulation process.

	Kind of Professional			
	Strategist	Both(*)	Decision Maker	General
Most (+)	Intellectual Competence	Knowledge Competence	Knowledge Competence	Knowledge Competence
		Intellectual Competence		Intellectual Competence
Importance	Knowledge Competence	Information Competence	Intellectual Competence	Information Competence
	Technical Competence			Technical Competence
	Information Competence			
	Emotional Competence	Technical Competence		Emotional Competence
Less (-)		Emotional Competence		

(*) Strategist and Decision Maker.

It seems to have a convergence among opinions of strategists and decision makers in terms of define the most important competences to formulate strategies: knowledge competence and intellectual competence. Those competences tend to be related to “thinking processes”, which is an important feature of the strategic formulation activity. This tendency could confirm the importance of knowledge management in strategic context.

Once more, the importance of each competence to strategic decision process is different as can be seen in Table 5.

There is a tendency to emphasize information competence and knowledge competence to decide strategically. Really, both competences are complimentary: information related to manage data inside and outside organization environment and knowledge related to the process of using information and exchanging ideas about possibilities and alternatives to a way of action.

It is interesting to observe that technical competence is the less competence considered important to strategic decision, while emotional competence is the less one to strategic formulation.

Table 5. Importance of competence versus kind of professional, considering strategic decision process.

	Kind of Professional			
	Strategist	Both(*)	Decision Maker	General
Most (+)	Knowledge Competence	Intellectual Competence	Knowledge Competence	Information Competence
Importance	Intellectual Competence	Information Competence	Information Competence	Knowledge Competence Intellectual Competence
	Emotional Competence	Knowledge Competence	Technical Competence	Emotional Competence
			Emotional Competence	
	Information Competence	Emotional Competence	Intellectual Competence	
Less (-)	Technical Competence	Technical Competence		Technical Competence

(*) Strategist and Decision Maker.

4.2.2. *Importance of Competence Versus Expertise*

The importance of each competence to both strategic formulation and strategic decision process is shown in Tables 6 and 7.

Table 6. Importance of competence versus expertise, considering strategic formulation process.

	Expertise		
	Novice	Expert	General
Most (+)	Knowledge Competence Intellectual Competence	Knowledge Competence Technical Competence	Knowledge Competence Intellectual Competence
Importance	Information Competence	Intellectual Competence	Information Competence
		Information Competence	Technical Competence
Less (-)	Technical Competence Emotional Competence	Emotional Competence	Emotional Competence

The importance of each competence to strategic decision process is different as can be seen in Table 7.

Table 7. Importance of competence versus expertise, considering strategic decision process.

	Expertise		
	Novice	Expert	General
Most (+)	Information Competence Knowledge Competence	Information Competence	Information Competence
Importance		Knowledge Competence Intellectual Competence	Knowledge Competence Intellectual Competence
	Intellectual Competence	Emotional Competence	Emotional Competence
	Emotional Competence		
Less (-)	Technical Competence	Technical Competence	Technical Competence

Likewise, results observed in Tables 4 and 5 are confirmed by novices and experts.

4.3. *Evaluation of Skills*

Similarly to the procedures adapted to present kinds of competence, skills will be shown in terms of kind of professional and expertise.

4.3.1. *Importance of Skill Versus Kind of Professional*

The importance of each skill to strategic formulation process is different as can be seen in Table 8.

Table 8. Importance of skill versus kind of professional, considering strategic formulation process.

	Kind of Professional			
	Strategist	Both(*)	Decision Maker	General
Most (+)	Judiciousness			
Importance	Expertise	Capacity		Capacity
	Formal Knowledge	Expertise		Expertise
	Capacity	Judiciousness		Formal Knowledge
				Judiciousness
	Attitude	Formal Knowledge		Attitude
	Network Relationship	Attitude		Network Relationship
Less (-)		Network Relationship		

(*) Strategist and Decision Maker.

As can be seen, in terms of skills, there is not a consensus about which competences are the most important to strategic formulation process. Strategists consider judiciousness and expertise the most important skills, but there were no tendencies observed by decision makers. Maybe it is because decision makers tend to valorize different skills to each decision situation.

On the other hand, when the subject is strategic decision process there is a clearer tendency in terms of the most important skill, as shown in Table 9.

Table 9. Importance of competence versus kind of professional, considering strategic decision process.

	Kind of Professional			
	Strategist	Both(*)	Decision Maker	General
Most (+)	Judiciousness		Formal Knowledge	Judiciousness
Importance	Expertise	Attitude/Capacity	Expertise	Expertise
	Network Relationship	Expertise Judiciousness	Judiciousness	Network Relationship
	Formal Knowledge		Capacity	Formal Knowledge Capacity
	Capacity Attitude	Formal Knowledge Network Relationship	Attitude Network Relationship	Attitude
Less (-)				

(*) Strategist and Decision Maker.

4.3.2. *Importance of Skill Versus Expertise*

Other interesting result refers to perception of which skills are most important to strategic decision process, showed in Table 10.

Although novices and experts consider judiciousness a very important skill to strategic formulation process, novices consider capacity the most important and experts, expertise. It seems to be coherent because novices tend to emphasize improvement and innovation and experts tend to focus on their own background, giving more importance to all the know-how achieved during their career.

Finally, novices and experts confirm results described in Table 9: judiciousness is the most important skill to strategic decision process. On the other hand, attitude tends to be less important to the process, as shown in Table 11.

Table 10. Importance of skill versus expertise, considering strategic formulation process.

	Expertise		
	Novice	Expert	General
Most (+)			
	Capacity	Expertise	Capacity
	Judiciousness	Judiciousness	
Importance	Expertise		Expertise
	Formal Knowledge	Formal Knowledge	Formal Knowledge
		Network Relationship	Judiciousness
	Attitude	Capacity	Attitude
Less (-)	Network Relationship	Attitude	Network Relationship

Table 11. Importance of skill versus expertise, considering strategic decision process.

	Expertise		
	Novice	Expert	General
Most (+)	Judiciousness	Judiciousness	Judiciousness
	Capacity	Expertise	Expertise
	Network Relationship	Network Relationship	Network Relationship
Importance	Formal Knowledge	Formal Knowledge	Formal Knowledge
	Expertise	Capacity	Capacity
	Attitude	Attitude	Attitude
Less (-)			

5. Conclusions

Preliminary results of this on-going research lead to interesting points, as follows:

i. There are different perceptions between strategists and decision makers, as well between novices and experts, related to main competences and skills involved with strategic decision and strategic formulation process.

ii. Strategists tend to consider intellectual competence, knowledge competence and judiciousness the most competences and skills to formulate strategies and take strategic decisions.

iii. Emotional competence, technical competence, attitude and network relationship tend to be the less important competencies and skills in SKM context.

At this point, there are some questions that might be formulated based on these results:

i. Is there some kind of relation between competences/skills and corporation size?
ii. Do strategists from big corporations think different of micro and small ones, in terms of strategic decision and strategic formulation process?
iii. When a person acts as strategist and decision maker simultaneously, which is often saw in small corporations, which competences and skills are most considered?

Those doubts open new fields of study in Strategic Knowledge Management context.

Overall, in order to congregate all primary results, Figures 1 and 2 shows principal competences and skills involved in SKM context, separated by kind of professional and expertise.

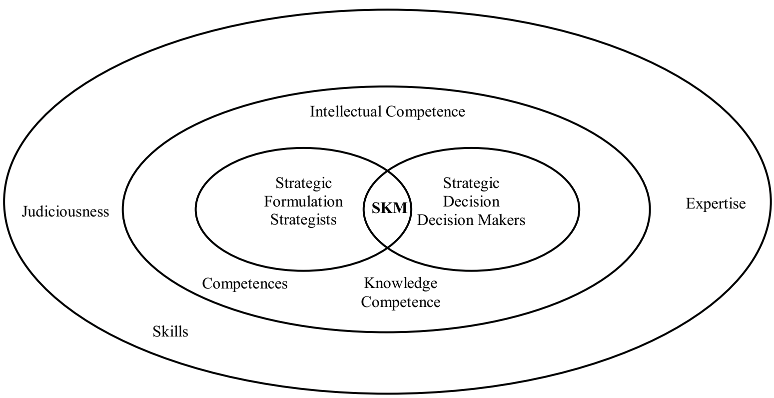


Figure 1. Most important competences and skills to SKM, according to strategists and decision makers.

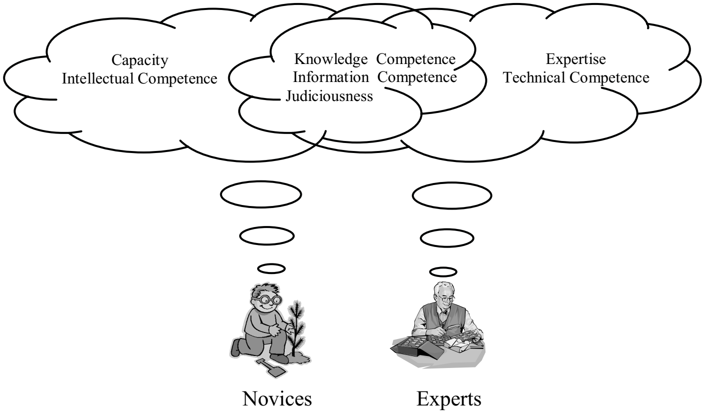


Figure 2. Most important competences and skills to SKM, according to novices and experts.

References

1. Fellbaum, C. (Ed.) (2005). Cognitive science laboratory. *WordNet* [Eletronic version]. Retrieved November 20, 2005, from http://wordnet.princeton.edu/perl/webwn.
2. Le Boterf, G. (1995). De la compétence – essai sur un attracteur étrange. In: *Les éditions d'organisations*. Paris: Quatrième Tirage.
3. Miranda, R.C.R. and Costa, S. M. S. (2005). Strategic knowledge management: a conceptual framework", In: Hawamdeh, S. (Ed.) *Proceedings of the 2005 International Conference on Knowledge Management—Nurturing Culture, Innovation and Technology* (pp. 417-427). Charlotte, NC: World Scientific.
4. Ntatt (2004). *Glossary of Terms*. Retrieved November 20, 2005, from National Training Agency Access http://www.ntatt.org/glossary.html.
5. Prenhall (2002). *Glossary*. Retrieved November 20, 2005, from Pearson Education Access: http://wps.prenhall.com/wps/media/objects/213/218150/glossary.html.
6. Resende, E. (2003). O livro das competências: Desenvolvimento das Competências: a melhor Auto-Ajuda para Pessoas, Organizações e Sociedade. Rio de Janeiro: Editora Qualimark.
7. Sveiby, K. E. (1997). The intangible assets monitor. *Journal of Human Resource Costing and Accounting,* 2(1), 73-97.
8. Sveiby, K. E. (2005). *Methods for Measuring Intangible Assets*, Retrieved November 22, 2005, from http://www.sveiby.com/articles/IntangibleMethods.htm.
9. Wikipedia (2005). *Competence.* Retrieved November 20, 2005, from http://en.wikipedia.org/wiki/Competence.

Chapter 14

MANAGING E-RECORDS IN NORTH AMERICAN COLLEGES AND UNIVERSITIES: AN OVERVIEW OF CURRENT PRACTICES

Lisl Zach
School of Library and Information Science
Louisiana State University
267 Coates Hall, Baton Rouge, LA 70803-3920
lzach@lsu.edu

This paper describes the goals, methods, and preliminary findings of an on-going research project funded by the U.S. National Archives and Records Administration's National Historical Publications and Records Commission (NHPRC) to investigate current practices among North American college and university archives and records management programs. The project used an on-line survey and follow-up interviews to document approaches to capturing, storing, organizing, and making available a broad range of materials such as administrative records, digital assets, email, and institutional publications and websites. The findings indicate that the field is less advanced in its efforts than many practitioners, particularly at small- and medium-sized academic institutions, have believed. For many the key to introducing even a rudimentary e-records management program is to focus on a single, often highly visible area and to modify existing policies and procedures to encompass e-records.

1. Introduction

In October 2005, the U.S. National Archives and Records Administration's National Historical Publications and Records Commission (NHPRC) provided funding for a research project to determine what patterns, if any, exist in current practices among North American college and university archives and records management programs regarding their approaches to capturing, storing, organizing, and making available institutional e-records. For the purpose of this project, the term 'e-records' covers a broad range of materials such as administrative records, digital assets, email, and institutional publications and websites. In addition, this project is investigating what categories of 'best practices' would be most useful to archivists and records managers in meeting institutional needs for sharing permanently valuable content and leveraging university resources to capture, store, organize, and deliver e-records.

More and more university records, of all types, have gone digital in recent years. This means that, by in large, records are not being archived. Most transaction-processing and information systems used by colleges and universities are not designed for permanent record keeping. For example, in some human resource systems, the capability of creating a print record is simply not available without months of modification. Therefore, users do not print out and retain hard copies. In addition, many campus publications that were once printed and retained are now being mounted on the Web as part of an active page that is updated routinely or produced in a PDF format that is replaced periodically. In neither case is a copy always captured for permanent storage. These and other campus records management issues result in a loss of institutional history and, in the case of some state institutions, a loss of state records.

To date, much of the theoretical research that has been done within the archival community on e-records has been to identify the data or functional requirements for metadata systems needed to support archival requirements, to define what constitutes an authentic and reliable record, and to determine the requirements for trustworthy recordkeeping systems. Significant projects such as the standards and guidance for keeping e-records developed by the National Archives of Australia and the work done by the Canadian National Data Archives Consultation, which has studied issues surrounding the preservation of research data, have helped to define the problems faced by archivists and records managers (Swan, Cunningham, & Robertson, 2002; Humphrey, 2005). However, while this research lays the essential foundation for the development of answers to the problems of long-term preservation of digital resources, it is still far from providing practitioners in the trenches with a practical solution for capturing, storing, organizing, and making available institutional e-records.

With some notable exceptions, such as the descriptions of the e-records management programs at Indiana University and the University of Michigan (Bantin, 1998; Bantin, 1999; Gilliland-Swetland, 1996), most examples of actual implementations have come from outside the United States and from outside of the academic environment (Smyth, 2005; Gregory, 2005; Johnston & Bowen, 2005; Maguire, 2005). A 2002 study of 15 U.S. and 15 Canadian universities identified a significant need for "institution-wide electronic records management policies and procedures developed in cooperation with senior administrators, information technology staff, university archivists, and records managers" (Schina & Wells, 2002). The on-going NHPRC research project examines current e-record management practices and addresses the extent to which this need has been met.

2. Approach

The study was divided into two distinct parts, including a survey phase and an interview phase. In November 2005, a 22-question on-line survey was sent to 638 archivists and records managers identified from the membership of the College and University Archives Section of the Society of American Archivists (SAA). The survey was pre-tested with 10 practitioners. Data were collected and analyzed using the tools provided by the SurveyMonkey software; responses to open-ended questions were analyzed manually. Data collected in the survey phase of the study were used to identify a baseline of current practices in the field and to solicit participants for the interview phase of the study.

Participants for the second phase of the study, conducted in spring and summer 2006, were chosen from those individuals who indicated on the survey that they/their institutions are actively implementing or planning an e-records program. Other selection criteria included: size/type of institution, type of program being implemented or planned, geographic location, and scheduling constraints. Use of selection criteria such as these is consistent with theoretical and/or purposive sampling techniques. Fifteen practitioners were identified for in-person and phone interviews.

The interview phase of the study was carried out using a multiple-case studies design (Zach, 2006). The multiple-case studies design allowed the researchers to use subsequent cases/interviews to confirm or disprove the patterns identified in earlier research. In practice, this means that the interview protocol may be revised during the course of the study in response to new information. The interviews were sequenced so that the first group of five interviewees had many characteristics (size/type of institution, type of program, etc.) in common (*literal replication*); these interviews provided (as far as possible) a baseline of current practices and identified priorities. The remaining ten cases/interviews were selected to explore and confirm or disprove the patterns identified in the initial interviews (*theoretical replication*)

3. Preliminary Results

Of the 638 requests for participation that were sent out, 57 went to bad email addresses, 40 people declined to participate, and 224 have responded to date. After the initial data collection period (including two follow-up requests), the researchers targeted specific individuals at key institutions on the list and contacted them by telephone. In some cases, the person who had originally received the survey then completed and returned it; in other case, he or she referred the researchers to a more appropriate person at the same institution. Of the remaining 317 who have not responded, many are from institutions with two or more members of College and University Archives section, where one member has already responded. In total, 418 institutions were represented in the sample pool; 193 institutions provided at least one response to the survey. Of the responses received, only 15 said that their institutions have a formal e-records management program. Another 81 said that their institutions have a program in the planning stages. While these results are not a surprise, they are a sad commentary on the state of managing e-records on college and university campuses.

Based on the data collected in the survey, it appears that email and administrative records have been the priorities among institutions that have formal e-records management programs in place. However, as shown in Figure 1, below, those institutions that are in the planning stages of an e-records management program are putting their attention primarily in the areas of websites and digital assets.

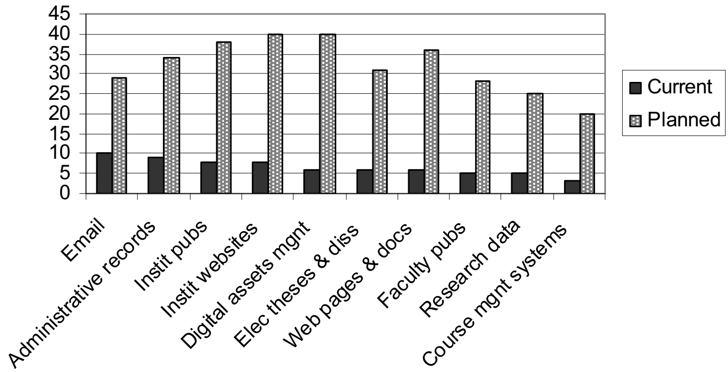


Figure 1. Priority areas for E-records management programs.

Although these results are too few to be more than suggestive about institutional priorities, the responses to the question about areas in which archivists/records managers would like to see addition 'best practice' guidance material supports the same conclusions (see Figure 2.).

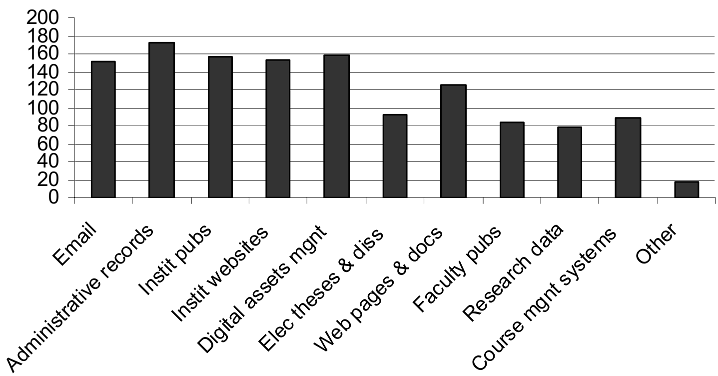


Figure 2. Areas in which 'best practice' guidance is desired.

These responses suggest that while email continues to be an area of concern to archivists and records managers, the primary focus of attention is more institutional, with administrative records being cited by almost 80 percent of the respondents as their top priority, followed by digital assets and institutional publications.

An encouraging finding from the initial survey is that most of the archivists/records managers at institutions with formal e-records management programs, either in place or in the planning stages, are involving stakeholders from other units on campus in the development of these programs. This begins to address the need for intra-institutional cooperation identified in the 2002 survey. Of those responding to the question about stakeholders, 60.5 percent said that the campus attorney had been or was involved in the development of policies. Sixty and one-half percent also reported the involvement of data managers and/or the CIO. Respondents also identified other stakeholders, including provosts, senior academic VPs, external board members, risk managers and auditors. The exact nature of the involvement of these various stakeholders was investigated in depth during the follow-up interviews

Of the 224 individual respondents to the survey, 212 (96 percent) work in institutions that have a designated archivist. Of these, 137 report to the director or dean of the library, 35 report to the director or head of special collections, and 38 report else where on campus, including to the provost, college president, associate vice-president for legal affairs, the associate vice-president for information technology, and the university secretariat. In general, the reporting structure indicates a level of recognition for the importance of the archival function on college and university campuses. However, only 78 of the respondents work in institutions with a designated records manager, and there does not appear to be a strong relationship between institutions that have a records manager and those that report having a records management program in place or in the planning stages. Of the 96 respondents with programs in place or in the planning stages, all but five have a designated archivist, but only 41 (43 percent) reported having a records manager. The majority of respondents feel that the records management function, where it is not mandated by legal requirements, is done unsystematically on a department-by-department basis. This is consistent with the findings of the 2002 survey and indicates the continued need for the development and implementation of formal, institution-wide policies and procedures.

The follow-up interviews with those agreeing to discuss their programs confirmed that few if any institutions have a fully developed e-records management program; most are only in the very early planning stages. In those institutions that do have some e-records management initiatives underway, those initiatives deal with only specific areas, e.g., administrative records, e-mail, institutional publications, etc. rather than providing an overall approach. Even at the best-funded and most prestigious institutions, there are no existing programs to use as models for the field. Further discussions with the interview participants determined that support at the institutional level, e.g. provost, president, board of trustees, is essential for an effective overall e-records management program; even piece-meal development requires some high level support to make any real progress. In addition, the interviews showed that successful archivists/records managers have made strategic alliances with key players outside of the library, such as the CIO and legal counsel.

4. Discussion

The preliminary results from the project do not, in themselves, represent significant new findings regarding the current state of college and university e-records management programs. Rather they confirm, from a much broader sample base, the anecdotal information that has been heard from practitioners over the last several years. The strength of the results lies in the areas they suggest as priorities for increased attention. The results of the interviews indicate that the field is less advanced in its efforts than many practitioners, particularly at small and medium sized academic institutions, have believed. Archivists and records managers find themselves in understaffed and under-funded environments. They cannot afford to be pioneers. For many, the key to introducing even a rudimentary e-records management program is to focus on a single, often highly visible, area and to modify existing policies and procedures to encompass e-records. To do even this, archivists and records managers must learn to develop strategic alliances with key players in other parts of the college/university system and to become publicists for the importance of e-records management as part of the larger institutional agenda.

References

1. Bantin, P. (1998). Developing a strategy for managing electronic records--the findings of the Indiana University Electronic Records Project. *The American Archivist*, 61(Fall), 328-364.
2. Bantin, P. C. (1999). The Indiana University Electronic Records Project revisited. *The American Archivist*, 62(Spring), 153-163.
3. Gilliland-Swetland. (1996). *Policies and politics: the archival implications of digital communications and culture at the University of Michigan*. A case study. (Report No. PC 307). The Society of American Archivists.
4. Gregory, K. (2005). Implementing an electronic records management system: a public sector case study. *Records Management Journal*, 15(2), 80-85.
5. Humphrey, C. (2005). *Preserving research data: the Canadian experience*. Unpublished paper presented by at ECURE 2005. Tempe, Arizona: Arizona State University. Retrieved May 25, 2005, from http://www.asu.edu/ecure/2005/humphrey/HumphreyCharles_bio.html.
6. Johnston, G. P., & Bowen, D. V. (2005). The benefits of electronic records management systems: a general review of published and some unpublished cases. *Records Management Journal*, 15(3), 131-140.
7. Maguire, R. (2005). Lessons learned from implementing an electronic records management system. *Records Management Journal*, 15(3), 150-157.
8. Schina, B., & Wells, G. (2002). University archives and records programs in the United States and Canada. *Archival Issues*, 27(1), 35-51.
9. Smyth, Z. A. (2005). Implementing EDRM: has it provided the benefits expected? *Records Management Journal*, 15(3), 141-149.
10. Swan, K., Cunningham, A., & Robertson, A. (2002). Establishing a high standard for electronic records management within the Australian public sector. *Records Management Journal*, 12(3), 79-86.
11. Zach, L. (2006). Using a Multiple-Case-Studies Design to Investigate the Information-Seeking Behavior of Arts Administrators. *Library Trends*, 55(1), 4-21.

Chapter 15

A STUDY ON MEASUREMENT OF INTELLECTUAL CAPITAL KERNEL BASED ON ARTIFICIAL NEURAL NETWORK AND EMPIRICAL ANALYSIS

Huang Ningyan
Institute of Scientific & Technique Information of China
15 Fuxing Road, Beijing 100038, China
0086-10-5888.2561
flowerofmay11@wanadoo.fr

Shuai Liguo
University of South-East
Nanjing 210096, China
0086-13813893943
liguo.shuai@gmail.com

Aiming at the measurement of "intellectual capital kernel", this paper proposes to construct a mathematic model with the artificial neural network (ANN) and demonstrate how the executive process of intellectual capital resource can be simulated by the ANN. This model will show the mathematical relationship between intellectual capital carrier instructors and firm performance. BP (Back propagation) network--considered classic and simple--is taken for granted in the case study of Sweden Skandia in order to measure intellectual capital kernels. The empirical study is done based on data from the Skandia intellectual capital report. A deep analysis is made for the results, which show that for its limited samples, the model operates normally. Additional case studies are anticipated to further validate our conclusions.

1. Introduction

Our present concept of economy is developing into the naming of a new era of commerce, namely the knowledge economy. With the rise of this "new economy", knowledge became the firm's most valuable resource (Drucker, 1993). Information and knowledge have become the most important resources in an organization, as well as the source for obtaining the predominance of sustainable competitiveness. Accepting this premise suggests that knowledge can be transferred, combined, and used (Grant, 1996), and it may be a potential source of sustainable competitive advantage (Nonaka & Takeuchi, 1995).

With the continuous increase of obtainable knowledge in organizations, the demand for managing and deploying knowledge effectively rises. Firms create value, combining different types of resources (tangibles and intangibles) and competences, and that value increases as much as those resources and competences interact (Cho & Bontis, 2002). Value creation resides at the very heart of strategic management literature, and it is the primary rationale of intellectual capital (Edvinsson & Sullivan, 1996). However, it should be recognized that management and deployment of knowledge is not an easy task. Difficulties include recognizing that tacit knowledge accounts for much value in the organization and that if a person essential to the organization leaves, knowledge will generally leave with him, among others. In face of these issues, Knowledge Management (KM) and Intellectual Capital (IC) management have developed during the last decade in an attempt to formulate new resolutions for improving organization adaptability, organization survival, and competence management.

Intellectual Capital is yet a concept hard to define. What gets measured gets managed (Seetharaman, 2002）. In order fully understand the concept of IC, it is necessary to grasp the status and the change of IC, as well as the relationship between IC and other phenomenon, by means of measurement. Although none of the current definitions for IC is universally accepted, since the mid-1990s, organizations have initiated the practice of IC measurement and not limited its concept to discussion and theory system. Beginning in 1994, Skandia--the pioneer in the practice of IC measurement--published its IC report as a supplement to their traditional annual report. Influenced by Skandia, a number of well-recognized organizations have since implemented the practice and reporting of IC measurement. Dow Chemical, for example, has been at the forefront in using patents as proxies for practical IC measurement. In 1996, Dow Chemical released its first public report as a supplement to its annual report (Bontis, 2001). Austria Research Center Seibersdorf (ARCS) adopted the concept of IC and has applied its model for measurement in the evaluation of the research institution, publicizing its IC report since 1999 (ARCS, 1999, 2000, 2001, 2002). German Aerospace Research Center and German Space Agency (DLR) probed into IC measurement as well.

2. Problem

The IC research field is beset with numerous difficulties, particularly the measurement of IC. A number of solutions have been suggested, but unfortunately, none are considered effective. Sveiby (2004) listed 28 methods for measuring intangibles, which are, in fact, just a part of present measurement methods for IC. The value of Sveiby's list lies in that it has contributed to most efforts creating the tools and methods that aim at the measurement of IC or intangibles.

Although not universally defined, the conceot of IC is often recognized as "knowledge that can be converted into value" (Edvinsson & Sullivan, 1996). In order to construct a base for our study, a definition of IC and a new point of view for IC are given as follows. "Intellectual Capital is the collection of intangible resources relating to human's knowledge or intelligence, which influence or have a potential influence to the organization performances or the innovation competences." (Huang, 2005)

Based on this definition, IC is divided into two layers--"IC carrier" and "IC kernel". (See Figure 1.) IC carrier is the "shell" of IC, represented by human resources, R&D input, R&D activities, databases, organizational structures, and other forms we can touch and which bear knowledge. IC Carriers then, are a part of IC that we can easily identify. It bears noting that IC carrier is for our purposes not absolutely IC per se. In fact, researchers became interested in knowing IC carrier first, when getting acquaintance with IC. During the last decade of IC study, researchers and practitioners generally believe that they measured IC, per se. We suggest in reality, these measurements merely reflect IC carriers. This misunderstanding comes from the lack of deep understanding of IC. IC kernel is the "core" of IC. According to our definition of IC, it is essential to understand that IC kernel is that which "influence or have a potential influence to the organization performances or the innovation competences". Therefore, IC kernel can be the function or the effect of resource elements on performance, especially innovative performance. Certainly, IC kernel can be expressed in other forms, and this is the basis for our exploration: that IC can be understood in terms of layers. In our research, we propose a special division for IC, not one defined as a new category, as such, but one subject to the disassembly of IC by layers. Our aim is to understand, express, and grasp the connotation of IC.

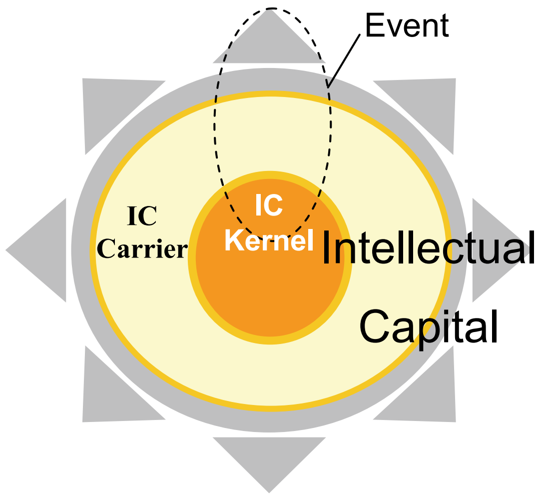


Figure 1. The division of intellectual capital: IC carrier and IC kernel.

IC carrier can be one of two types, namely persons or objects which carry knowledge and competence, e.g. personnel, organization infrastructure, and activities or performances which demonstrate the value of IC, e.g. projects, patents, papers, high-tech products, etc). Compared with IC carrier, IC kernel is a relatively abstract concept, one intangible by nature and with no entity. Thus, it is difficult to specify IC kernel; however, that which it exposes is the profound connotation of IC, namely that which IC researchers and practitioners seek with great effort. IC kernel, therefore, is no doubt essential for IC researches and practitioners.

In practice, IC kernel measurement is a new field to dig and mine. This so, limited research on IC kernel measurement has already appeared. However, none indicates the aim of their model is IC kernel. Bontis (1998) had already proved and shown in a pilot study that there is a strong and positive relationship between Likert-type measures of IC and business performance. The explanatory power of the final specified model was highly significant and substantive (R^2=56.0%, p-value<0.001). Cabrita (2005) did an empirical study on the industry of Portuguese banks with the method of Partial Least Squares (PLS). Andreou (2005) made use of the same method in his empirical study. These efforts are valuable, but far from sufficient. More research and practice is needed at this level.

3. Methodology

IC cannot exist without IC carrier. We believe that IC carrier leads the way to knowing IC kernel, and that there is certainly relevancy between IC Carriers and I". Our goal is to obtain IC kernel by making use of the instructors of IC carrier and those of performance. To realize this, we introduce the method of artificial neural network (ANN).

3.1. *Significance of Introducing ANN*

Both cerebra and firms can be deemed organizations. It can be said that the extreme level (utmost limit) for organization development is to simulate the mode of cerebra. The cerebra are self-organized, self-learning (Qiu, 2001). The popularity of theories like agile manufacture and learning organization embodies the desire of organization to enforce the adaptation and existence as a learning organization. One of the tasks for IC measurement is to establish the mathematical relationship, for instance, of how IC factors influence the innovative performance and between which the IC factors and the elements of management concern. However, most often an organization is regarded as a sort of "black box", in which questions like these are difficult to answer.

3.2. *ANN*

Artificial Neural Network (ANN) is a mathematical tool for studying the relationship of cause and effect. ANN, generally, is superior for studying complicated questions, or for those in which the data not integrated or which cannot be well resolved by commonly used methods. Particularly when the interior of a problem is not well known and when it is difficult to judge the type of relation among factors, it is reasonable to choose ANN for research. Thus, we selected ANN for IC kernel measurement.

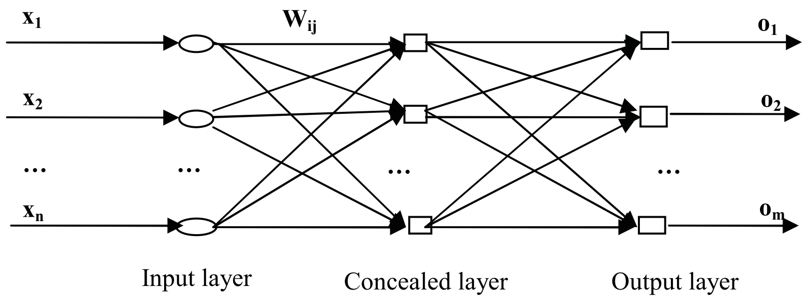


Figure 2. The structure of artificial neural network.

4. Model

A model for IC kernel measurement based on ANN is herein proposed. The rationalization comes from the comparison between organizational management process and the characteristics of ANN.

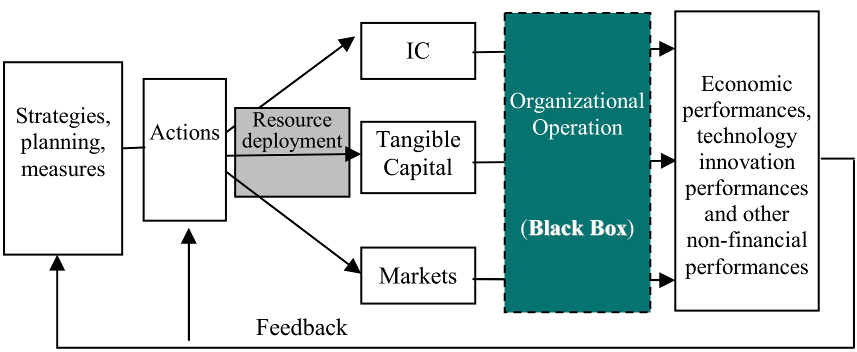


Figure 3. Process of resource circulation for an organization.

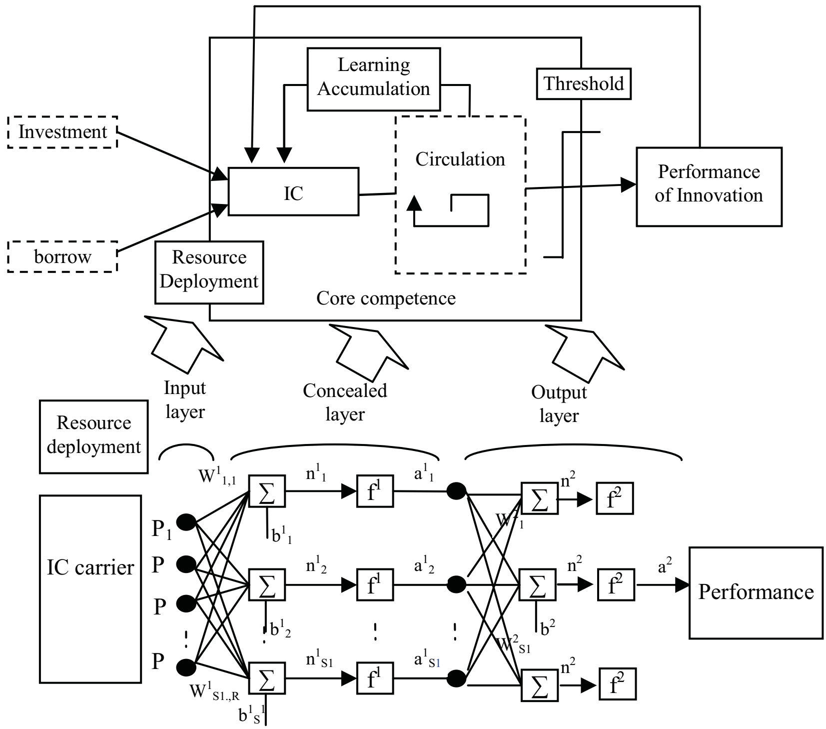


Figure 4. The simulation of IC process and ANN.

The process of management can be regarded as that of resource operation. (See Figure 3.) Generally, the process of management starts at the strategy, planning or measures. Then firms take action according to management decisions to deploy necessary resources. These resources include three principal types, namely tangible capital, IC, and market resources. In the industrial era, material capital and monetary capital were rare resources. For firms with a correspondent level in management and technology resource, the core competence lies in obtaining the capital needed for operation production. If the level of production is correspondent, marketing becomes the core competence of enterprise. Today, the quantity and quality of IC become the new core competence which draws people's attention, especially for knowledge-based firms.

What interests us is the resource of IC, so that we can extract and draw the IC process in Figure 4. (Note the upper side of Figure 4.)

Generally, there is a threshold for the input of resource to obtain a result in performance. In other words, if the quantity of input resources are inferior to the threshold, the process of management will not be influenced, that is to say, a result of performance may be achieved.

Through comparing the upside and the downside of Figure 4, we found that the IC process strongly resembles ANN. The input level relates to the deployment of IC resource; the concealed level refers to the "black box" of organization; and the output level analogizes organizational performance. Therefore, ANN model is reasonable for the simulation of IC process.

4.1. *An Empirical Study Based on BP Network*

4.1.1. *Data Collection*

Only two establishments, namely Skandia of Sweden and Austria Research Center Seibersdorf (ARCS), have published IC reports during the past four years, [15] [3] In terms of our definition of IC, both reports relate to IC carrier. Skandia had publicized its IC report as a supplement to its annual report for five years beginning in 1994. ARCS had publicized its IC report for four years, as well. Therefore, we decided to utilize the data of Skandia in order to demonstrate of our ANN model for IC measurement. It is important to note that since 1999, Skandia has unfortunately changed its policy for publicizing IC measurement. Rather, Environmental and other factors have become the subject their supplementary report. This said, we utilize the data from reports for 1994 to 1998. (See Table 1.)

The IC model of Skandia is named "navigator", and is based on the notion that it is an instrument that navigates us into the future and inspires our renewal and development (Edvinsson and Malone, 1997). The navigator includes five focuses, namely finance, custom, human process, renewal, and development.

The data are reported according to these five focuses in the supplement of the annual report. In our study, the instructors of data are treated for the ANN simulation according to the following rules: the

Table 1. The supplementary reports of Skandia since 1994.

Year	Supplementary annual report of Skandia
1994	• Visualizing Intellectual Capital
1995	• Renewal and Development • Value-Creating Processes
1996	• Customer Value • Power of Innovation
1997	• Intelligent enterprising
1998	• Human capital transformation
1999	• Environmental Report 1999
2000	• Environmental Report 2000
2001	• US GAAP American Skandia Inc • Environmental report 2001
2003	• Memo Compensation Matters

Source: http://www.skandia.com.

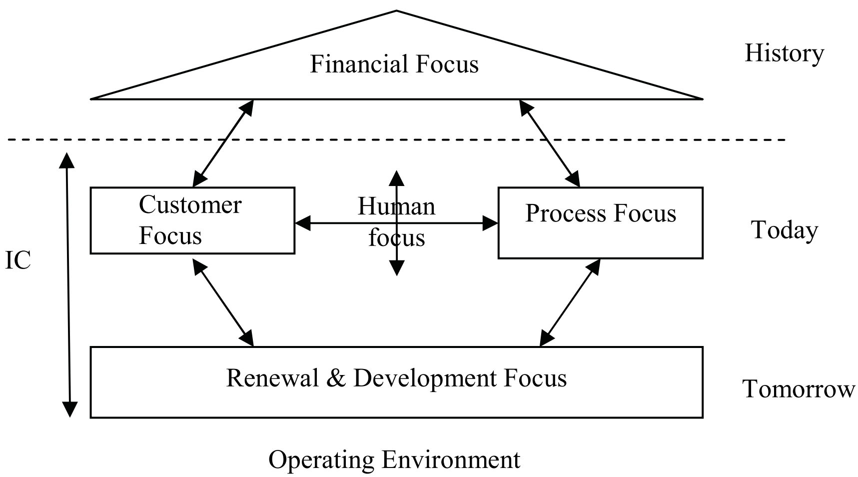


Figure 5. Skandia Navigator.

instructors who pertain to the result of performance are selected as output variables; and some instructors who do not pertain to IC carrier are eliminated. The remainder is considered input variables. (See Table 2.)

Table 2. Data of Skandia for ANN simulation.

	Indicators	1997	1996	1995	1994	1993
Input Layer	Savings/contract (SEK 000s)	499.0	396.0	360.0	333.0	371.0
	Surrender ratio (%)	4.4	4.4	4.1	4.2	3.6
	Points of sale	45881.0	33287.0	18012.0	11573.0	4805.0
	Number of employees, full-time	599.0	418.0	300.0	220.0	133.0
	Number of managers	88.0	86.0	81.0	62.0	37.5
	Training expense/ employee (SEK 000s)	2.7	15.4	2.5	9.8	10.6
	Adm. exp./gross premiums written (%)	3.5	2.9	3.3	2.9	2.6
	IT expense/admin. expense (%)	8.1	12.5	13.1	8.8	4.7
	Share of gross premiums written from new launches (%)	0.9	23.7	49.2	11.1	5.2
	Increase in net premiums written (%)	31.9	113.7	29.9	17.8	204.8
	Development expense/ adm. exp. (%)	9.8	9.9	10.1	11.6	9.8
	Share of staff under 40 years (%)	76.0	78.0	81.0	72.0	74.0
Output Layer	Return on capital employed (%)	21.9	27.1	28.7	12.2	24.3
	Operating result (MSEK)	1027.0	579.0	355.0	115.0	96.0
	Value added/employee (SEK 000s)	2616.0	2206.0	1904.0	1666.0	1982.0
	Number of contracts	189104.0	133641.0	87836.0	59089.0	31997.0

Source: Data of American Skandia, Supplement to Skandia's Annual Report, 1994–1998.

4.1.2. *ANN Model and Learning Arithmetic*

The BP (Back propagation) network fits well with the case study of Skandia. BP refers to a classic arithmetic in ANN. BP is a sort of "feedforward neural network" and is the most widely used in application. Feedforward neural networks are the first and arguably simplest type of ANN devised. In this network, the information moves in only one direction--forward--from the input nodes, through the hidden nodes (if any), and to the output nodes. There are no cycles or loops in the network.

BP is considered as a network no feedback. "No feedback" in this case supposes that the input of material capital and the marketing elements are stable, and that the feedback of learning either does not greatly influence the status of IC or the influence can be neglected.

The strength of BP network lies in its adaptability and effectiveness. Research shows that for a network with two layers (the input layer not included), the Sigmoid function used in the through layer and the linear function in the output layer, almost all functions can be approached with any precision using the BP network (Hagen, 2002).

Since BP can meet the requirements of most engineering applications and because the precision requirement for the domain of IC is lower than the engineering domain, BP can be well applied for IC measurement.

The model of BP network for Skandia is shown in Figure 6. IC carrier instructors of Skandia are defined as the input layer, and the performance instructors as the output layer. Sigmoid function ($f(x)=1/1+e^{-x}$) is applied between the input layer and the concealed layer, and linear function between the concealed layer and the output layer.

4.1.3. *Software*

Matlab provides ANN functions in its toolkits. Some software platforms are developed based on Matlab. We utilized software named as "NNBP ANN Simulation Tool" for calculation, which is executed in two principal steps. One is network training. After having standardized the input instructors and the output instructors, the standard data are utilized for input layer and output layer. By training these samples, a matrix of

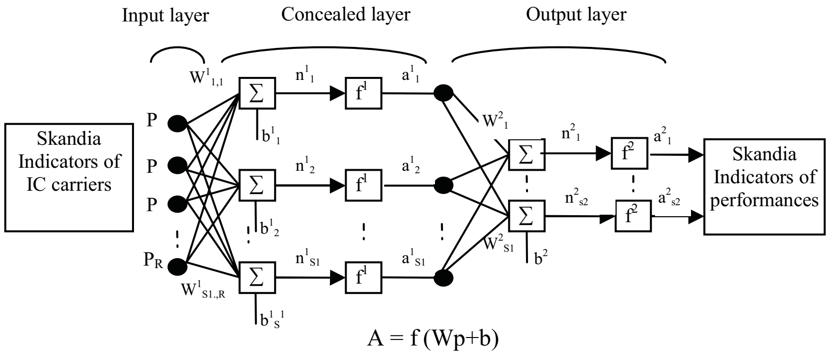


Figure 6. The BP model of Skandia for IC kernel measurement.

weights and a matrix of thresholds can be constructed. A second is network simulation. In virtue of the matrixes by sample training, the test data is simulated to get the network output.

4.1.4. *Calculation and Analysis*

Standardization process of sample data

The Sigmoid transfer function is generally utilized for the concealed layer. In order to improve the training speed and delicacy and to avoid the saturation section of Sigmoid function, the input values are required within 0~1. For realizing it, the paper adopts the extremum mode of standardization. The formula is as follows:

$$x' = \frac{X - X_{min}}{X - X_{max}}$$

In order to make the model extrapolated in some extension, the standardized data should be regulated within [0.2-0.8]. Therefore, the formula is modified as follows:

$$x = \frac{x' - x'_{min}}{x'_{max} - x'_{min}} \times d_1 + d_2$$

Thereinto, d_1 equals to 0.6, and d_2 equals to 0.2.

4.1.4.1. The calculation and results

Since we can only dispose five groups of sample data for each calculation, we use four for training and one for simulation and test. In this way, we can do the training and the simulation five times and get five different results. The parameters for network training are as follows: the input layer: 12 nodes; the concealed layer: 20 nodes; the output layer: 4 nodes; the maximum times of training: 10,000; the precision requirement: 1e-006; step length: 0.001. The result is shown in Table 3.

Table 3. The comparison of simulation results and true values.

Tested samples Indicators of output level		1997	1996	1995	1994	1993
True values	1. Return on capital employed	0.552727	0.741818	0.8	0.2	0.64
	2. Operating result	0.8	0.511278	0.366917	0.212245	0.2
	3, Value added/ employee	0.8	0.541053	0.350316	0.2	0.399579
	4. Number of contracts/ employee	0.8	0.588184	0.413252	0.303466	0.2
Simulation results	1. Return on capital employed	0.444011	0.704917	0.659067	0.690492	0.448495
	2. Operating result	0.409721	0.388866	0.506639	0.611026	0.2708
	3. Value added/ employee	0.500716	0.466958	0.478771	0.535355	0.379494
	4. Number of contracts/ employee	0.453774	0.41206	0.662153	0.541961	0.439796

According to the results, the relative errors for the five experimentations are calculated, and then the average relative error for every group of test is obtained. (See Table 4.)

Table 4. Average relative errors of simulation results.

Tested samples	1997	1996	1995	1994	1993
Average relative errors	37.3%	18.1%	38.1%	169.9%	47.6%

These results show that the average relative error for the tested sample "1994" is the largest, and that the average relative errors for the others range from 18-48%. In statistics, generally speaking, an average relative error below 30% is acceptable, whereas the training is executed by such a small quantity of samples that errors are basically inevitable.

4.1.4.2. Analysis

It is understandable that the result of output is different from the real value, even if the error is sometimes quite large[1]. Two reasons are offered. One, the precision of network represents a concept as a whole. The precision of ANN does not lie only in the precision of several samples, but in the overall precision of network on the condition that sufficient samples are included in training the network. Two, the error of

---

[1] We also did such a experiment: training all the five samples, and selecting by haphazard one group of data for testing the model:

The outputs of simulation : 0.6400419 0.199819 0.3994663 0.2004151

The true outputs of the 5th samples: 0.64 0.2 0.399579 0.2

It shows that the results accord with the true outputs very well. That's because the samples tested have already participated in training. That is to say, this sample (if we compare it to a point in a mathematical curve) has been simulated on the curve (which represents the mathematical relation). So, the results of simulation are certainly conform to the real values. Whereas, these results just show that the model has succeeded in training the samples.

result for an individual test cannot reflect the precision of model. Only when most test outputs conform to the true outputs, can we acknowledge that the model is precise. On the contrary, if the tests have limited execution, it is not appropriate to conclude that a model is not precise.

In our case, the samples for training are very limited because of the reality of the IC research. Thus, it is natural that there appeared relatively large errors in the tests of simulation. However, with the increase of the number of samples, the precision of model will surely improve. Alternatively, with regard to small samples, suppose that the result of error is not large. Can we conclude, then, that the model is highly precise? The answer is “no”. For small samples, high precision is only by accident. Generally speaking, the fewer the samples, the lower the precision. In summary, for our case study, a conclusion cannot be made according to our results.

5. Conclusions

IC is a new domain valuable to explore and probe, but IC also faces numerous difficulties, particularly in IC measurement. Researchers often believe the factors used for instructors for IC measurement merely refer to IC. Indeed, only when a relationship between factors is firmly proved, then approved by an academic community, can IC instructors be named. A great number of empirical studies are required before anything becomes certain.

To measure IC kernel through IC carrier is an interesting proposition. The significance of this empirical study lies in that it has explored the study of IC measurement by applying the method of ANN. This exploration accumulates the experience for greater scale study. It is necessary to indicate that the selection of topology of network should depend on the concrete problems. In our case study, a simple network BP is selected. Two conclusions can be made from our case study of Skandia:

1) The calculation result shows that the errors are acceptable according to the number of samples. Therefore, it is possible to use the ANN method for a new case study and to forecast.

2) For other firms, it is feasible to try to construct an ANN model in order to obtain the mathematical relation between IC carrier instructors and organizational performances.

To improve the precision of model for measuring the IC kernel, the study should have more samples to deepen the application of this method by studying more empirical cases. We would like to find enterprises who have an interest in this method. In addition, the authors hope this new method of analysis will provoke discussion and result in an echo from the academic community.

In fact, researchers concentrating on IC are fewer in number than those concentrating on knowledge management in a narrow sense. It is evident that more research and more practices are required to make significant progress in IC research.

References

1. Andreou, A. (2005, January 19-21). A suggested human capital knowledge asset index (HC-KAI). In *Proceeding of 26th McMaster World Congress*, Hamilton: Ontario, Canada.
2. Austria Research Center Seibersdorf (ARCS). (1999; 2000; 2001; 2002), Intellectual Capital Report.
3. Bontis, N. (2001). Assessing knowledge assets: a review of the models used to measure intellectual capital. *International Journal of Management Reviews, 3(1),* 41-60.
4. Bontis, N. (1998). Intellectual capital: an exploratory study that develops measures and models. *Management Decision*, 36(2), 63-76.
5. Bagozzi, R. & Phillips, L. (1982). Representing and testing organizational theories: a holistic construal. *Administrative Science Quarterly*, 27, 459-489.
6. Cabrita, M. (2005, January 19-21). Intellectual capital and business performance in Portuguese banking industry. In *Proceeding of 26th McMaster World Congress,* Hamilton: Ontario, Canada.
7. Choo, C. & Bontis, N. (2002). *The strategic management of intellectual capital and organizational knowledge.* New York: Oxford University Press, Inc.
8. Cong, S. (1998). *The theory and application of artificial neural network oriented the toolkits of MATLAB.* Hefei: Press of University of Science & Technology of China. In Chinese.
9. DLR (German Aerospace Research Center and German Space Agency). (2000; 2001). *Intellectual Capital Report.*
10. Drucker, P. (1993). *Post-capitalist society.* Oxford: Butterworth-Heinemann Ltd.
11. Edvinsson, L. & Malone, M. (1997). *Intellectual capital: realizing your company's true value by finding its hidden brainpower.* New York: Harper Business.

12. Edvinsson, L. & Sullivan, P. (1996). Developing a model for managing intellectual capital. *European Management Journal, 14(4),* 356-364.
13. Grant, R.M. (1996). Toward a knowledge-based theory of the firm. *Strategic Management Journal, 17,* 108-122.
14. Hagen (2002). *The design of neural network.* Beijing: Press of Mechanism Industry. Chinese version.
15. Huang N. (2005, May). *A theoretical exploration on intellectual capital and a proposition of a model framework for intellectual capital measurement.* Dissertation of Chinese Academy of Sciences. In Chinese.
16. Nonaka, I. & Takeuchi, H. (1995). *The knowledge creating company: How Japanese companies manage the dynamics of innovation.* Oxford: Oxford University Press.
17. Qiu G. (2001). Measuring intellectual capital with applying ANN. *Chinese Journal of Management Science,* (Suppl.), 683-688. In Chinese.
18. Serenko, A. & Bontis, N. (2004). Meta-review of knowledge management and intellectual capital literature, *Knowledge and Process* Management, 11(3), 185-198.
19. Seetharaman, A. et al. (2002). Intellectual capital accounting and reporting in the knowledge economy. *Journal of Intellectual Capital, 3(2),* 128-148.
20. Skandia. (1994; 1995; 1996; 1997; 1998). *Intellectual capital report.* Stockholm.
21. Sveiby, K. E. (2004). *Methods for measuring intangible assets.* Retrieved from http://www.sveiby. com/articles/IntangibleMethods.html.
22. Zhu Y. & Yuan Y. (2002). *Application of ANN in intellectual capital measurement.* (CAD/CAM and Informatization of Manufacture, No. 2). In Chinese.

This page intentionally left blank

Chapter 16

THE COMPILATION OF MATH PATTERN FOR PRODUCTIVITY IN INDUSTRIAL ORGANIZATIONS BASED ON ORGANIZATIONAL CLIMATE

Fattah Nazem
Islamic Azad University, Roudehen Branch, Iran
f_Nazem@Yahoo.com

The purpose of this research is to investigate the relationship between organizational climate and the productivity of managers' services in industrial organizations with the goal of presenting a suitable management pattern. The subjects were 349 managers who were randomly selected from a research population comprised of all managers who work in the companies of the National Petro-Chemical Industries. The research tools are questionnaires given to staff regarding organizational climate (Litwin & Stringer, 1968) and management productivity (Smith et al., 1998). For each manager, three clerks (totaling 1,047 subjects) were chosen to complete the questionnaires and evaluate their immediate managers. The results of applying the analysis of multi-variation regression are as follows:

A) There is a relationship between the organizational climate and the productivity of the managers' services.
B) The math pattern for the productivity of the managers' services in industrial organizations is demonstrated

Productivity = 0.65 × organizational climate +(-29.19)

1. Introduction

Perfection is the state for which humanity strives, and this notion serves as a driving force in the historical development of life. Productivity is but one of the indexes toward reaching perfection in man-made systems, e.g. intentional activities of humans. More than ever, the promoting of productivity is not only a goal, but it is a necessity in all aspects of life when considering the lack of resources and increased competition in the world economy. Countries worldwide try to increase their index of national, industrial, commercial, and educational productivity in a number of ways, recognizing that higher productivity leads to better social security and welfare. Because earnings increase when resources and products are used appropriately, increases in production and competition are also achieved. Success, in this way, leads to increased business and ultimately to a higher quality of life. Such a state is not realized without an acceptable attitude toward productivity and the promotion thereof (Kopelman, 1984). In this, managers have an important role. Peter Drucker (1974) believes that mangers are the most rare and most valuable resources of organizations. He defines them as the main and decisive factor in society where they fulfill a basic role in the development and flourishing of such. Qualified and knowledgeable managers can attain an organization's goals by utilizing their internal capabilities, specialized knowledge, and professional experiences all the while using fewer resources. In this way, managers can increase the organization's efficiency and effectiveness.

Making the most of these salient characteristics, managers should therefore apply given data and the results of management studies with respect to the organization's condition, to select suitable methods and then provide a favorable climate in which the organization's goals may be attained.

2. Literature Review

According to French et al. (1985) the organizational climate of a collection has a direct and steady relation with the perception of the organization members about its cultural features. This perception affects people's feelings, attitude, and behavior in their workplace. Boulden (1992) also believes that the organizational climate is an environment in which people work, as well as a reflection of staff attitude and the style of organizational management. The organizational climate, therefore, consists of a system of values which define the methods for doing something and distinguishes acceptable behaviors. According to Schneider (1990) the organizational climate has been broadly defined as the common perception of policies, activities, and organizational procedures either formal or informal which members can observe.

According to Owens (1991) the concepts of both organizational culture and organizational climate are structures that deal with the same fact, and people's behavior in organizations is not the result of interaction with direct and tangible events, but rather the result of interaction with the intangible powers in the environment.

Boone (1991) believes that organizational culture relates to nature, beliefs, and expectations of organizational life, while organizational climate is an index for defining and achieving these beliefs and expectations. Chandan (1995) states that organizational climate reflects the people's attitude toward the organization in which they are interested, and it is a collection of features and factors considered as the main force that is effective in defining the staff's behavior. Koene (1996) believes that organizational climate mostly depends on the description of work atmosphere as the members of the organizations understand it. The work atmosphere measures can be considered as environment assessment by the staff. On the other hand, organizational culture is considered as the description of the staff's behavioral preferences. In other words, the researcher who work in the field of organizational culture are not interested about what the people think, but they look for devices which affect the perception, motivation, and performance of the whole members of the organization. As examples, power distance and aloofness (both aspects of organizational climate) can be mentioned. The first one illustrates the tendency and normalcy of a person toward the acceptance of misbalance of power; and the second one refers to the staff's perception of a boss who is aloof. Denison (1996) believes that organizational climate has two distinct concepts. One is the people's common reflection or perception of a situation. Therefore, it may pose the atmosphere of satisfaction, resistance, conflict, etc. The other is the totality of condition, which affects people's behavior. The objective features of a social system are consistent with its reflections and those features.

The organization size is among the effective structural factors, which affect organizational climate. The results of a study show that in smaller organizations, there is a more open and truthful environment, as well as a friendly organizational climate. In addition, the professional status of an individual in the organizational hierarchy can affect his perception of organizational climate (e.g. Steers, 1977; Sofianos, 2005; Erbisch, 2004; Fouts, 2004). The study carried out by Burns and Stalker (1961) shows that the repetitive technology-like production line causes a kind of organizational climate, which is inflexible and rule based; therefore, staff innovation noticeably decreases. There is little information about the effect of the organization's external environment on its internal environment. Nevertheless, environmental changes can affect organizational climate. For example, in terms of economical condition, an organization may have to dismiss some staff, and this makes the staff feel that the organizational climate is threatening and not supportive (Ekrami, 2002). There are many studies (Seraj, 2005; Karami, 2005; Wilson, 2005; Durcikova, Galletta & Butler, 2004; Lambert, 2004; Fouts, 2004; DeMeritt, 2005; Jackson, 2005) that confirm the effect of management policies and activities on organizational climate. For instance, it has been known that the following factors cause staff to take responsibility in achieving the group and organization goals: feedback to the staff, individual independence, job identification, attention to environmental changes in the work place, staff placement in suitable posts, salary, relation, reward, and encouragement.

Cherrington (1989) believes that the factors affecting organizational climate include management values, leadership style, economical situation, organizational structure, staff characteristics, forming unions, organization size, and work nature. According to French et al. (1985), available theories and the results of studies that have been done are not enough to introduce the best organizational climate. Litwin et al. (1968) state that it is meaningless to tell the manager how to manage things. Instead, the manager chooses a climate that is necessary for highly favorable performance. The manager first gathers related information, e.g. the gap between the favorable climate and the present climate in his unit, then chooses the programs and operations needed for reaching such a climate. Steers (1977) believes that if the organization goal is to achieve favorable feedback and performance, the climate facilitating success will be more suitable; likewise, if the organization aims to satisfy its staff, a friendly climate will mostly suit it. The results of a study done by Timm and Peterson (1986) show that in a strong organizational climate there are factors like trust, shared decision making, attention to reports made by high level people, and adherence to the high performance goals (Ekrami, 2002).

In a study, Halpin and Croft (1963) identified six kinds of organizational climate that are on a continuum. Moran and Volkwein (1992) quote these as follows: open, closed, autonomous, controlled, paternal, and familiar. In the field of searching and investigating organizational climate, Likert also has presented the diagram of organizational characteristics called “profile of organizational characteristics” (POC) which contains eight dimensions of the fundamental characteristics of an organization. Regarding his theory of the four management systems, these dimensions consist of leadership process, motivating forces, relation process, penetration-interaction process, decision-making process, the order of goals, the process of supervision, and achieving educational goals (Gibson et al., 1973). In addition, Stern and Steinhoff have presented another formulation of organizational climate (Steers & Porter, 1979). Stern, like Halpin, believed that the human personality can be compared to the personality of an organization. Based on the studies of psychologist Henry A. Murray, he took the concept of “need-press” as a factor, which forms the human personality. Murray believed that the personality is the result of dynamic interaction between (internal) need and press (Owens, 1991).

After executing numerous researches in the field of organizational climate, Litwin and Stringer (1968) compiled a 50-item questionnaire that contains dimensions like structure, responsibility, reward, risk taking, warmth, support, conflict, standard, and identity, to name a few. Litwin and Stringer have showed that leadership style is among the factors that affect organizational climate. In addition to this, there are stable evidences of the relationship between organizational climate and other factors like emotional intelligence (Rogers, 2005), organizational learning (Jimenez, 2004), and job satisfaction (Stevens, 2005). Therefore, in order to fully achieve organizational goals, managers should supervise the organizational climate and create a suitable work environment in order to increase the productivity of the organization.

The word “productivity” was posed for the first time in 1776 by Quesnay (Sumanth, 1998). In that same year, Adam Smith shared his ideas about achieving work productivity, assigning work tasks and specialty for profit rise, reducing tiredness, and growing the use of technology (Nayudamma, 1980). Regarding the concept of productivity, Smith refers to efficiency and specialty and believes that work should be assigned based on people’s efficiency and productivity. Economists like Sinver defined productivity based on the worker’s physical, mental, rational, and intelligent quality, as well as his physical and mental power and skill. However, the most revolutionary research in productivity was accomplished by Taylor beginning in 1881, which can be considered as the history of formal and scientific studies about productivity management (Taylor, 1947). In the 19th century, Litter defined productivity as the power of production (Sumanth, 1998). Mahoney (1988) believes that productivity includes efficiency, effectiveness, and change. Likewise, scientists like Mescon et al. (1986), Boone and Kurtz (1991), Monga (1997), Robbins (1991), Ranftl (1989), Koontz et al. (1986), Stoner and Freeman (1992), Schermerhorn and John (1989), and Landel (1986) believe that productivity includes efficiency and effectiveness of performance, and increase in the productivity level in an organization is the result of the efficiency of management, which equals good management.

Studies show that productivity has meaningful relationship with the factors such as the styles of leadership and supervision (Foroutan, 2005; Akbari, 2005; Engle, 2004; Poxes, 2004), the quality of work life (Karimvand, 2004), managers’ education (Zamani, 2004), the observation of the principles of human relations (Chenari, 2004), the emphasis on team work (Chiu, 2005), the thought styles of managers (analytic) (Rahmati, 2005), the organizational climate (Barari, 2005), and entrepreneurship of managers (Hosseinzadeh, 2005).

French (1986) believes that the success and survival of organizations depend on the managers’ attention to the internal (organizational) and external environments and outcome (results). These factors affect each other. For instance, when the organizational management strengthens and supports educational programs (an internal factor), it affects the legal disciplines of employment (an external factor). As another example, the effectiveness of the organization (outcome) affects the quality and reasonable price of its products, thus increases the demand for that product (an external factor). "The International Organization of Work" has divided the factors that affect the organization productivity into two main groups:

A. External factors (uncontrollable): Those factors that affect the organization from outside and are not under the control of people or managers of that organization. In order to improve the performance of the organization, the external factors affecting the organizational management and efficiency should be pinpointed, and the organization should adapt itself to the changes of these external changes. These factors include financial rules and regulations, international politics, tax rules and statements, political, economical, and social relations.

B. Internal factors (controllable): These factors are under the control of people and managers of the organization. They can be used to promote the productivity of the organization if the contemplative managers apply them correctly (Prokopenko, 1992).

In his investigations, Sumanth (1998) has found 70 ways through which managers can promote the productivity of organization. These ways have been classified as five main groups based on technology, human resources, product, work (process), and material. This classification includes all ways based on the engineering of traditional industry, market (buying and selling), controlling systems, research on operation, computer engineering, management, psychology, behavioral sciences, and so on.

The importance of organizational climate as one of the characteristics of industrial environments and its effect on the productivity of the managers' services of Petro-Chemical Industries, leads the present research to form and examine the following questions:

1. Is there a relationship between the organizational climate and the productivity of the managers' services of the national companies of Petro-Chemical Industries?

2. What is a suitable math pattern for calculating the productivity of the managers' services of the National Company of Petro-Chemical Industries?

3. Research Methodology

In this research, the statistical population consists of all managers and supervisors (with educational diploma or higher) from 13 companies of the National Petro-Chemical Industries--petro-chemical companies of Arak, Shiraz, Bandar Imam, Isfahan, Razi, Tabriz, Tehran, Khark, Khorasan, Bazargani, Non- Industrial Operations, Development Management, and Oromieh--at an organizational rank of 15 to D who manage at least three staff

3.1. *Statistical Sample*

The volume of sample has been calculated to be 349 people using the formula: $n = \frac{z^2 \partial^2}{d^2}$. For each manager, three clerks are selected, totaling 1047 (349 × 3) people who have posed their ideas about organizational climate and the productivity of the managers' services in the specified industries.

3.2. *Research Tools*

The Litwin and Stringer questionnaire of organizational climate includs 50 multiple items (the questionnaire considers nine dimensions: 1) structure, 2) responsibility, 3) reward, 4) support, 5) risk, 6) warmth, 7) standard, 8) conflicts, and 9) identity. The study for productivity of managers' services was designed by Smith, Mccall, and Stoll and includes 17 multiple items--all which measure the productivity factor and refer to the efficiency and effectiveness of managers' performance. These were randomly given to 16 managers and their 48 clerks in Tabriz complex. After removing the problems and probable ambiguity of the items, the questionnaires were adjusted and administered. Based the calculations, the index Corenbakh's Alpha--which is the internal similarity index of items--equals $\propto = 0.98$ for organizational climate and $\propto = 0.94$ for the productivity of the managers' services. In this research and in order to find answers to our aforementioned research questions, the Pearson's Corelation Coefficient formula and linear multi-variation regression using SPSS software were applied.

4. Research Findings

In the present study and in order to answer the aforementioned research questions, the Pearson's correlation coefficient, and linear multi-variation regression have been used in conjunction with SPSS software.

Conducting the tests and data analysis, the following results were obtained:

1. There is a relationship between the organizational climate and the productivity of the managers' services of the National Company of Petro-Chemical Industries.
2. In suggested pattern, the nine dimensions--in order of their importance and with respect to their productivity in promoting the balance of the managers' services--are as follows: support, reward, warmth, identity, standard, conflict, risk, responsibility, and structure.
3. The standardized indexes (βi), resulting from the linear multi-variation regression, are shown in Table 1 in order to present the importance of the effect of the nine factors of organizational climate in the productivity of the managers' services.

Based on the research findings (and the researcher's thoughts and experience), this interactive pattern is presented for the management of the National Petro-Chemical Industries.

In order to predict the productivity of the managers' services of industries and based on the organizational climate of work environments using linear multi-variation regression, the following formula can be used:

Productivity = 0. 65 × organizational climate + (- 29.19)

Table 1. Standardized indexes (βi).

Dimensions	Standardized Indexes (βi) (Percent)
support	21
reward	17
warmth	16.7
identity	15
standard	14.7
conflict	6
risk taking	5.1
responsibility	3.7
structure	0.8

In addition to this, the productivity of the managers' services of the National Petro-Chemical Industries can be calculated by considering the effective variables on organizational climate as follows:

Productivity = 0.21 × support + 0.17 × reward + 0.17 × warmth + 0.15 × identity + 0.15 × standard + 0.06 × conflict + 0.05 × risktaking + 0.04 × responsibility + 0.008 × structure + (-18.30)

5. Conclusion and Discussion

Promoting productivity and protecting its growth is one of management's main goals. In fact, the basis of productivity management is to provide suitable conditions for acquiring the highest performance. The process of productivity management basically comprises change, and change can never occur easily. In order to achieve desired changes, the needed background should be provided, and the organization should experience some serious variations. In addition, the management ought to recognize and cope with the factors that are against these desired changes.

Regarding the subject under investigation and the factor of organizational behavior in explaining the managers' productivity, the obtained results show that there is a significant relationship between organizational climate and the productivity of the managers' services in different centers. As French et al. (1985), Boulden (1992), and Schneider (1990) show that organizational climate depends on the organization members' perception of the characteristics of the organizational culture and the staff's attitude toward it, we can emphasize the effect of this important factor on productivity promotion. However, the results of studies done by Burns and Stalker (1961) show the effect of the organizational climate--which has tendency toward rules and is not flexible--to be a reduction of innovation in the staff's performance. The existence of a friendly atmosphere (Steers, 1977) and the prominence of critical factors--trust, shared decision making, support, the willingness of top staff to have relationship with staff under their supervision, the attentive listening to reports from high ranking people, and the consideration of the goal of high performance (Timm & Peterson, 1986)--can lead to job satisfaction and improvement in the staff's performance, which ultimately lead to productivity promotion. Therefore, the findings of the present study are in agreement with the results obtained from the aforementioned studies.

Furthermore, Findings of the present study are in agreement with other studies concerning the components of organizational climate and their effects on the level of productivity; the results of studies about dimensions like leadership style (Seraj, 2005; DeMeritt, 2005); the dimensions of structure, risk taking, conflict, responsibility (Barari, 2005); trust, respect, and less worry about creating relationship (Erbisch, 2004); and salary, reward and appreciation, assessing tools for performance, relationship, structure, and leadership style (Fouts, 2004; Poxes, 2004; Karimvand, 2004; Chenari, 2004; Akbari, 2005). On the whole, the results of the present study and other studies show that productivity has a high correlation with the factor of organizational climate and its related components.

Regarding the results of this study, and recognition of the factors which are related to productivity of the managers' services in educational and industrial centers as well as the centers of higher education, it is concluded that the factors which can be effective in achieving the highest productivity are as follows: the existence of a suitable organizational climate in which there are cooperation and assistance among managers and other employees; bilateral support; trust creation; assistance for employees to accomplish their tasks; suitable rewards based on services provided; the appreciation of suitable jobs; the oversight of the mistakes that employees make without purpose; the creation of a friendly relationship; the acceptance of criticism; and the encouragement of employees to express their ideas.

Moreover, the creation of a friendly climate, the formation of informal groups, and the effort to make staff feel that the organization belongs to them will ultimately lead to productivity promotion.

References

1. Akbari, A. (2005). The relationship between the leadership style the managers' control position with the productivity of the staff who in the districts of 1 and 2 in towns of Soltaniyeh, Tarome Olya, Ijeroud, the Zanjan Zone, Zanjan Province (Masters dissertation, Islamic Azad University, Roudehen Branch, 2005). *MA Dissertation Abstract.*
2. Barari, F. (2005). The study of relationship between organizational climate with the productivity of the managers who work in municipalities of the east of Tehran province (Masters dissertation, Islamic Azad University, Roudehen Branch, 2005). *MA Dissertation Abstract.*
3. Boone, L. E., & Kurtz, D. L. (1991*).Management.* New York: McGraw-Hill.
4. Boulden, G. P. (1992). Productivity linked to profitability: The basis for improved work life. *In Better Quality of Work Life through Productivity*: *International Productivity Congress* (pp.69-70). Tokyo: Asian Productivity Organizations.
5. Bryman, A. (1992). Charisma and leadership in organizations. Sage Publications.
6. Burns, T., & Stalker, G. M. (1961). *The management of innovation*. London: Tavistock .
7. Chandan, J. S. (1996). *Organizational behaviour*. New Delhi, India: Vikas Publishing House.
8. Chenari, A. (2004). The relationship between the extent to which managers observe the principles of human relations with their productivity in boys' governmental high schools (Tehran) (Masters dissertation, Islamic Azad University, Roudehen Branch, 2005). *MA Dissertation Abstract.*
9. Cherrington, D. (1989). *Organizational behaviour.* Boston, MA: Allyn and Bacon Publishers.
10. Chiu, S. K. (2005). The linkage of job performance to goal setting, work motivation, team building, and organizational commitment in the high-teach industry in Taiwan (China) (Doctoral dissertation, Nova South-eastern University, 2005). *DBA Dissertation Abstract.*
11. DeMeritt, E. G. (2005). An examination of the effect of organizational storytelling on the administrative climate of church–based schools as a model for the leadership of nonprofit organizations (Doctoral dissertation, Capella University, 2005). *PhD Dissertation Abstract.*

12. Denison, D. R. (1996). What is the difference between organizational culture and organizational climate? A native's point of view on a decade of paradigm wars. *Academy of Management Review*, 21(3), 619-654.
13. Drucker, P. F. (1954). *The practice of management*. New York: Harper and Row.
14. Durcikova, A., Galletta, D., & Butler, B. (2004, August). The role of organizational climate in the use of knowledge management systems to support problem solving. *Proceedings of Americas Conference on Information Systems* (pp. 2225-2229). New York, NY.
15. Ekrami, M. (2005). Formulating a mathematical pattern for effective leadership in state universities based on organizational climate and managers' characteristics (Doctoral dissertation, Islamic Azad University, Reserarch & Science Branch, 2005). *PhD Dissertation Abstract.*
16. Engle, M. C. (2004). Changes in job satisfaction and productivity between team-directed and single- leader work groups (Doctoral dissertation, Capella University, 2005). *PhD Dissertation Abstract.*
17. Erbisch, P. M. (2004). Management style, organizational climate, and organizational performance in a public mental health agency: An integral model (Doctoral dissertation, Western Michigan University, 2005). *PhD Dissertation Abstract.*
18. Foroutan, M. H. (2005). The relationship between the managers controlling style with the productivity of the staff in Kanoun Parvaresh-e-fekri Koudakan va Nojavanan (Masters dissertation, Islamic Azad University, Reserarch & Science Branch, 2005). *MA Dissertation Abstract.*
19. Fouts, H. M. (2004). Organizational climate of North Carolina cooperative extension (Doctoral dissertation, North Carolina State University, 2005). *EdD Dissertation Abstract.*
20. French, W. L., Kast, F. E., & Rosenzwieg, J. E. (1985). *Understanding human behavior in organization.* San Francisco, CA: Harper and Row.
21. Gibson, J. L., Ivancevich, J. M. & Donnelly, J. H. (1973). *Organizations: structure, process, behavior.* JR.
22. Halpin, A. W., & Croft, D. B. (1963). *The organizational climate of schools*. Chicago: Midwest Administration Center of the University of Chicago.
23. Hersey, P., & Belanchard, K. H. (1988). *Management of organizational behavior: Utilizing human resources.* Englewood Cliffs, NJ: Prentice-Hall.
24. Hosseinzadeh, A. (2005). The relationship between managers' entrepreneurship and organizational climate with the high school teachers' performance (Ahvaz) (Masters dissertation, Islamic Azad University, Reserarch & Science Branch, 2005). *MA Dissertation Abstract.*
25. Jackson - Malik, P. J. (2005). Organizational climate and hospital nurses' job satisfaction, burnout, and intent to leave (Doctoral dissertation, University of Pennsylvania, 2005). *PhD Dissertation Abstract.*
26. Jimenez, B. M. (2004). Organizational climate and organizational learning in schools (New York) (Doctoral dissertation, School of Education and Human Services, ST. John's University 2004). *EdD Dissertation Abstract.*
27. Karami, K. M. (2005). The study of the relationship between managers' characteristics and the managers' controlling procedures with the productivity of the education offices

(Ministry of Education) (Tehran) (Masters dissertation, Islamic Azad University, Reserarch & Science Branch, 2005). *MA Dissertation Abstract.*

28. Karimvand, S. (2004). The study of the relationship between the staff's life quality with their productivity in the education offices (Ministry of Education) (Tehran) (Masters dissertation, Islamic Azad University, Reserarch & Science Branch, 2004). *MA Dissertation Abstract.*
29. Koene, B.A.S. (1996). Organizational culture, leadership, and performance in context: Trust and rationality in organizations (Doctoral dissertation, Masstricht University, 1996). *PhD Dissertation Abstract.*
30. Koonts, H., O'Donnell, C., & Weihrich, H. (1986). *Management*. New York: McGraw-Hill.
31. Lambert, W. E. (2004). Servant leadership qualities of principals, organizational climate, and student achievement: A correlational study (Doctoral dissertation, Nova South-eastern University, 2004). *EdD Dissertation Abstract.*
32. Landel, R. (1986). *Managing productivity through people: An operations perspective: text and cases.* Sponsors of the Colgate Darden graduate school of business, University of Virginia.
33. Litwin, G. H., & Stringer, R. A. Jr. (1968). *Motivation and organizational climate.* Boston, MA: Harvard University Press.
34. Mahony, T. A. (1988). Productivity defined: The relativity of efficiency, effectiveness, and change. In J. P. Campbell, & R. J. Campbell (Eds.), *Proceedings of the Productivity in Organizations: New perspectives form Industrial and Organizational Psychology.* (pp. 13-39). San Francisco, CA: Jossey-Bass.
35. Mescon, M. H., Albert, M., & Khedouri, F. (1988). *Management.* 3rd Ed. New York: Harper & Row.
36. Monga, R.C. (1997). The emerging definition of productivity: A proposal. *APO Productivity Journal,* Summer 1997, Nordica International LIT, 123-124.
37. Moran, E., & Volkwein. J. f. (1992). The cultural approach to the formation of organizational climate. *Human Relations*, 45 (1), 19-47.
38. Nayudamma, Y. (1980). Factors which hinder or help productivity improvement in the Asian region. Tokyo: Asian Productivity Organization.
39. Owens, R. G. (1991*).Organizational behavior in education*. 4th Ed. Boston, MA: Allyn and Bacon.
40. Poxes, A. M. Jr. (2004). Communication and organizational climate in the administration of NCAA intercollegiate athletic programs in the southeast United States (Doctoral dissertation, The University of Southern Mississippi, 2004). *PhD Dissertation Abstract.*
41. Prokopenko, J. (1992). *Productivity management: A practical handbook*. Genevea: International Labour Organization (ILO).
42. Rahmati, Z. (2004). The Study of the Relationship between Managers' Thinking Style and Organizational Climate with the Productivity of the Staff in Education Offices (Ministry of Education) (Tehran) (Masters dissertation, The Islamic Azad University, Roudehen Branch, 2004). MA Dissertation Abstract.
43. Ranftl, R. M., (1989). Seven keys to high productivity. In A.D. Timpe (Ed.), *Productivity: The Art and Science of Business Management* (pp. 84-99). New York: Kend Publishing.

44. Robbins, S. P. (1991). *Oganizational behavior: concepts, controversies, and applications*. New Delhi: Prentice- Hall of India Private Limited.
45. Rogers Gerrish, S. (2005). A study of the relationship of principal emotional intelligence competencies to middle school organizational climate and health in the state of Washington (Doctoral dissertation, Seattle Pacific University, 2005). *EdD Dissertation Abstract.*
46. Schermerhorn, J. R. Jr. (1989). *Management for productivity.* New York: John Wiley & Sons, Inc.
47. Schneider, B. (1990). *Organizational climate and culture*. San Francisco, CA: Jossey-Bass.
48. Seraj, A. (2005). The relationship between the creativity and leadership style with the productivity of the educational managers in secondary and high schools (Savadkoh) (Masters dissertation, Islamic Azad University, Roudehen Branch, 2005). *MA Dissertation Abstract.*
49. Silver, P. (1991). Educational administration. Theoretical perspectives on practice and research. New York: Harper and Row.
50. Smith, I., McCall, J., & Stoll L. (1998). The school management scale. *Management in Education*, 12(2), 27.
51. Sofianos, T. J. (2005). The relationship between organizational climate and job satisfaction as reported by community college executive secretaries and/or associates to the president (Doctoral dissertation, University of Florida, 2005). *PhD Dissertation Abstract.*
52. Steers, R. M (1977). *Organizational effectiveness. A behavioral view*. Santa Monica, CA: Good Year Publishing.
53. Steers, R. M., & Porter, L. W. (1979). *Motivation and work behavior.* New York: McGraw-Hill.
54. Stevens, D. J. (2005). The correlation of job satisfaction and worker productivity in the market data industry (Doctoral dissertation, Capella University, 2005). *PhD Dissertation Abstract.*
55. Stoner, J. A. F., & Freeman, R. E. (1992). *Management.* 5th Ed. Englewood Cliffs, NJ: Prentice Hall.
56. Sumanth, D. J. (1998). Total productivity management: A systemic and quantitative approach to compete in quality, price, and time. Boca Raton, FL: St. Lucie Press.
57. Sumanth, D. J. (1985). Productivity engineering and management: Productivity measurement, evaluation, planning, and improvement in manufacturing and service organizations. New York: McGraw- Hill.
58. Taylor, F. W. (1947). *The principles of scientific management.* NewYork, Evanston, and London: Harper and Row.
59. Timm, P. R. & Peterson, B. D. (1986). *People at work: Human relations in organizations*. West Publishing Company.
60. Wilson, C. L. (2005). Principal leadership, school climate, and the distribution of leadership within the school community (Doctoral dissertation, University of Montana, 2005). *EdD Dissertation Abstract.*

61. Wright, W. L. (1989). Overcoming Barriers to Productivity. In A. D. Timpe, (Ed.), *Productivity: The Art and Science of Business Management* (pp. 339-341). New York: Kend Publishing.

62. Zamani, A. (2004). The comparison between the productivity of the high school managers who enjoy specialized education in the filed of management with other high school managers whose fields of study are not management from the high school teachers' point of wiew (Shahroud) (Masters dissertation, Islamic Azad University, Roudehen Branch, 2004). *MA Dissertation Abstract.*

Chapter 17

FROM A PROPERTY OF THE AVERAGE OF FRACTIONS TO A TEXT-PROCESSING INTERFACE

Guillermo Oyarce
Texas Center for Digital Knowledge
School of Library and Information Sciences
University of North Texas
P.O. Box 311068, Denton, TX 76203
940-565-2445
oyga@unt.edu

The average of two natural numbers always falls between those two numbers. Partitioning a document set into two non overlapping subsets, the words in the set will appear only in one subset or on both. These properties can be used to present users with choices that can allow them to build a phrase where chosen terms have context. The average frequency of a term can be used to study relevance by comparing it to the same term's average in the relevant and the non-relevant subsets. A coefficient of variability (VAR) is defined as the normalized distance between these two values. The vast majority of words seem not to be significant because VAR as their relative comparative value is minimal. But a few words show very high values. This could be exploited by a system to find strong word instances that represent relevant concepts. It is possible to imagine an iterative procedure based on these properties through which a user identifies significant words with high VAR values. Such procedure would be desirable for a diversity of text-related computer-based tasks such as content analysis, thesauri construction, data mining, computer-based indexing and feature selection. A software instance to help users build context has been developed as a prototype to show the concept. Knowledge is always related to a given context and requires a support structure which has some cognitive elements such as other knowledge, data, concepts, information, etc. Using objective and subjective measures, users derive conceptual relationships. Users gauge and build topical relevance by engaging the system, which can then offer more suggestions. There is great advantage in reducing cognitive load and extraneous information. Users deal directly with easier to identify words, phrases and their combinations to form information capsules.

1. Introduction

A broad categorization of organizational knowledge distinguishes between recorded knowledge, also known as explicit knowledge, and the procedural knowledge captured at the individual user level, or tacit knowledge. There is a third category, implicit knowledge, or knowledge that cannot be articulated, but for the purposes of this article it will be treated as synonym for tacit knowledge. Management of Information Systems (MIS) and Information and Communication Technologies (ICT) theorist and practitioners have successfully provided business solutions related to the information that results in explicit knowledge, such as information organization, storage and diffusion, etc. With tacit knowledge, however, there is the opposite situation. One of the biggest challenges for Knowledge Management (KM) in a modern organization is related to the identification and effective use of tacit knowledge. Some enterprise-wide applications allow sharing, generation and management of explicit knowledge, such as inter-functional communication resources. New tools and approaches are necessary if incorporation of tacit knowledge is expected across ICT and MIS operations.

One such approach would be the development of automatic systems that elicit from the user and/or enterprise the appropriate tacit knowledge required to complete given tasks. One of those tasks is the focus of this article, (e.g. the identification of explicit knowledge in the internal repositories of the enterprise), also known by their generic name of information retrieval (IR) systems. These exist as either stand-alone or as part of larger systems, such as content management, digital libraries, and heterogeneous document repositories. Since the vast majority of IR systems use text to represent content, even in the case of multimedia, the current discussion will consider content as text-based. IR addresses the problem of identifying information relevant to a given information need. The information must be stored and organized for later access. While an apparently simple task, the information need is not always specific, or even explicit. Moreover, the impossibility to predetermine all uses of any one piece of information leads to a locked and rigid scheme of information storage and organization. Any searches that need the information that fall outside the initial assumptions will fail.

The purpose of this article is not to describe all the components of an IR system but rather to address a property of word frequencies that can facilitate user-system interaction by assisting users and systems to match their inherent tacit knowledge (e.g. what has been embedded in the system and what is owned by the user). An important aspect of this interaction is the system interface, which should allow users to directly manipulate the lexicon and also be able to identify or construct the appropriate context for their information need. The user-system interface is an important component of this particular human-computer interaction task, and it is supported by mathematical properties also introduced in this article. More to the point, the approach presented in this article can be customized for many supervised computer-based text processing tasks which require tapping into the tacit knowledge of the enterprise.

Feature selection in document processing refers to representation of objects through suitable terms or phrases (Ingwersen and Willett, 1995). At the core of this task is the problem of anticipating future needs and uses of the object. Relevant documents may remain invisible because a specific use was not foreseen during indexing (Lewis, 1992; Swanson, 1988). Understanding the strengths and weakness of this task is critical to computer-based text processing (Korfhage, Lin, and Dubin, 1995; Rorvig and Fitzpatrick, 1997; Rorvig, Sullivan and Oyarce, 1998). Reduction of the representation set is desirable. All of these issues are critical to minimize the user's information overload. Appropriate and efficient use of tacit knowledge, as an inherent cognitive characteristic of the human resource in the enterprise, stands to benefit from any reduction of potential overload. This paper examines the practical application of a mathematical property of lexicon, which could be used to reduce the feature set of the representation (Oyarce, 2000).

Current methods of feature selection calculate a weight of importance for each term in each one of the documents. The weight is normally a function of three parameters: the term's frequency in the document, the term's frequency in the collection, and the number of documents where it appears. These functions exacerbate a problem unique to document characterization - while high frequency terms are considered important and have more weight, those in very few and in most of the documents are considered to carry little information. Desirable terms have certain specific characteristics (Luhn, 1957; Spark-Jones, 1971; Lewis, 1992; Oyarce, 2000).

Documents and judgments of TREC-8 are used (Voorhees and Harman, 2000). The average word frequency across the document set is used as a weighting value for each term. Calculated in the relevant and non-relevant subsets, the difference between the two values shows potential for indexing to select representative words that are significant and discriminative. This article reports on the results.

2. Literature Review

In addressing this problem, Kui-Lam Kwok (1996) worked with the average term frequency (*avtf*) as a way to differentiate among terms with the same term frequencies. The idea is that the average frequency indicates significance and descriptiveness: "the *avtf*idf* weighting already produces good improvements to retrieval results" (Kwok, 1996, p. 192). These results are consistent with those found by other researchers, most notably Ari Pirkola (2002). However, the absence of data supporting their empirical assumption makes the resulting findings unclear. The intuition that the *avtf* value "seems a reasonable value to employ for differentiating term importance in short queries" (Kwok, 1996, p. 190), however promising, requires more testing. The literature shows that while the average term frequency value has been included in a variety of experiments, there has not been a systematic attempt to understand its behavior.

As IR evolved from simple exact Boolean matches in the early days, more complex probabilistic processes have advanced. Modern IR includes tasks to evaluate degrees of relevance due to the probabilistic nature of the vocabulary. The problem is, however, that these tasks are "often developed without reference to the data", and because they are not tested independently of the IR system, "it is difficult to decide whether unsatisfactory retrieval is due to the retrieval system or the underlying model" (Teevan and Karger, 2003, pg. 18). Likewise, theoretical frameworks that include term frequency, though not necessarily void of empirical support, tend to avoid examining the relationship between document frequency and term frequency with respect to relevance. Research that includes average term frequency looks for the most part at some of the performance measures or other metrics related to IR in general rather than to the statistics of the underlying distributions (Cronen-Townsend, Zhou, and Croft, 2002; Teevan and Karger, 2003). The present work sets forth to examine the distribution and behavior of the average term frequency in light of one of its key mathematical properties.

3. Theory

Notwithstanding meaning, a set of documents is made up of individual words. For a document ***d***, the frequency ***f*** of term ***t*** in the document it is always a positive integer greater than or equal to 1. Term Density ($TD_{\mathbf{T}}$) is defined as the average frequency of a term over a set of documents. In equation 1, F_{ALL} is the total frequency of a term in the set of documents and D_T is the total number of documents which contain that term.

$$TD_{t,d} = \frac{\sum f}{D_T} = \frac{F_T}{D_T} = \overline{TD_T} \qquad (1)$$

In the case of a retrieved set, assuming binary relevance, a document will only belong to one of two non-overlapping subsets: relevant or non-relevant. On the other hand, each term may appear only in relevant documents, only in non-relevant documents or in both relevant and non-relevant documents. Four partitions emerge with respect to each word and the relevance of documents:

1. Relevant documents that contain the word
2. Relevant documents that do not contain the word
3. Non-relevant documents that contain the word
4. Non-relevant documents that do not contain the word

$$F_{ALL} = F_R + F_N \qquad | \text{ for each term} \qquad (2)$$

$$D_{ALL} = D_R + D_N \qquad | \text{ for each term} \qquad (3)$$

R and N denote the relevant and the non-relevant subsets respectively. By definition:

$$\overline{TDall} \mid \frac{F_{\mathrm{ALL}}}{D_{\mathrm{ALL}}} \mid \frac{F_R \; 2 \; F_N}{D_R \; 2 \; D_N} \tag{4}$$

$$\overline{TD_R} = \frac{F_R}{D_R} \tag{5}$$

$$\overline{TD_N} = \frac{F_N}{D_N} \tag{6}$$

Inequality 7 is known to hold true (see proof in the appendix)

$$\frac{a_A}{b_A} \leq \frac{a_A + a_B}{b_A + b_B} \leq \frac{a_B}{b_B} \tag{7}$$

Simplifying (7) according to formulas (4), (5) and (6):

$$\overline{TD_A} \leq \overline{TDall} \leq \overline{TD_B} \quad | \quad \text{A and B can be indistinctively R or N} \tag{8}$$

All ***F*** and ***D*** values are non-negative natural numbers greater than or equal to 1; ***F*** is always greater than or equal to ***D***, therefore the lower boundary of TD is always 1. The right-side term in (8) has no upper boundary. Notice that the subscripts A and B in equation 8 can indistinctly refer to {R} or {N}.

To compare the relative significance of a particular term in both subsets, a coefficient of variability (VAR_D) is defined as the Euclidean distance between $TD_{\mathbf{R}}$ and $TD_{\mathbf{N}}$. The normalized value is shown in equation 9. Euclidean distance is routinely used throughout classification tasks for discriminant analysis. Negative values of VAR_D should be interpreted as a term which is more significant in the non-relevant subset than in the relevant subset. Division by zero is discussed in a later section.

$$VAR\_D_{CT} \mid \frac{TD_R \; 4 \; TD_N}{TD_{\mathrm{ALL}}} \tag{9}$$

This function plots as a rotated S-shape curve, showing that there is only a minority of terms that can be considered suitable for the purposes reported in this article. However, any combination of top ranked terms might produce good results.

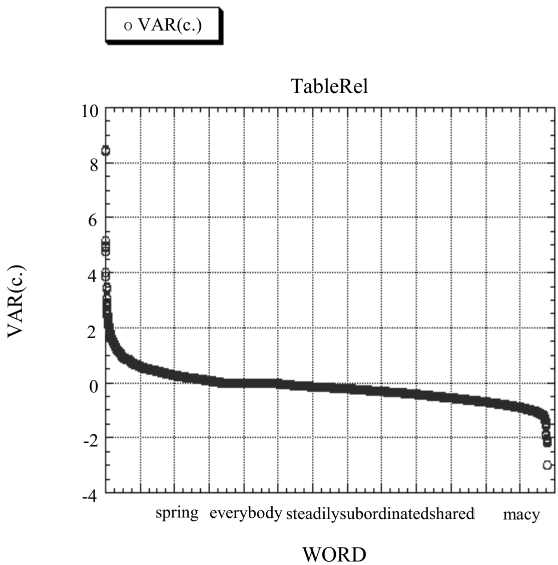


Figure 1. Distribution of VAR_D_{CT}.

This research seeks insight into the potential relationships of term statistics across documents in light of the term and document properties outlined above. A semantic cloud, defined as the network of concepts captured by the linguistic structures in a set of documents, may be examined. Expressed as a question, it would be: could the metrics presented in this discussion be used to identify words and concepts that cluster relevant documents while pushing away non-relevant documents? If so, how? In the experiment reported here, insight is sought regarding whether the proposed VAR_D_{CT} metrics capture significance of terms in document sets and the characterization of document sets to determine a-priori relevance.

4. Method

A suitable dataset comes from TREC 8, a total of 50 sets of documents corresponding to the 50 topics of the filtering task. They provide a ready-to-use relevance-judged dataset. The TREC program and its data are explained elsewhere (Voorhees and Harman, 2000). In TREC terminology, a topic is a specific trial where an information-need document is written outlining the boundaries of the documents that must be identified by the system. There were 50 topics in the filtering task of TREC 8. Important to this discussion is that the information needs for each TREC topic was prepared by experts on those topics and that these same experts also served as judges of the relevance of the retrieved documents. Each topic generated a set of judged documents. For our purposes, binary relevance is deemed sufficient.

Term dictionaries and document frequency tables were created for each topic. Figures 1–24 show the distributions of the different frequency and document metrics for the TREC 8 topic number 417. The distributions are prototypic for all the other topics. An entry for the word accident in topic 417 looks like this:

accidents: 71:271:2:3:69:268

There are three pairs of values for the word. The first pair of values refers to the whole document set; the next pair refers to the relevant subset and the last pair to the non-relevant subset. Each pair is the word frequency and the document frequency for that word. They are used to calculate the VAR_{CT} metrics according to formulas (9) and (10).

5. Division by Zero

Division by zero only takes place when TD_N would be 0, that is, a term doesn't exist in the non-relevant subset. This appears in the TD_N value in equation (10). For our purposes, the TD value of a term in a subset where the term doesn't exist will be assigned a small value, 0.001. When TD_N would be zero, the value 0.001 will be used. When TD doesn't exist in one subset, its value in the other subset is always equal to TD_{ALL}, therefore $VAR\_D_{CT} = 1$.

$$\mathrm{VAR\_D}_{\mathbf{CT}} = (\mathrm{TD}_{\mathbf{R}} - 0.001) / \mathrm{TD}_{\mathbf{ALL}} \approx 1 \quad | \quad \mathrm{TD}_{\mathbf{ALL}} \approx \mathrm{TD}_{\mathbf{R}},$$
$$\mathrm{D}_{\mathbf{N}} = 0,\ \mathrm{TD}_{\mathbf{N}} = 0.001 \qquad (10)$$

6. Rules

Straightforward heuristics are followed for each word in the dictionary. Since relevance is considered binary, the TREC dataset provides a great opportunity to examine different term distributions. Of particular interest is discrimination for relevance among significant words in a dataset. Towards that objective, the set of words is divided into three groups: words that only appear in relevant documents, words that only appear in non-relevant documents, and words that appear in both. These are three disjoint, non-overlapping subsets.

This approach allows the examination of the assigned values in each subset and the operational question of how to determine a-priori relevance. Although words that only appear in the relevant or in the non-relevant subset can be considered a-priori candidates of significance for {R} and {N} respectively, there is no probabilistic assurance that this is so because the word only exists in one sample (the relevant or non-relevant subset). In fact, this mistaken assumption has been a fundamental problem in text processing tasks such as automatic indexing.

The set of words that appear in both relevant and non-relevant documents is the particular focus of this article. It provides a way to observe the concentration of relevant words among those with highest $\mathrm{VAR}_{\mathbf{CT}}$ and that belong to relevant documents. Any conclusions can only be generalized across the same metrics, as shown in equations (9) and (10), not across all of the different distributions.

No prescriptive relationship among term frequency, document frequency, TD and VAR values is assumed. While the distributions of these values are predictable, neither one of the relationships between these distributions is assumed to be probabilistically sound and predictable. For these reasons, some arbitrary parameters are chosen to determine thresholds for each one of them. For each value, including word frequency, document frequency, TD and VAR, an ordinal list is formed so that the higher TD and VAR values are first. The rank (ordinal data type) is included in the dictionary as the index of the word's place in the decreasing monotonic list. Normalization is done as needed through percentage calculation for values to be compared to each other.

At the core of this experiment, a document set is segmented to examine characterization rules of optimal representative word sets that have high significance. The average term frequency is used as the weight of each term in an experiment designed to test this assertion. Using the documents retrieved for topic 417 at TREC 8, five random relevant documents were selected as the training set. Features were selected from all bigrams (two consecutive words). These features were then used to search the remained documents in a simple keyword search with no relevance feedback. The 417 topic was selected because out of 2992 documents retrieved there are only 75 (2.5 percent) relevant documents; that is, it mimics a real situation where the vast majority of documents are non-relevant.

The heuristic produces a set of terms that are used to identify target documents, which are decomposed into vocabulary components and compared to the baseline distribution of TD_{ALL}. Documents with at least one term whose frequency is higher than TD_{ALL} (thus VAR values are positive) are considered relevant. The reasoning for this determination is that such a term is part of the relevant information capsule within the document, even if no other information capsules in that document are relevant. This process is repeated for each term starting from the highest ranked term and down the list. Recall and precision values are calculated and plotted.

7. Results and Discussion

The results can be seen in Table 1, which shows the traditional recall and precision values. Several noteworthy points appear, such as the inverse relationship between recall and precision. While these localized results are not enough to generalize any trends, notice the quick descent of precision from a maximum value.

Table 1. Recall and precision values.

Rank	R	P
1	0.48%	18%
2	0.51%	19%
3	0.55%	20%
4	0.69%	8%
5	0.72%	8%
6	0.75%	9%
7	1.44%	6%
8	1.47%	6%
9	1.92%	5%
10	2.09%	5%

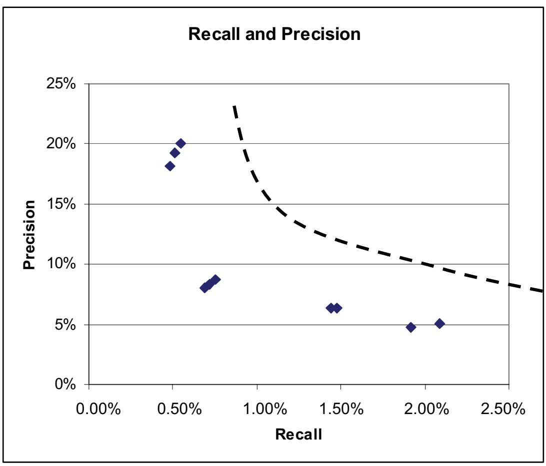


Figure 2. Recall and precision curve.

Since machine learning algorithms rely on a set of judged documents, an obvious result is that word ranking in descending order of VAR_{CT} can be used to identify suitable feature candidates as good representatives for machine learning. A more ambitious proposal would aim to present the logic and methodology in this research as part of an automatic learning algorithm.

Size reduction can also be accomplished through the setting of a certain cutoff point in the VAR_{CT} values. The sharp descent of the distributions seen in Figure 1 may be related to the sharp descent also seen in Figure 2. In general, more testing is required to find which one of the VAR_{CT} metrics performs better, under what conditions and for what purposes.

In addition to the suggestions above for future research, user-centered studies could be designed to investigate how well these measures behave in interactive settings. Looking into whether the standard deviation of a word, as an optimization measure of variation to partitioning sets, can be used to identify relevance of individual documents in the set seems promising.

From the plot in Figure 1, the vast majority of words appear to be weak discriminators, (e.g. have a similar value in both subsets). A few words, however, show very high discriminative value, some for relevance and some for non-relevance. This could be exploited by a system to find strong word instances to represent relevant concepts in the set. It is possible to imagine a procedure based on these properties through which is possible to identify significant words with high discriminating power in the document set. Such procedure would be desirable for a diversity of text-related computer-based tasks:

- Content analysis
- Thesauri construction
- Data mining
- Computer based indexing
- Feature selection

A software instance to use this property and helps users to build context has been developed as a prototype to show the concept. Information does not exist in isolation, it needs contexts. Similarly knowledge requires a supporting structure of certain cognitive elements, other knowledge, data, concepts, information, etc. Through this interface, users would evaluate term relevance within the context they are able to select. Adjacent words are built into phrases of that carry relevant significance. As an example, the words cat, house, and wild form different combinations:

- Cat house
- House cat
- Wild cat
- Cat wild
- Wild house
- House wild
- Wild house cat
- Wild cat house

A good representation must serve two purposes, it must describe content and it must discriminate for relevance among similar documents. In the above list, some phrases make more sense than others, but all of them have a distinctive meaning. When words acquire context, ambiguity is reduced and specificity increases. More words in a phrase contribute to specificity: compare the phrase ***the orange cat on the roof*** and the words ***cat*** or **orange**.

8. Brief Description of the Interface

An interface that takes advantage of the interactive potential of relevance feedback can be envisioned. Using recalculation to present users with relationship of terms in their current selections can assist them in finding adjacent words in order to construct relevant phrases. The construction of these phrases can be accomplished by several parameters related to term and document frequency, as well as the VAR_D coefficient introduced in this article. While some interfaces attempt to use graphics to represent semantic relationships, meanings are not universal. Enterprise culture and tacit knowledge have an effect on what constitutes relevant information, which is not necessarily constant in time, space, or across people.

An interface that allows users to access the vocabulary and form phrases engages the system, which can then offer more suggestions – or the existing adjacent words from which the user may select idoneous meanings. An important cognitive advantage is the reduction of extraneous information. Users deal directly with easier to identify words, phrases and their combinations, which are put together on-the-fly as information capsules. An instance of this idea has been presented at the American Society and Information Sciences (Oyarce, 2004a; Oyarce, 2004b).

References

1. Cronen-Townsend, S., Zhou, Y., & Croft, W. B. (2002). Predicting query performance. *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 299 – 306.
2. Ingwersen, P., & Willett, P. (1995). *An introduction to algorithmic and cognitive approaches for information retrieval.* Libri.
3. Korfhage, R., Lin X., & Dubin, D. (1995). VIRI: visual information retrieval interfaces. *Proceedings of the 18th International Conference on Research and Development in Information Retrieval (ACM SIGIR).* Danvers, MA: ACM.
4. Kwok, K. L. (1996). A new method of weighting query terms for ad-hoc retrieval. *Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp.187–195). Zurich, Switzerland.
5. Kwok, K.L (2005). An attempt to identify weakest and strongest queries. *Proceedings of SIGIR 2005, Workshop on Predicting Query Difficulty.* August 19, Salvador, Bahia, Brasil. (http://www.haifa.il.ibm.com/sigir05-qp/papers/kwok.pdf)
6. Lewis, D. D. (1998). Naive (Bayes) at forty: The independence assumption in information retrieval. *Proceedings of the 10th European Conference on Machine Learning,* Lecture Notes in Computer Science; 1398, 4-15.
7. Lewis, D. D. (1992). Representation and learning in information retrieval. (Doctoral dissertation, University of Massachusetts, 1992). *Dissertation Abstracts International, 53(02), SecB,* 937. (UMI Dissertation Services, A Bell and Howell Co.)
8. Li, W., Candan, K.S., Vu, Q. & Agrawal, D. (2001). Retrieving and Organizing Web Pages by ”Information Unit”. *Proceedings of the Tenth International WWW Conference* (pp. 230-244). Hong Kong, China.
9. Luhn, H.P. (1957). A statistical approach to the mechanized encoding and searching of literary information, *IBM Journal of Research and Development, 1,* 309-317.
10. Luhn, H.P. (1958). A business intelligence system. *IBM Journal of Research and Development, 2,* 314-319.
11. Oyarce, G.A. (2000). *A study of graphically chosen features for representation of TREC topic-document sets.* Unpublished doctoral dissertation, University of North Texas.

12. Oyarce, G. A. (2004a), Document discrimination modeling for interface design, *National Core Journal for the Information Management and Technology Profession. Documentation and Information Center, Chinese Academy of Sciences 10,* 24.
13. Oyarce, G. A. (2004b, November 13-18), *Kontextual Interactive Support System (KISS).* Paper presented at the annual meeting of ASIST 2004: "Managing and Enhancing Information: Cultures and Conflicts", Providence, RI.
14. Pirkola, A. (2002). The RATF formula (Kwok's formula): exploiting average term frequency in cross-language retrieval. *Information Research, 7(2).*
15. Rorvig, M.E., & Fitzpatrick, S. (1997). Visualization and scaling of TREC topic-document sets. *Information Processing and Management, 34(2-3),* 135-149.
16. Rorvig, M., Sullivan, T. & Oyarce, G. (1998, October 25-29). A visualization case study of feature vector and stemmer effects on TREC topic-document subsets. In C.M. Preston (Ed.), *Proceedings of the 1998 Annual Meeting of the American Society for Information Science, Information Access in the Global Information Economy, Pittsburgh, PA, 35,* 130-142.
17. Spark Jones, K. 1971. *Automatic keyword classification for information retrieval.* Connecticut: Archon Books.
18. Swanson, D.R. (1988). Historical note: Information retrieval and the future of an illusion. *Journal of the American Society for Information Science, 39(2),* 92-98.
19. Teevan, J. & Karger, D. (2003). *Empirical development of an exponential probabilistic model for text retrieval: Using textual analysis to build a better model.* Paper presented at the meeting of the 26th Annual ACM Conference on Research and Development in Information Retrieval.
20. Voorhees, E.M., & Harman, D. (2000). Overview of the eighth text retrieval conference (TREC-8). In *The 8th Text Retrieval Conference (TREC-8)* (pp. 1-24). Gaithersburg, MD: National Institute of Science and Technology.

This page intentionally left blank

Chapter 18

ANALYZING COMPUTER MEDIATED COMMUNICATION LOGS USING A MARKOV MODEL

Shantanu Pai and Qiping Zhang
Drexel University
College of Information Science and Technology
3141 Chestnut Street, Philadelphia, PA 19104 USA

One of the challenges in knowledge management in a computer-mediated environment is how to analyze the huge log file. Markov chain models have been widely used for analyzing several types of sequential activities such as web navigation and human-computer interaction analysis. However, the Markov chain analysis of computer-mediated communication is still unexplored. In this paper we consider the dyadic activity of computer mediated negotiation as sequential and having a temporal structure. The communication process is represented as a sequence of transition between states using a Markov model. We show that it is possible to discover probabilistic patterns in the negotiation activity by discovering sequences that may characterize the negotiation activity.

1. Introduction

With the rapid growth of knowledge management system, information exploration becomes a serious issue. It is almost impractical for individual users to search the huge documents in the systems to effectively find things useful to them (Sarukkai, 2000). Different approaches have been adopted to address the navigation problem in the search (Sarukkai 2000, Cheung, 1997; Shahabi et al., 1997). However such approaches do not provide temporal sequential information in the search.

In the area of computer mediated communication, cross-cultural communication has recently been the focus (Sarker, 2005; Zhang, et al., 2005; Zhang, Olson & Olson 2004; Setlock, Fussell & Neuwirth, 2004; Kersten, Koeszegi, Vetschera, 2003). One advantage of computer-based communication research is that it allows researchers to discretely and unobtrusively record communication data. However, such data is often qualitative in nature (e.g., chat transcripts), which makes drawing consistent conclusions a bit difficult.

In this paper we will discuss the use of Markov models for analysis of communication data. Using an example, we will first show how communication patterns can be represented as a Markov chain. In so doing, we will discuss the steps involved in transforming the raw data from the transcript into a format suitable for Markov analysis. Further, we will show how the use of Markov chains allows us to identify critical points in a communication activity. This Markov chain analysis can also be applied in analyzing log file for use in knowledge management system.

2. Markov Chains

A Markov chain is a stochastic (random) and sequential process in which the future state depends on the immediately preceding state (Smith, Olekalns and Weingart, 2005). Mathematically, this can be represented as follows:

$$p_{ij} = P\{X(t_n) = j \mid X(t_{n-1}) = i\}$$

where p_{ij} is the transition probability of going from state "E_i" at time t_{n-1} to "E_j" at time t_n. Transition probabilities from E_i to E_j (i= 0, 1, 2, 3…, j=0,1,2,3….) can be expressed in a matrix forms as:

$$P = \begin{bmatrix} p_{00} & p_{01} & p_{02} & \cdots \\ p_{10}\cdots & p_{11} & p_{12} & \cdots \\ p_{20} & p_{21} & p_{22} & \cdots \\ \vdots & \vdots & \vdots & \end{bmatrix}$$

The transition probability matrix above must satisfy the following two conditions:

$$\sum p_{ij} = 1, \text{ for all i and}$$

$$p_{ij} \geq 0, \text{ for all i and j.}$$

Markov chain models have been widely used for analyzing several types of sequential activities such as web navigation (Sarukkai, 2000) and human-computer interaction analysis (Thimbleby, Cairns and Jones, 2001).

Smith, Olekalns and Weingart (2005) recently used a Markov chain model to analyze communication processes in a negotiation activity. However, they used only two categories to code their raw communication transcripts, namely integrative and distributive. Distributive strategy is one in which the participants try to get as much as possible for themselves, while integrative strategy is one in which participants work towards achieving overall benefit. Use of additional categories may allow the exploration of tactics used by participants to arrive at their strategy.

In the following sections we will use data from our recent experiment to show our coding method, as well as analyze this data utilizing a Markov model.

3. Analyzing Communication Data with a Markov Chain Model

We used raw data from a two-person negotiation activity conducted as a part of a research experiment for analyzing cultural issuses in computer mediated communication (Zhang, Olson & Olson 2004). In the negotiation task, participants were asked to play the roles of marketing managers of two competitive companies in which they had to agree on prices for their three common drugs: Alpha (for diet), Beta (for cancer), and Gamma (for flu). Each drug had nine possible prices and corresponding profits. Each participant had a table (made visible only to them) showing their own profits. The profits were constructed in such a way that for one participant, a particular price would result in a great payoff; for the other, that same price would result in a loss. The goal set forth for the participants was to gain as much personal payoff as possible. In this scenario, the greatest monetary gain overall, notwithstanding individual profit, is achieved by way of choosing a set of three prices that are reciprocally beneficial. In such, one participant will receive a high payoff while the other realizes a loss; the reverse would be true for another product.

First the communication process was transcribed verbatim. A set of negotiation activity codes was developed based on the research literature in negotiation (Fisher 1997; Fisher & Ertel, 1995; Leung, 1997; Poole, Shannon & DeSanctis, 1992) and small group behavior (Moran & Carroll, 1996). Although these categories were developed for a specific negotiation task, their generic nature makes them applicable in analysis of verbatim transcripts in other domains. This coding method in computer-mediated communication research has been used in other research studies (Olson et al., 1993; Fu et al., 2000). Table 1 below shows the coding scheme of different communication activities used for our data.

Table 1. Categories for classifying participant statements in a negotiation task.

Category	Definition
Goal (g)	Statements of what is trying be achieved in the negotiation, general evaluation of negotiation activity
Option (o)	Statements or questions of possible solutions to the agreement
Criteria (c)	Evaluative statements with stories
Clarification (cl)	Clarification of previous statement
Role-playing (rp)	Stories related to the experiment roles, but irrelevant to negotiation, mainly occurring at the beginning or the end of the meeting
Story (s)	Stories occurring in the categories other than criteria or clarification of criteria
Persuasion (p)	Convincing opponent on one's strategy
Relationship (rel)	Statements concerning their relationship and future cooperation
Framework (fr)	General product prices
Fairness (f)	Considering opponent's view while negotiating
Meeting Management (mm)	Statements that move negotiation ahead

Second, the transcript is coded into one of the categories listed in Table 1. Several meaningful conclusions can be drawn from analysis of these categories. For example, a higher count on "fairness" may reflect participants' fairness to each other. However, frequencies of occurrence do not depict temporal structure in the negotiation process. For example, in studying cultural differences in a negotiation pattern, we would be interested in comparing how participants reciprocate "fairness" from their opponent. Markov analysis can help us answer such questions.

Third, the frequency of activity code transitions is calculated. For Markov analysis, each of the aforementioned categories is considered as a state. The frequency of transition from one state to another is first tabulated following the coding of the raw transcript. For example, for a two person negotiation activity, the sequence "GCCPOOCCPPGCOOC" represents the following sequences of statements:

- (Coded as G) Participants 1: So, our goal is to set a price for drug "A",
- (Coded as C) Participant 2: Can we just avoid "A", because I am interested in "B" now?
- (Coded as C) Participant 1: But "A" is my priority.
- (Coded as P) Participant 2: I really hope that you will first support me on "B".
- (Coded as O) Participant 1: Can we set these aside and discuss drugs "C"?

The frequency of transition tabulation is shown in Table 2 below. In the example of Table 2, after G(oal) statement there were 2 C(riteria) statements; after C(riteria) statement, there were 1 C(criteria), 1 P(ersuasion), and 1 O(ption) statements; after P(ersuasion) statement there were 1 G(oal), 1 P(ersuasion), 1 O(ption) statements; and after O(ption) statement, there were 2 C(riteria), and 2 O(ption) statements.

Fourth, the transition probability matrix is calculated by dividing each of the row frequency by the respective row sum. Table 3 is a transition probability matrix of the example above.

Table 2. Frequency transition matrix.

	G	C	P	O
G	0	2	0	0
C	0	1	1	1
P	1	0	1	1
O	0	2	0	2

Table 3. Transition probability matrix.

	G	C	P	O
G	0	2/2	0	0
C	0	1/3	1/3	1/3
P	1/3	0	1/3	1/3
O	0	2/4	0	2/4

Fifth, the transition probability matrix can be visualized in a state transition diagram for visualizing patterns in communication. Figure 1 below shows a state transition diagram for transcript from one entire negotiation session between a pair of participants in our study, which has much richer information than the brief conversation example used above in Table 2 & 3. Different states are represented in ovals. The numbers represent transition probabilities. A preceding "C" to each code stands for clarification statements. For example, "CMM" represents a statement clarifying (C) a previous meeting management (MM) statement. "CC" represents a statement clarifying (C) a previous criteria (C) statement.

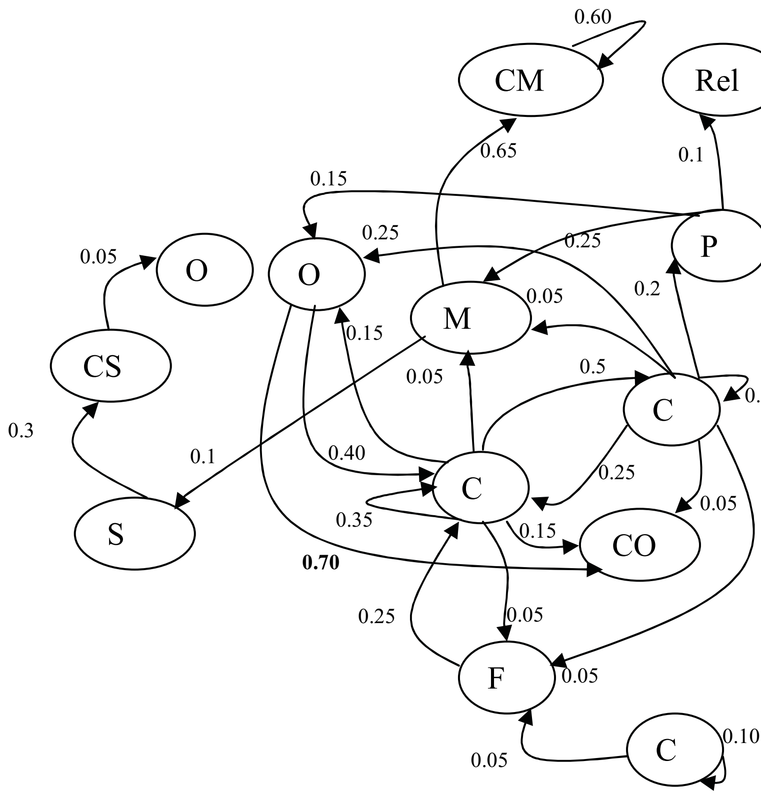


Figure 1. State transition diagram

It is possible to draw the following conclusions from the state transition diagram shown above in Figure 1.

1. The most probable transition from "Options" (O) is to a state of "Clarification of Options" (CO), given the changeover high transition probability (p_{ij} = 0.70). This may mean that understanding their options is clearly a priority for both participants.
2. "Options" (O) are reciprocated by "Criteria" (C) very frequently (p_{ij} = 0.40). This indicates that both participants like to evaluate their options against the criteria so as to maximize their gains.

3. The action persuasion (P) does not always generate options (O), as evident by the low transition probability ($p_{ij} = 0.15$).

Several other meaningful conclusions can be drawn using the transition probabilities values. Note that the transition probabilities shown above are for a single negotiation activity transcript and for illustration purposes only. The values will have more validity when we calculate transition probabilities using the average of the data from all the participants' pairs in our experiment.

4. Discussion and Implications

We consider the communication process as having a sequential structure. The goal of sequential analysis, then, is to find sequences that characterize our data. We used several code categories (Table 1) and were interested in finding if these codes have any probabilistic patterns. To discover this, we used the Markov model as a tool.

The Markov model discussed here is of the most basic kind and is often referred to as a first order Markov chain because the current state is dependent on only the immediately preceding state. A second order Markov chain would consider the last two states while arriving at the current state, and so on. However, our goal in this study was to show the use of Markov models and avoid higher-level mathematics.

It may not be enough to say that a sequence follows a first order Markov chain or that it has temporal structure. For example, in our experiment we were interested in finding differences in negotiation patterns for people with two different cultural backgrounds, namely American and Chinese. For this comparison, we need to find two separate-state transition diagrams for the two sets of data and subsequently look for differences in patterns. Also, log linear analysis can be performed on the frequency transition table to discover how frequencies vary as a function of the main effects. Readers are referred to Olson, Herbsleb, and Rueter (1994) for a detailed discussion on the use of log linear analysis for sequential data.

We believe that the analysis technique discussed here can be used to discover patterns in communication data from other domains such as emails, chat logs from customer support, doctor-patient dialogues, and the like, and further exploration merits attention.

References

1. Chueng, D. W., Kao, B. & Lee, J.W. (1997). Discovering user access patterns on the world-wide web, *Proceedings of First Pacific-Asia Conference on Knowledge Discovery and Data Mining* (PAKDD-97).
2. Fu, X.L., Zhang, K., Zhang, Q., Olson, G. M. & Olson, J. S. (2000). A comparison of quality and process of design activities of Chinese and American groups, *Ergonomics*, *6(1),* 1-5.
3. Fisher, R. (1997). *Coping with international conflict: a systematic approach to influence in international negotiation*. Upper Saddle River, NJ: Prentice Hall.
4. Fisher, R., & Ertel, D. (1995). *Getting ready to negotiate: the getting to yes workbook*. New York: Penguin Books.
5. Kersten, G., Koeszegi, S. T., & Vetschera, R. (2003). The effects of culture in computer-mediated negotiations. *Journal of Information Technology Theory and Application, Vol. 5* (2), 1 - 28.
6. Leung, K. (1997). Negotiation and reward allocations across cultures. (P. C. Earley & M. Erez Eds.). *New perspectives on international industrial/organizational psychology* (pp. 640-675): Jossey-Bass.
7. Moran, T. P. & Carroll, J. M. (Ed.). (1996). *Design rationale: concepts, techniques, and use.* Hillsdale, NJ: Lawrence Erlbaum Associates.
8. Poole, M. S., Shannon, D. L., & DeSanctis, G. (1992). Communication media and negotiation processes. (L. Putnam & M. E. Roloff Eds.). *Communication and negotiation: sage annual reviews of communication research,* 20, 46-66, Newbury Park, CA: Sage Publications.
9. Olson, G.M., Herbsleb, J.D., & Rueter, H. (1994). Characterizing the sequential structure of interactive behaviors through statistical and grammatical techniques. *Human-Computer Interaction, 9,* 42-472.
10. Olson, J.S., Olson, G.M., Storrøsten, M., & Carter, M. (1993). Groupwork close up: A comparison of the group design process with and without a simple group editor. *ACM Transactions on Information Systems, 11*, 321-348.
11. Sarker, S. (2005). Knowledge transfer and collaboration in distributed U.S.-Thai teams. *Journal of Computer-Mediated Communication, 10(4).*
12. Sarukkai, R. R. (2000). Link prediction and path analysis using Markov chains. *Computer Networks, 33,* 377–386.
13. Setlock, L. D., Fussell, S. R., & Neuwirth, C. (2004). Taking it out of context: Collaborating within and across cultures in face-to-face settings and via instant messaging. *Proceedings of CSCW 2004* (pp. 604-613). NY: ACM Press.
14. Shahabi, C., Zarkesh, A., Adibi, J. & Shah, V. (1997). Knowledge discovery from users web page navigation, *Proceedings of the 7th International Workshop on Research Issues in Data Engineering(RIDE'97) High Performance Database Management for Large-Scale Applications,* 20-29.

15. Smith, P. L., Olekalns, M., & Weingart, L. R. (2005). Markov chain models of communication processes in negotiation. *International Negotiation, 10,* 97-113. Hague, The Netherlands: Kluwer Law International.
16. Thimbleby, H., Cairns, P., and Jones, M. (2001). Usability analysis with Markov models. *ACM Transactions on Computer-Human Interactions*, *8*, 99–132.
17. Zhang, Q., Chintakovid, T., Sun, X., Ge, Y., and Zhang, K. (2005). Saving face or sharing personal information? A cross-cultural study on knowledge sharing. *Journal of Information & Knowledge Management, 5(1),* 73-80.
18. Zhang, Q., Olson, G.M., and Olson, J.S. (2004). Does video matter more for long distance collaborators? *Proceedings XXVIII International Congress of Psychology* (CD-ROM).

Chapter 19

ADAPTIVE AND CONTEXT-SENSITIVE INFORMATION RETRIEVAL

Axel-Cyrille Ngonga Ngomo
University of Leipzig
Department of Computer Sciences
Augustusplatz 10-11
04109 Leipzig, Germany
ngonga@informatik.uni-leipzig.de

Retrieving information stored electronically has become a more and more time consuming task, especially in IT-driven companies. We assume this is partially due to the commonly accepted "one-fits-all" construction of search engines. In order to increase the quality of Information Retrieval (IR) in the company environment, we developed a system that increases the accuracy of standard IR engines by two means: One, expanding user queries using a description of the user context and two, learning terms describing user contexts out of implicit user feedback. This paper presents a formal description of the architecture of the system, the implementation details, and our results.

1. Introduction

The use of digital data storage has grown in popularity during the last four decades, and this trend continues. Yet retrieving desired data has become one of the most time-consuming tasks for knowledge workers, especially those in a company environment. Most Information Retrieval (IR) engines have been developed to retrieve information using classical models, such as the vector space and the Boolean model, which operate by searching through the data set relevant to words in the user query. These models, however, are based on the assumption that the pieces of information needed are the same for all users. In this paper, we suggest that IR using exclusively these methods is one reason for the poor quality of IR we observed when studying the use of IR engines by knowledge workers in company environments. We therefore propose an IR-technique that incorporates the user context during the retrieval process. The context of a user is described as a set of ontological concepts that is transformed into a vector of terms used to expand the user's queries. Implicitly collected user feedback is used to adjust the vector weights and, thus, to improve the retrieval quality. The outline of this paper is as follows: After presenting related work exploring semantic information retrieval and user feedback, we will formally describe the method used. We will then delineate the IR-process, including implicit user feedback. Lastly, we will present implementation details and results.

2. Related Work

The topic of this paper is linked to the areas of semantic IR and user feedback. Using thesauri or other forms of semantic knowledge to improve IR has previously been done in different settings in which most try to expand user queries and/or document descriptions. An early experiment to evaluate the use of thesauri was carried out by Wang et al. (1985) and showed that its use for query expansion may noticeably improve both precision and recall. Yet experiments conducted by Voorhees and Hou (1992) and Voorhees (1994) on subsets of the TREC collections showed not only an improvement but also a degradation of the precision or recall can result from using thesauri for IR. In more recent papers, authors like Carpineto et al. (2001) showed that query expansion may considerably improve the recall and precision of IR engines.

Some approaches use formal semantic models. For example, the Thyssen-Krupp-Community-World project (Mueller, 1994) exploits a distributed semantic model that connects information sources and people in a knowledge map that can be used to retrieve relevant information in an efficient manner.

Other approaches aiming at improving retrieval quality use methods based on user feedback. Since the invention of relevance feedback (Rocchio, 1971), different methods to explicitly collect and process user feedback have been developed. However, such methods usually do not produce a representative amount of data because users are normally reticent to evaluate retrieval results. Hence implicit methods were developed to collect data from the user without requiring the user to explicitly evaluate documents. Most techniques of this type monitor the time a user spends reading a document and performing basic direct interactions such as clicking, scrolling, printing etc. (Morita and Shinoda, 1994; Claypool et al., 2001). Kelly and Teevan (2003) offer a good overview of actual techniques used in this domain.

3. Formal Specification

We developed a software tool that combines both semantic query expansion and implicit user feedback to improve the retrieval quality of a search engine. Figure 1 shows that the core of the system can be subdivided into four layers, namely the *business process layer,* containing the role and the task of the user (which are used as context); the *ontology layer,* describing the semantic world of each role; the *adaptation layer,* connecting each concept with semantically close terms; and the *term-net layer,* displaying the semantic similarity of terms.

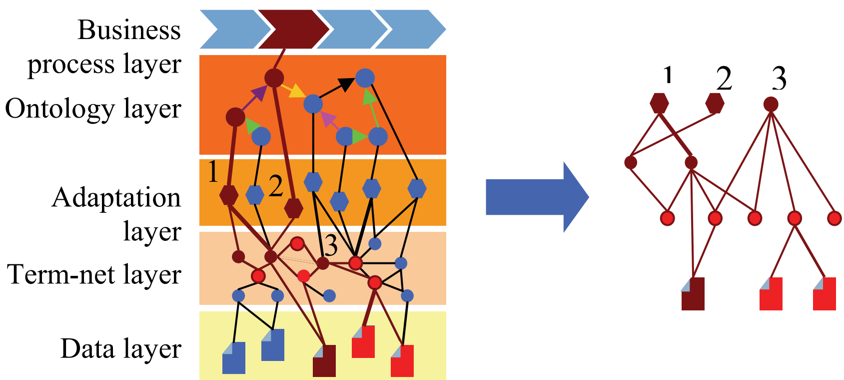


Figure 1. System architecture and retrieval path.

The *ontology layer* contains a family O of ontologies O_r, r being an element of the set R of roles. Let T be the set of all tasks and $t \in T$. We define an ontology O_r as a subset of $C_r \times Rel_r \times C_r$, C_r being the set of all concepts related to the role r and Rel_r the set of all semantic relations between them. Each user context (r, t) is related to a subset $C_{t,r}$ of C_r.

The *adaptation layer* A is a bipartite, weighted and directed graph from the sets C_r to the set of terms K: $A = \{A_r, A_r \subset C_r \times \Re + \times K, r \in R\}$. Let ω_A be the weight function of the adaptation layer. The computation of the initial weight of the edges between a concept $c \in C_r$ and the corresponding terms will be explained in the following section.

The *term-net* W is a weighted and directed graph. In this subset of $K \times \Re + \times K$, $a \in K \Rightarrow (a,0,a) \in W$ holds.

The initial weight of each edge is computed using distributional feature vectors. We first compute the frequency distribution of words in the corpus. Supposing a Poisson distribution of words and considering a window of w = 2 to 3 words around the $\phi = 200$ most frequent ensures a text coverage of more than 50% when using a corpus of 500 MB or more. The feature vector of a word w is the concatenation of ϕ vectors twice as wide as the window size, in which the number of occurrences of w in the relative positions between -w and w relative to each high frequent word are stored. The distance between the vectors is computed using the well-known cosine measure. The weight of the edge linking a word w and a concept c is obtained by normalizing the inverse of the distance between the label of c and w by the sum of the inverse distances between c and its k (in our case k = 4) nearest neighbors. Figure 2 shows an example of such a graph computed using the OHSU-TREC-9 corpus[1] for filtering. An edge directed from A to B means that B is one of the nearest neighbors of A.

The initial weight $\omega_A(c, k)$ does not entirely reflect the semantic closeness of the concept c and the term k. Although the computation described above gives us a hint towards its value, computing the closeness between a concept and terms is yet highly complex, if not impossible, primarily due to the different semantic complexity of the objects found in the ontology (high-level concepts) and the term-net layer (simple terms). According to the componential analysis described (Goodenough, 1957), each concept can be characterized by a set of

[1] The graph viewer was implemented using Touchgraph. See www.touchgraph.com.

semantic features (or components). Thus, concepts correspond to a whole set of terminological entities, each expressing the semantics of the concept, which to a certain degree is itself highly context-dependent. We therefore propose to learn the ω_A-values from implicit user feedback using a query-dependent neural network issued from the retrieval path in the core of our system.

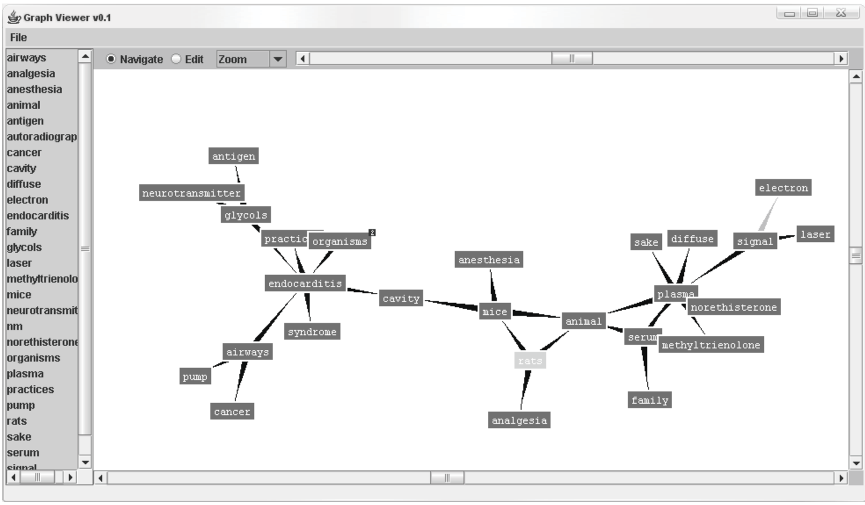


Figure 2. Excerpt from a similarity graph.

4. The Retrieval Process

The retrieval process consists of the following steps:

(1) Expand the user query by adding the description of the user's context to his initial query.
(2) Monitor some user interactions with the retrieved documents and generate a feedback value.
(3) Generate a neural network equivalent to the query path; propagate the feedback back into the network; and save the new adaptation weights. (See right side of Figure 1.)

4.1. *Query Expansion*

A query Q can be represented as a vector in both the vector space of terms and the vector space of concepts: $Q \in [0, 1]^{|Cr|} \times [0, 1]^{|K|}$. We define the partial vector of Q in the concept space as being $Q_C \in [0, 1]^{|Cr|}$ and the partial vector $Q_T = [0, 1]^{|K|}$ in the term space:

$$Q_C = \sum_{i=1}^{|C_r|} q_i^c \overrightarrow{c_{r,i}}, \overrightarrow{c_{r,i}} \in C_r$$

Q_C is identified by a naive string-matching algorithm that compares each of the keywords input by the user with the list of concept names. The vector C_r^t is added to Q_C by following linear combination:

$$Q'_C = (1-\alpha) Q_C + \alpha C_r^t$$

where α defines the degree to which the context is considered during the retrieval process. (In our tests α = 0.2.) The next step consists of transforming Q'_C into a vector in the term space using the adaptation layer:

$$Q'_T = \sum_{i=1}^{|C_r|} \left(q_i^c \sum_{j=1}^{|K|} \omega_A(c_i^r, k_j) . \overrightarrow{k_j} \right)$$

Q'_T is combined with the term vector Q_T originally derived from the query:

$$Q' = Q_T + Q'_T$$

The last step of the query expansion consists of adding word similarity knowledge to Q′. The term-net W can be described by an adjacency matrix ¡. First this matrix is normalized:

$$\Gamma'_{i,j} = \frac{\Gamma_{i,j}}{\sum_{i=1}^{|K|} \Gamma_{i,j}}$$

To avoid an over-activation of terms, the edges between activated terms are deleted:

$$\Gamma''_{i,j} = \begin{cases} 0 \; if \;\; q'_i > 0 \; and \; q'_j > 0 \\ \Gamma'_{i,j} \;\; else \end{cases}$$

The terms similar to those in Q′ are then added to Q′. Graphically, this is equivalent to spreading the activation of each term in Q′ to its closest neighbors in the term-net according to the relative weights of the edges between them. The query Q′ is eventually transformed into Q″:

$$Q'' = Q'(I + \Gamma''^T)$$

where I is the identity matrix and Γ''^T is the transposed of Γ″. The expanded query Q" is finally forwarded to a search engine that implements the vector space model (Salton et al. 1975) e.g. the Hummingbird Search Server[2], which is used in our prototype.

4.2. *Feedback*

The results of the retrieval described above are displayed in a modified browser window. Each of the documents in the data layer is segmented into paragraphs (so called "information chunks"), of which the most relevant are presented as retrieval results. Moreover, it is possible for the user to interoperate with the chunks. He can display the entire document or the chapter in which the chunk can be found, as well as the paragraphs before and after the one in which the chunk is found; go back to the last state of the window; zoom in or out of the document; and even close or download the entire document (see Fig. 3). While the user interacts with the chunks, each operation is mapped to a feedback value between -1 and 1, which is then cumulated and saved (together with the user context and the query) into a MySQL 5[3] database. When the database contains a specified number of entries, the learning process is triggered.

---

[2] http://www.hummingbird.com.
[3] http://www.mysql.com.

Figure 3. The modified browser window.

4.3. *Learning*

Our system "learns" using standard neural networks. The left half of Figure 1 shows an example of a retrieval path in our system. The network generated is always equivalent to a feed forward network with shortcut edges. Its exact topology depends on the user query where the input layer consists of both the concepts and the terms activated by the initial query. The two hidden layers consist of the other activated terms. The output layer is composed of the retrieved documents. Shortcut edges exist between retrieved documents and terms of the original query or terms directly connected to a concept because the two are separated by a hidden layer.

The activation of the document nodes in the network is proportional to the ranking given to the document by the search engine. Thus, using a standard learning technique to adjust the weights of the edges between the adaptation layer and the term-net should adapt the ranking in the sense of the user. We use the batch version of the standard back-propagation algorithm without momentum (Zell, 1997) to update the weights of the edges between the adaptation and the term-net layer, the input vector being the expanded query and the error vector being

$$\xi=\varpi(1+\varepsilon),$$

where ϖ is the vector describing the activation of each chunk node and ε the feedback values generated for each chunk. After data gathering, the query/feedback pairs are sequentially retrieved and used to dynamically generate and train an equivalent neural network. The network itself is implemented using the JoOne[4]-library. Since this library cannot directly handle shortcut connections (because of the layer-paradigm it uses to represent networks), we add temporary neurons to ensure correct connections between the neurons. The input synapse layer was modeled as linear; all others as sigmoid layers.

5. Results

The evaluation of our system is manifold. Firstly, we evaluated the algorithm computing the word similarity (CLIC). The evaluation was carried out in two steps. The pre-evaluation step was conducted manually on 50 randomly selected clusters that were computed using the "Corpus of Late Modern English Texts" (CLMET). We had two independent evaluators, both fluent in English. Both evaluated CLIC to perform 35% better than the standard algorithm k nearest neighbors (kNN). The automatic evaluation was carried out by comparing the clusters of similar words generated by CLIC and those generated by kNN on the OHSU-TREC-9 corpus for filtering with the GALEN ontology for diseases. The evaluation showed that our algorithm outperforms kNN in precision. As Figure 4 shows (the x-axis representing the different values of k and the y-axis representing the quotient of the precisions in percent), we achieve better clustering results with higher values of k relatively to kNN (robustness).

[4] http://www.jooneworld.com.

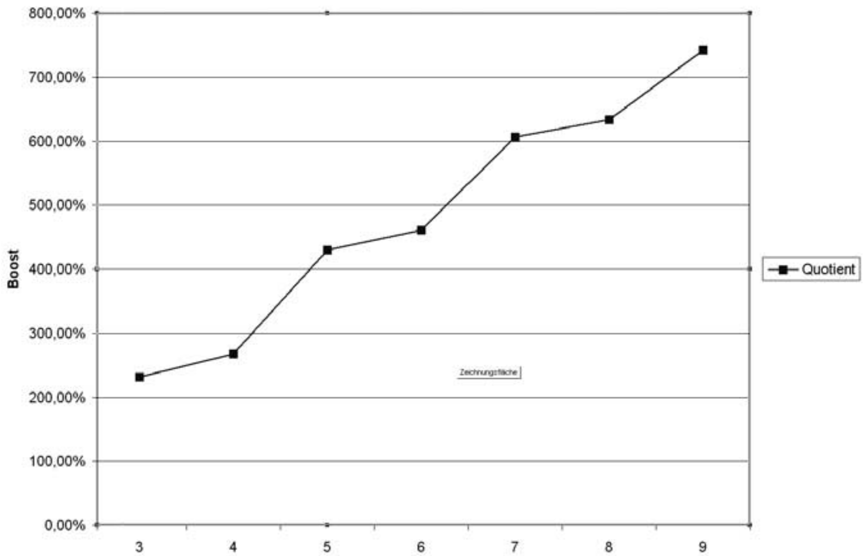


Figure 4. Relative comparison of the precision of kNN and CLIC.

Our entire system was evaluated using a corpus of firm-intern data on machine construction. We obtained better retrieval results with an augmentation of the precision by 0 to 30% (the mean value being 15%) using a small set of queries on 20 MB of data on machine engineering and without considering the feedback.

A complete evaluation of the feedback module could not yet be carried out. The use of neural networks guarantees an augmentation of the retrieval quality. We studied the evolution of document rankings, exclusively using positive feedback. Figure 5 displays our results: We considered the seventh information chunk (blue line) to be the document that suited our information request best. This was learned by the system after eight cycles. The document at the fourth position (yellow line) had the same vector description but was shorter. Thus it was considered to be more relevant because tf-idf favorises shorter documents. The document that was ranked first (red line) by the first retrieval was eighth after twelve learning cycles.

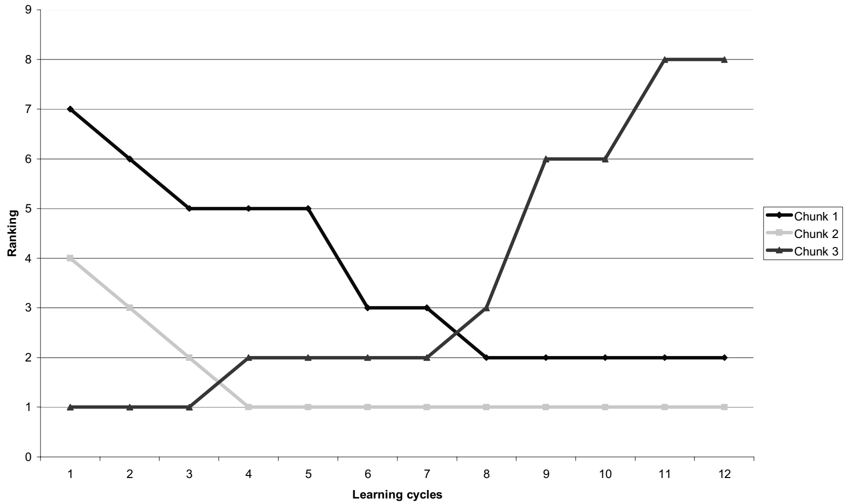


Figure 5. Evolution of document ranking with learning.

A long-time study will be necessary to evaluate the best values for parameters such as the learning rate, the impact of the feedback module on the precision and recall of our engine, and the best value(s) for the context parameter α.

References

1. Carpineto, C., de Mori, R., Romano, G., & Bigi, B. (2001). An information-theoretic approach to automatic query expansion. *ACM Transactions on Information Systems (TOIS), 19(1),* 1-27.
2. Claypool, M., Le, P., Wased, M., & Brown, D. (2001). Implicit interest indicators. In *Proceedings of the 6th International Conference on Intelligent User Interfaces* (pp. 33-40). Dublin, Ireland: ACM/Springer Press.
3. Goodenough, W. (1957). Componential analysis and the study of meaning. *Language, 32,* 195-216.
4. Kelly, D. & Teevan, J. (2003). Implicit feedback for inferring user preference: a bibliography. *SIGIR Forum*, *37(2),* 18-28.
5. Morita, M. & Shinoda, Y. (1994). Information filtering based on user behaviour analysis and best match text retrieval. In W. Bruce Croft and C. J. van Rijsbergen, editors, *Proceedings of the 17th Annual International ACM-SIGIR Conference on Research and*

Development in Information Retrieval (pp. 272-281). Dublin, Ireland: ACM/Springer Press.

6. Muehlhoff, T. (2004). Ein Konzern will wissen, was er weiss. In Petkoff B. Gronau N. and T. Schildbauer, editors, *Proceedings of KnowTech 2004* (pp. 45-53). Munich, Germany: GITO-Verlag.
7. J. J. Rocchio (1971). *The SMART retrieval system, experiments in automatic document processing*. Upper Saddle River, NJ: Prentice Hall.
8. Salton, G., Wong, A., & Yang, C. (1975). A vector space model for automatic indexing. *Communications of the ACM*, 18, 613-620.
9. Taghva, K., Borsack, J., & Condit, A. (1998). *The effectiveness of thesauri-aided retrieval.* Las Vegas: Information Science Research Institute of the University of Nevada. (Technical Report TR-98-01).
10. Voorhees, E. &d Hou, Y. (1992). Vector Expansion in a Large Collection. In Harman D.K., editor, *Proceedings of the TREC-1* (pp. 343-352). Gaithersburg, MD: National Institute of Standards and Technology.
11. Voorhees, E. (1994). On expanding query vectors with lexically related words. In Harman D.K., editor, *Proceedings of the TREC-2* (pp. 223-231). Gaithersburg, MD: National Institute of Standards and Technology.
12. Wang Y., Vandendorpe, J., & Evens, M. (1985). Relational thesauri in information retrieval. *J. American Society for Information Sciences, 36(1),* 15-27.
13. Zell, A. (1997). *Simulation neuronaler Netze.* Kirchheim, Germany: Oldenbourg.

Chapter 20

CAN BLOGGING BE USED TO IMPROVE MEDICATION ERROR COLLECTION AS PART OF HEALTH INFORMATICS KNOWLEDGE MANAGEMENT?

Deborah E. Swain
School of Library and Information Sciences
NC Central University, PO Box 19586, Durham, NC 27707
919-833-9028
dswain@nccu.edu

Blogs (web logs) seem to be growing in popularity and are a phenomenon of the Internet. Could blogging be an innovative tool for knowledge management (KM)? This paper looks at the concept of using blogs to store and share knowledge about healthcare. A general research question is whether information systems (IS) incorporating intranet-based or restricted blogs might provide an innovative, user-friendly method for identifying and preventing medical errors. This research looks at ways to uncover explicit and tacit knowledge about human-system situated processes associated with nursing care in a hospital. Information concerning errors must to be uncovered without threatening healthcare professionals in terms of ethics, legal problems or job security. To conduct the research, a blog was established in a protected, anonymous-posting system and was provided as an interactive medium where novice nurses could share information and develop knowledge. The study provided proof of concept and showed that blogs can support communities of practice (CoP). Therefore, web-based sharing (blogs, wikis, and forums) might provide effective tools for discovering medical-errors. The collected information and data then could be used with analytical tools to create knowledge-based reports that improve procedures and provide quality assurance. Blog data was analyzed for information concerning procedures, errors, quality assurance, and job definitions. The research showed there are potential benefits in regularly using blogging tools in a medical environment (or other domains) as a framework for KM. Blogging can support sharing and reusing personal knowledge to improve applications, enterprises, and human-system environments.

1. Introduction

As a form of knowledge management (KM), a blog is a Web page "with reverse chronological sequences of dated entries, usually with sidebars of profile information and usually maintained and published with the help of a popular blog authoring tool" (Kumar et al., 2004). A Web log user (a "blogger") creates an internet-based journal and/or responds ("posts") to statements by someone else on the Web. Blogs provide a specific form of personal communication with the public. On the Internet, they have even become a telecommunications channel for the mass media (newspapers, magazines, and television). However, blogs are not just a one-way presentation of personal opinions, events, or interests; they can also be an interactive medium for a community.

Health informatics is defined as "the systematic application of information and computer science and technology to public health practice, research, and learning" (Yasnoff, 2000).

The project described in this paper investigated blogging as a KM tool for a healthcare environment in the United States, where automated information systems (IS) are being introduced to new hospitals and clinics at a fast pace. Emerging issues associated with healthcare management and improving medical care worldwide make this research seem pertinent and timely.

Furthermore, as O'Carroll (2002) has defined capability or competency levels for IS in public health, the second level of competency in health informatics is the management of information technology projects to improve the effectiveness of care, measurement of quality, or for research. Many medication errors are not reported in full or critical elements may not be captured due to traditional, labor-intensive reporting mechanisms and/or data entry techniques. Yet, quality improvements and effective cost cutting in healthcare depend on the accurate analysis and continual correction of errors in treatment and care.

2. Purpose

The application of KM in medical and healthcare environments is a growing field. The assumption to be proved or disproved is that restricted-access blogs can be useful in KM health informatics at hospitals. The first objective of this study was collecting blog data from student nurses at a hospital, for one month, in order to analyze the text for health informatics and communications data about medical error collection and process improvements. Each participant was a consenting volunteer, and all names and identities were kept confidential. This study gave the investigator a chance to validate and verify the data collection methodology and the concept of blog use by healthcare professionals.

As Steve Cayzer (2004) discovered at Hewlett-Packard, blogs can be the basis for a semantic network and provide a tool for decentralized, informal KM. In business and other environments, regularly using blogging tools can provide a framework to share and reuse information and data across applications, enterprises, and user communities. In healthcare facilities, specific standards and practices must be followed, but processes are constantly revised as part of the learning environment associated with quality assurance in medical care. Hospitals and clinics must be learning organizations to survive as businesses and to attract both patients and healthcare workers.

Because of the emphasis placed on regulations, charts, and procedures, it was proposed that a blog's chat-like environment might help staff share both explicit and tacit knowledge more easily. Volunteers were recruited, a blog with limited access was established, and the healthcare workers were invited to share questions and discoveries. (Note: all volunteers were interested in new ways to conduct quality assurance.) The purpose of the study then was to see if a blog can be an innovative tool for sharing medical error knowledge. In the future data mining and analytics could be applied to the data collected from restricted, confidential blogs.

3. Significance

The rising cost of medical care has created an emerging issue that requires the improvement of knowledge concerning healthcare practices that can provide better quality care with cost-efficiencies. Estimates show that 70 percent of injuries in hospitals and clinics are the result of medication errors (Burns *et al.*, 2004).

The need for better medical care information and communications systems is apparent. Many healthcare professionals fear reporting medication errors or find documentation difficult. In the nursing profession, reasons not to write up current reports on errors include lack of readily available forms, no time on a busy shift, being unskilled with fax machines or emailing, or simply forgetting after a shift. Evidence of such gaps in data collection suggests the need for a user-friendly method to identify human-system situated errors in hospitals and clinics. If collected information does not include explicit and tacit knowledge, it will not be possible to do thorough root-cause analysis in follow-up procedures for nursing knowledge.

4. Theoretical Framework

The cognitive and psychological approach to computer-based KM systems emphasizes how humans interact with machines for communication. Understanding such interactions has led to conceptual frameworks for machines (computers) that mirror human processing. For example, the Model Human Processor (MHP) presented by Card, Moran, and Newell in the early 1980s at Carnegie Mellon used computer metaphors to understand how humans process information. The goal was to design better interfaces on the computer that help people complete tasks more easily.

Concerning blogs and KM, the research on online personal logs is in a period of early theoretical formulation. The variety of blogs and number of users is astounding. Blog research can involve as many as one million blogs worldwide (Kumar et al., 2004). However, they are a form of communication that allows all levels of stakeholders to contribute ideas about processes and systems. For more than 20 years, group interactions and strategies for improving communication effectiveness have been used in the design and development of information systems (Carroll, Thomas, and Malhotra, 1980; Guinan, 1986; and Walz, 1988). Given that medical error detection and correction are an important aspect of health care systems as learning organizations, the theoretical question is whether or not communications research on blogs can be applied to KM in health care environments or other learning organizations.

5. Research Method

A qualitative data-gathering approach based on sense making (Dervin, 1992) was used as the research method of choice. Nine participants were recruited from a volunteer student nursing class. Overtime, 23 users signed in and established usernames. Seven of the nine vounteers (78 percent) contributed regularly to the blog. Others read and downloaded information but did not visibly participate as bloggers. All volunteers were instructed to use the blog freely to find answers or close knowledge gaps on any subjects that might come up. The blogs were their personal, community journals. Entries from the study's blogs were collected for one month. Qualitative analysis was done weekly on entries to determine categories of topics and areas of knowledge development.

The blog utilized a blog "reserved space" server. Space reservation allows users to edit and collaborate on the same content, but restricts access to members only. Service was supported by a network-focused company as opposed to an operating system-based company (Gilster, 2006). The process of making blog entries and posting responses was similar to posting articles in an online, quick-encyclopedia or "wiki."

The users established nine subject categories or information "topics" for their postings:

- "Support, please."
- Challenging Patients
- Help Finding a Job.
- Job Contacts
- Medications
- NCLEX
- Patient Rights
- Public
- Test Reviews
- Uncategorized

Profile analysis of the blog data led to the classification of postings into three types:

- Concerns and curiosity about the *tool* and blogging as an act.
- *Medical topics* of interest to the bloggers.
- General learning and sharing of *professional* news and advice.

Specific information about process and qualitative data were collected through textual analysis of all the blog entries and comments.

6. Data Collection

More than 30 postings per week were collected as text. The data included initial blog entries, attached documents, and any comments. Weekly collections were made. Data that was not collected for analysis included news on tool upgrades and appreciation to subjects for participation from the Principal Investigator.

After initial text files were manually analyzed (week one), the data was classified into three basic functional types as described above. Then, ongoing sorting and pattern mining of the textual data was done at the end of weeks two and four. A table was constructed from the collected data for iterative analysis of the month-long blogging. See Table 1 for excerpted samples from each category.

6.1. *Data Analysis*

Using the text-based table for qualitative data analysis, the investigation showed a tendency towards knowledge sharing both on a daily cycle and long-term during the month-long study. Primarily two knowledge management concepts were verified:

- A repeated information flow towards Knowledge Sharing.
- Growing into a learning-based Community of Practice.

Table 1 shows blog entry samples of each topic area and the researcher's early analysis comments on the study. [The numbers in brackets represent coded IDs of participants.]

From the beginning, the study group showed they belonged to an already existing learning-based organization or community of practice (CoP) that was integrated and self-defining. Blogging entries illustrated a tendency towards information exchange that led to sustained knowledge sharing (see row 2 in Table 1). As an aggregate the group also developed as a community of practice (CoP). They enhanced their blogging activity by also sharing nursing certification test documents to use for studying. [It should be noted that the subjects were student nurses and part of a learning community before and after the study.

Table 1. Textual data analysis table.

TOPIC	Original Posting	Comments on Posting	Analysis
Concerns and curiosity about *tool* and blogging as an act.	3/21: “first blog attempt” Let’s see if this works, I’m getting so hi-tech it’s scary!! [5]	3/21: I think I messed up my username….sigh. [1] 3/21: Yay!!! I fixed it :-) [1]	Starting a CoP
Medical topics of interest to the bloggers	3/22: …The RN from the previous shift stated in report that she had not applied the patient’s nitro paste when it was scheduled, but said the patient wasn’t complaining of any chest pain. The patient was also taking Lasix and Toprol. On assessment the patient was experiencing shortness of breath and had extremely wet lung sounds, and was having minimal urine output. After I applied the nitro paste, the patient dramatically improved within 1 hour, with clearer lung sounds, less difficulty breathing, and high urine output… In this case the Nitro paste wasn’t ordered for chest pain, but was designed to increase cardiac output and kidney perfusion. [5]	3/30: Excellent catch! Shows how a drug ordered for one reason, actually has multiple effects on many other body systems…. This stresses the importance of timely med administration also. Thanks for the info. [3] 3/30: Wonder **why** she felt she could hold the med when the order didn’t expressly state it was for CP?? Hmmmmmmm…Great lesson for us all!! [1] 3-31: This is a lesson for all of us. It is evidence that the meds we give are not just because they are ordered and on the MAR, but because patients DO actually NEED them and for different reasons…[6]	Clear <u>knowledge</u> development about procedures for medication delivery Self-editing to retain sense of discovery for whole group/class. Brainstorm as <u>COP</u> Liberating <u>learning</u> as stand up for procedure

Table 1. (*Continued*)

TOPIC	Original Posting	Comments on Posting	Analysis
General learning and sharing of *professional* news and advice	3/23: … Anybody know any good contacts at XX Hospital? … I saw a few jobs at XX that looked interesting too. Too many possibilities! What a 'nice' situation to be in. :-) [3]	3/23: Sorry…the only person I know at XX is in security :-) [1] 3/24: My colleague in Nursing …has contacts at XX. Feel free to call …[4]	Information Gap Query about jobs by professional; shows *COP*

7. Social Network Analysis

Communication threads of networking were tracked using a basic Social Network Analysis (SNA) method. During week one, five participants exchanged information and interacted by posting 27 blogs. Before the end of the study, seven bloggers participated. Fourteen original entries were posted, and 22 comments were made on blog entries. See Figure 1 for illustration of the network established.

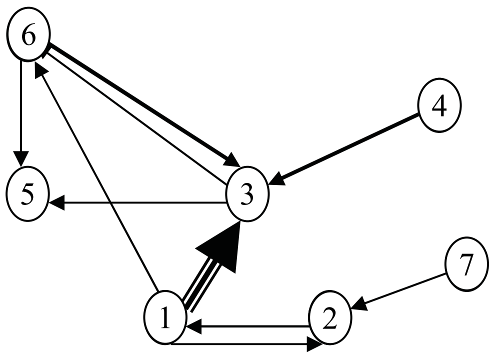


Figure 1. Social network analysis of blog comment activity.

Early analysis of communication patterns and the social network identified two very active knowledge participants posting original entries and commenting on entries. The majority of subjects only participated about five times. However, participants No.1 and No.3 sent comments and received responses at least 12 times each. Blog comments sent by subject No.1 represent 45 percent of the total responses to entries. Subject No.3 received 40 percent of the comments to six original postings and one response to a comment on the subject's own posting. Their nodes can be seen to be the main nodes in the network based on comments. Subject No.1 sent 10 comments, including one to own comments. Subject No. 3 is the primary knowledge gatekeeper, making six original postings and receiving nine comments. Furthermore, the strongest binary relationship on the blog network was between subjects No.1 and No.3 (making six connections to each other). Data suggests that 20 percent of team committed and excited about blogging.

8. Results

Research on medical error collection might include discovery of both explicit and tacit knowledge as shown by this study. Blogs have potential as data interfaces for capturing explicit knowledge as well as tacit understandings. The project described here was only a small study developed to validate the concept of blogging in the medical community. The results were positive. Lessons were learned about the application of blogging by healthcare professionals:

- Protect participants' privacy (Anonymous postings are required to assure participants' privacy and to support free discussion.)
- Ensure integrity of postings (More work is needed to be sure postings are not edited by others; this is a tools development issue.)
- New nurses who are familiar with using automated information systems in hospitals (not trained or loyal to paper-based charts and forms) enjoy sharing knowledge on blogs.

More research is appropriate and should include nursing graduates. This study's results show that tacit knowledge can be exposed in a blog and learned by others. For example, a blogger shared the following tacit knowledge when logging information about how quickly sepsis can take over: "[I] am struck by the importance of noticing subtle changes in your patient's loc, arousability, gut instincts, etc in order to quickly catch and correct a problem." The nurse spent a full day from morning to midnight monitoring this patient and took the time to document in the blog explicit details about the drugs used and hourly changes in patient status. Not only learning the explicit procedures used, a commenting blogger recognized that tacit, hidden knowledge had been shared: "…and sometimes it's our observations that make the difference in the patient's outcome. What an awesome responsibility we have!!"

The two "errors" logged into the blog were committed by healthcare workers not participating in the study. This may suggest that bloggers will hesitate to list their own errors. At least, anonymity allows sharing and informative discussion of errors. For first reference to a possible error, see row 2 in Table 1, and discussion about unexpected effects. The second "error" involved human-to-human behavior. After a patient under the blogger's nursing care over heard a rude comment in the hall, the patient was described as being livid and "upset that others in the hall had also heard him. She said she had never been so disrespected as a person! She had a few choice words for the tech…." The blogging community shared the indignation in comments, such as, "I have seen some things, and heard some things from nurses I work with that would make your 'hair curl'… not everyone in this business is equipped with compassion and patience." Both lessons seem to have been learned by the bloggers and documented in knowledge sharing comments: "Excellent catch!" and "I think you handled it well."

The results of this study suggest there is potential for using blogs to store and share knowledge about healthcare and procedures.

9. Discussion

KM can be utilized to improve healthcare quality. Various Information Systems (IS) currently gather error and near miss data in medical environments. Some of these established procedures for medication error incident reporting introduce extraneous variables that were not accounted for in this health informatics study because our research looked at web-based data and analytics. Handwritten forms, for example, are used by many hospitals, clinics, and healthcare facilities, and they introduce variables to an IS that are not related to computer-to-computer applications. Throughout the United States, reporting and data collection procedures for health records in hospitals are being revised as new IS systems and electronic, relational databases are built to serve the industry. New health informatics systems might be able to integrate data from blogs into knowledge bases.

Blogs seem to have potential as a KM innovation for improving the quality of healthcare. Blog text records similar to the data collected during this research could be fed into a knowledge-based repository. Then analytics of keywords could sort medication error incidents and near misses. Reported results might provide nursing and healthcare managers with useful knowledge about preventing future errors. As Holt Anderson, Director of the non-profit North Carolina Healthcare Information and Communications, Alliance, Inc. (NCHICA) has noted, "We need research to promote the reporting of medical errors" (1/25/2005). He believes fear of employment repercussions and malpractice suits holds many back. Such behavior is a major issue in the healthcare field that this study will address.

The long-term KM goal of this study is to contribute to refining health informatics practices, to designing healthcare IS, and to discovering innovative ways to improve healthcare quality. Healthcare practitioners might use knowledge developed from blog data to prevent errors in medication administration and nursing procedures. In conclusion, the analysis of blogging has potential as an innovative KM tool in medical healthcare and perhaps other domains.

References

1. Anderson, H., (2005, January 25). *Interview in research Triangle Park.* Paper presented at the meeting of the North Carolina Healthcare Information and Communications, Alliance, Inc. (NCHICA), North Carolina.
2. Boettger, S., (2005). Compliance & knowledge management IT Toolbox, *KM BLOG,* Retrieved February, 2005 at http://blogs.ittoolbox.com/km/pragmatics/.
3. Burns, M., N. Perin, W. Wei, K. Ames, T. Clark & A. Molnar, (2004). *Increasing medication error reporting in the United States.* Duke University Advisor, Arthur Spengler (Policy Memorandum to NCHICA).
4. Card, S., Moran, T., & Newell, A. (1983). *The psychology of human-Computer Interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.
5. Carroll, J. M., Thomas, J. C., & Malhotra, A. (1980). Presentation and representation in design problem solving. *British Journal of Psychology*, 71, 143-153.
6. Cayzer, S. (2004). Semantic blogging and decentralized knowledge management. *Communications of the ACM*, 47(12), 47-55.
7. Dervin, B., (1992). From the mind's eye of the user: The sense-making qualitative-quantitative methodology. In Glazier, J., & Powell, R. (Eds.), *Qualitative Research in Information Management* (pp. 61-84). Columbia, MO: Libraries Unlimited.
8. Efimova, L., (2003, May 23). *Blogs: The stickiness factor.* Paper presented at Blog Talk: A European Conference on Weblogs, Vienna, Austria.
9. Gilster, P., (2006, March 8). Collaborate or start a soapbox. *The News and Observer*, 3D.
10. Guinan, P. J. (1986). *Specialist-generalist communication competence: A field experiment investigating the communication behavior of information systems developers.* Unpublished doctoral dissertation, Indiana University, Bloomington, IN.
11. Herring, S., L. Scheidt, S. Bonus, & E. Wright, (2004, January 5-8). *Bridging the gap: A genre analysis of Weblogs.* Paper presented at the meeting of the 37th Annual Hawaii International Conference on System Sciences, WHERE?
12. Kumar, R., J. Novak, R. Raghavan, & A. Tomkins, (2004). Structure and evolution of blogspace, *Communications of the ACM*. 47(12), 35-39.

13. Nardi, B., D. Schiano, M. Gumbrecht, & L. Swartz, (2004). Why we blog, *Communications of the ACM*, 47(12), 41-46.
14. O'Carroll, P. W. (2002, August). Informatics Competencies for Public Health Professionals. Paper presented at Northwest Center for Public Health Practice, NEED CITY.
15. Walz, D. (1988). *A longitudinal study of the group design process.* Unpublished Dissertation, The University of Texas, Austin.
16. Yasnoff, W., P. O'Carroll, D. Koo, R. Linkins, & E. Kilbourne, (2000). Public health informatics: Improving and transforming public health in the information age. *J Public Health Management Practice* 6(6), 67-75.

This page intentionally left blank

Chapter 21

FACE WORK AND CREDIBILITY ON THE WEB

Takashi Nakamura
Faculty of Humanities, Niigata University
2-8050, Igarashi, Niigata 950-2181, Japan
+81-25-262-6452
takashi-nakamura@human.niigata-u.ac.jp

Hiromi Yuki
Graduate School of Modern Society and Culture, Niigata University
2-8050, Igarashi, Niigata 950-2181, Japan
+81-25-262-6452
z05c022a@mail.cc.niigata-u.ac.jp

In this paper, the authors applied Goffman's sociological theory of "Face Work" to an analysis of the Web. The authors stated that a user's ability to judge the credibility of related sites may be enhanced by examining the hyperlinks named on link pages and comparing those to the ones named on other pages. It was discussed that hyperlinks on link pages offer distinct meaning. The authors analyzed only Web communities that contained hyperlinks on link pages, noting the distribution between the number of in-coming hyperlinks (in-degree) and the number of out-going ones (out-degree). The reconstructed Web community indicated that no site existed with a high in-degree, as well as a high out-degree. The variation of the evaluated values of the link pages (e.g. the in-degree point) suggested that credible sites could be hyperlinked to other credible sites through link pages. The results of the two analyses suggested that a credible site (an authority) carefully selects the hyperlinks named on its link pages. This qualified selection could be founded on the site administrator's Face Work, which strongly constrains his or her observable conduct. The authors suggested that Face Work could be valid on the Web. They claimed that the observable credibility implied by related sites could be utilized in order to obtain knowledge about an unfamiliar topic.

1. Introduction

At present, browsing through pages or sites on the World Wide Web (WWW) is a popular way to obtain information; however, not all information discovered on the WWW is useful. Therefore, we must establish a way to discern between available and unavailable Web pages. The credibility of a Web page is one of the most important aspects for classifying it. Yet, there is no standard criterion to classify pages or sites when seeking information about an unfamiliar topic. Therefore, in order to aid our judgment about source reliability, we must recognize universal clues. Firstly, we suggest using the link structure between related pages or sites; alternately, we suggest utilizing the result of a ranking algorithm. In both cases, hyperlinks (especially those to related pages or sites which appear to be well acquired) offer clues to credibility.

When tracing a page or site link structure, the organization of its associated Web community is utilized to ensure credibility. Web communities are clusters of Web sites with similar interests that are linked together with hyperlinks. By graph theory, Web communities can be characterized by dense directed bipartite sub-graphs where member sites and hyperlinks are perceived as nodes and links (Kleinberg, 1999; Kleinberg, Kumar, Raghavan, Rajagopalan, & Tomkins, 1999). Previous studies demonstrate that authority weights and/or hub weights may be assigned to Web sites with regard to their topics. Kumar, Raghavan, Rajagopalan, and Tomkins (1999) comprised a list of 100,000 emerging Web communities by using a trawling method. Research for identifying Web communities has supposed that the sites of these communities can be divided into phases. Thus, researchers succeeded in designing a feature of the WWW, although they did not explain the emergence of such in order or hierarchy. If hyperlinks between related pages or sites are traced and rooted to an authority site, confidential information could possibly be discovered. Searching in such a way implicitly utilizes the structure of a Web community. However, searching should be distracted so that only a few sites favoring a particular subject would randomly compile the source files of the related pages or sites and automatically make hyperlinks to them. Thus, it is still unclear whether the authority site that is comprised of hyperlinks from all community hub sites is or is not credible.

The results obtained from the ranking algorithms used in many search engine sites can be utilized to rank Web sites by using rational numbers. For example, the values calculated by the Page Rank algorithm and the extended metrics are the probabilities that a random surfer visits the pages. Thus, a hyperlink would be interpreted as a directed path that is chosen during the random surfing (Brin & Page, 1998; Haveliwala, 2002; Jeh & Widom, 2003). Certain tools have realized efficient and effective information retrieval from the Web. It is possible that higher ranked sites would be credible since they serve as hyperlinks with many other sites. However, the number of hyperlinks from the other sites does not directly reflect the reliability of a site; furthermore, not all the pages or sites with higher ranks can be credible. The implicit structure of the Web community and the result of the ranking algorithm can possibly be useful to a certain degree; however, they cannot assure the credibility of related pages or sites.

The desirable order of Web pages or sites should more directly reflect their credibility and include the mutual evaluations between related pages or sites. Moreover, the order provided should be justified. In this paper, we focused on a Web site link page and discussed the order of the related Web sites. We applied the "Face Work" sociological theory (Goffman, 1967) to explain why hyperlinks are used in link pages and why the related pages or sites need to be ordered. We reconstructed Web communities that contained only hyperlinks on link pages and verified the efficiency of the sociological understanding.

2. Hyperlinks on a Link Page

2.1. *Characteristics of a Link Page*

The authors focused on Web site link pages because these offer unique characteristics, and because each hyperlink on a link page has a distinct meaning. On a page other than a link page, the purpose of hyperlinks is to enrich partial descriptions, e.g. words, phrases, symbols, technical terms, pictures, and photographs. Furthermore, these descriptions enrich the overall content of the page (Figure 1, right). In order to do this, the administrator should evaluate the pages to be linked and determine whether or not it is appropriate to create hyperlinks from his or her own page. This is true except in the case of automatic editing.

Contrary to this, a link page is edited in order to link it to related sites. It is not necessary to enrich the content of the link page since it consists exclusively of the list of hyperlinks. In fact, hyperlinks on a link page are obtained from the link page itself; however, they are created in almost the same manner as those from the entire site (Figure 1, left). The administrator of the link page would evaluate the entirety of site to be linked. He or she would then judge whether the entire content of the site alone is sufficiently appropriate to create hyperlinks from his or her site. A critical difference between hyperlinks found on link pages and those on other pages is that those on a link page link the entire site content to the entire content of a related site.

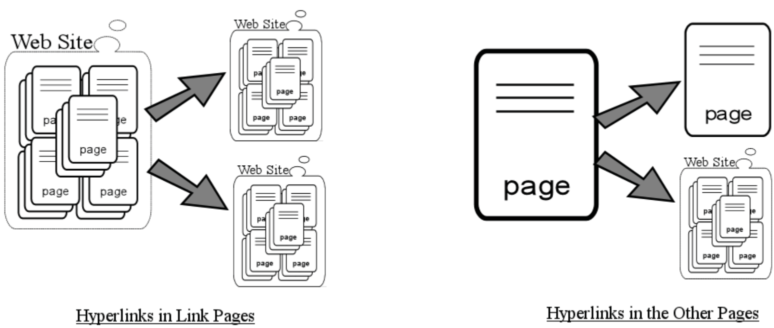


Figure 1. Difference between link pages and the other pages.

2.2. *Power-law Distribution and the Scale-free Model*

The number of in-coming hyperlinks (i.e., the in-degree) and the distribution of sites in the entire WWW follow a power law (Broder, Kumar, Maghoul, Raghavan, Rajagopalan, Stata, Tomkins & Weiner, 2000). Barabási and his colleges formulated “the rich get richer” model (the scale-free model) to simulate the growth of a linked network in order to satisfy the power law (Barabási, 2002). The model starts from a small seed network structure. In each time step, a node representing a Web site is increased. This new node links to a finite number of older nodes. The older nodes are selected according to the proportion of the weights of their in-degree values, which denote the number of in-coming hyperlinks. In time, the in-degree values and the distribution of the nodes follow the power law. This model, then, simulated the growth of the Web structure that satisfies a power law, clarifying the fact that older sites on the Web are more advantageous than newer ones. However, it is critical to note that in this model there is no difference between the hyperlinks on link pages and those on other pages. The distribution of the number of in-coming hyperlinks satisfied the power law; however, the hyperlinks from both link pages and others were investigated simultaneously. A hyperlink that enriches a partial statement or a figure on a page other than a link page may be irrelevant to the evaluation of the entire site. The richness in this case does not necessarily imply credibility.

Hyperlinks on a link page can be attributed to the site administrator's evaluation and confidence in related sites. Thus, utilizing a hyperlink found on a link page is important for evaluating knowledge available on the Web. In the following analysis, the authors focused on hyperlinks on link pages.

3. Goffman's Theory

3.1. *Stigma*

The authors applied Goffman's "Face Work" theory to analyze the Web. Goffman is a respected sociologist who focused on personal physicality and particularity. He believed that an individual's conduct is strongly affected by his or her personal physicality and particularity. In his famous work "Stigma" (Goffman, 1963b), he discusses the manner of accepting inborn physical specificity (e.g., physical handicap) and an acquired disposition (e.g., psychopaths or rap sheet). He claimed that in many cases, such individuals' characteristics generate the difference in norm, and this in turn drives them toward an exceptional position.

The Stigma theory can be considered to be a critique of the theory of Durkheim, who is recognized as a great initiator of sociology. Durkheim concluded that “personality” (personne) was the crucial kernel or core of human society (e.g., Durkheim, 1912). He considered personality to be general and uniform among all individuals, and it was this requirement that enabled both the participation in human society and the power to control an individual’s actions. The objective of his theory was the organization of human society; however, this theory was not sufficient to explain the behavioral norm influenced by the social relativity between individuals.

To the contrary, Goffman insisted that an individual’s characteristics are deterministic. They were neither general nor uniform. To enforce or strengthen his opinion, he considered radical subjects e.g. physically challenged individuals, psychotic persons, and offenders. Their individual characteristics would often cause other individuals to either be too polite or too critical toward the radical subject. Even more so, the radical subjects would be treated with a certain degree of bias by others regardless of their personalities. Social conducts should be conditioned by the characteristics of an individual.

3.2. *Face Work*

Inequalities can emerge in situations other than the radical subject observations discussed above. The Face Work study revealed that the behavioral norm of individuals strongly depends on the characteristics of the person and their social relations (Goffman, 1967). In this case, the “face” implies the portrayal of decency and honor. It includes the acceptance of an individual’s proper conduct according to their social position and relation to specific others. Goffman advocated that each person accepts another person’s behavior and determines his or her personal action for as long as the face of every person requires acknowledgement. Disrespecting someone’s face is considered to be ill-mannered. The behavior of each individual should be restricted according to his or her social position and relation with others. Occasionally, the behavior of an individual could be a ritual carried out in order to maintain his or her social relationship with others, thus not disregarding someone's face. Durkheim (1912) stated implicit ritual conduct to honor absolute existence; however, Goffman emphasized explicit ritual conduct to respect relative positions and relations between persons.

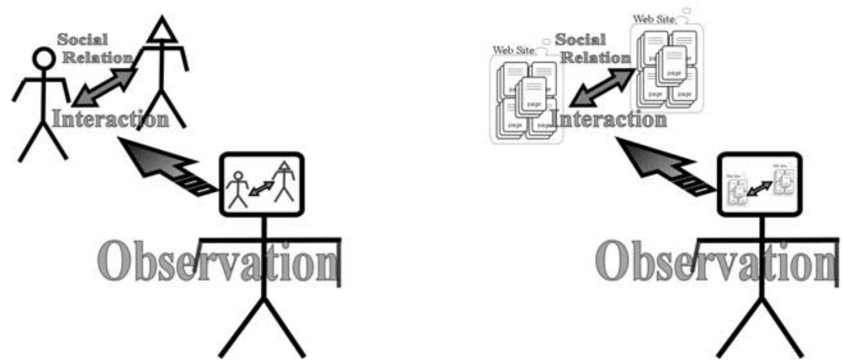
In the Face Work theory, "demeanor" implies that an "actor" succeeds in taking a well-mannered action so as not to disregard another's face. In other words, the actor's face can be acknowledged as long as he maintains his or her demeanor. In this paper, the authors applied the concept of demeanor to the analysis of link pages on the Web. If the hyperlinks on a link page include the site administrator's evaluation, then hyperlinks to a more credible site should be selected over those from less credible sites. Failure to do so would diminish the actor's face. Thus, a credible site would appropriately select credible hyperlinks which were determined by the evaluation of its own and related sites.

4. Social Behavior on the World Wide Web (WWW)

4.1. *Social Interaction in a Concurrent Space*

When a particular space is shared by many individuals the interaction between any two is observed by others (Figure 2, left). A person should first observe the interaction between people in the space before he or she acts. An observer would—rightly or wrongly—suggest that the social relation between the individuals observed should govern the nature of their interaction. The observer would then decide their attitude and behavior depending on his or her conjecture. In turn, the observer would also be observed and evaluated by others (Goffman, 1963a) and expected to behave in accordance with his or her own social position so as not to confuse other persons in the shared space.

The Face Work theory emphasizes that an observer is also an actor and that he or she should behave in accordance with his or her social position. In particular, a respectable person must select his or her action very carefully in a shared space. If he or she behaves inappropriately in the space, an observer could misinterpret the interaction. The observer would then decide his or her action depending on their ill-founded apprehension and perhaps act incorrectly. The subsequent, ill-mannered action could be observed by yet another observer, and this failure would generate another failure. A respectable person must remain well mannered so as not to confuse other individuals in the shared space. An actor is nearly compelled to understand his or her social position accurately and maintain his or her demeanor. However, physical actions can be interpreted in various ways, thus often be misinterpreted. Therefore, a person may select a ritual action in a shared space in order to avoid embarrassment (Goffman, 1967).



Observation in a physical space   Observation in the WWW

Figure 2. Observation of the interaction in a physical space and in the WWW.

4.2. *Hyperlink as an Interaction*

Our study applied the Face Work theory to analyze a Web community. In the WWW, individuals who enter a “space” can act as both actors and observers. They may consider hyperlinks as interactions between Web sites (Figure 2, right image). They may also interpret the hyperlinks on the link page of a credible site’s as the administrator’s deference.

A credible Web site must be managed appropriately in order not to confuse observers. The content of a credible Web site must be chosen in a manner that befits its authority. The interaction with other Web sites, which includes creating hyperlinks, must also be conducted in a similar manner. If the link page in a credible site is edited wrongly, the related sites would lose their face. Thus, the choice of hyperlinks could be interpreted as Face Work. In fact, the hyperlinks on a link page are expected to reflect the decision of the administrator.

An administrator of a credible Web site must select hyperlinks cautiously so as not to be misunderstood. If he or she is to maintain a respectable demeanor, he or she must "defer" to other credible Web sites by creating hyperlinks to them. In addition, he or she must not create links to any less credible Web site even if that Web site has linked to their site. The choice to link or not to link should be made carefully and in accordance with the social position presumed about him or her. On a credible Web site, therefore, a link page could reflect social relations, as well as the administrator's evaluation.

5. Analysis

5.1. *Two Suppositions*

We propose to apply the Face Work theory in order to analyze hyperlinks on link pages. As mentioned above, a credible Web site is expected to provide links on its link page to other credible Web sites. It must not provide hyperlinks to less credible sites. In doing so, other credible sites in the Web community would lose their face, which in turn could cause the marring of its own face. The administrator of a credible site must carefully observe the contents of other sites, accurately portray his or her social position, and appropriately select hyperlinks to the other credible sites.

It is important to note that a less credible site could provide hyperlinks on their link page to both credible and less credible sites, providing hyperlinks to any number of sites because his or her ill-mannered behavior would not confuse those persons on the WWW. The list of hyperlinks on a link page of a less credible site might be extensive and innumerable. In comparison, however, the hyperlinks chosen by an administrator of a credible site are more restricted than those of a less credible site. The choice constraints on the hyperlinks of a link page on a more credible site are greater than those of a less credible site.

Based on this argument, two suppositions can be made:

1. The more credible a Web site, the smaller the number of Web sites to which it can provide hyperlinks.
2. Web sites linked to a credible Web site are also expected also to be credible.

The following two analyses assured these two suppositions.

5.2. *Data*

In June 2005 and for the purpose of writing this paper, the authors analyzed two Web communities: one themed "environmental problems" and the other "atomic energy". The Web sites related to these themes were collated by a link page search on Google, beginning with the Web sites of many famous NGOs. Only Japanese Web sites were selected, and portals were excluded. We first assumed that the main purpose of link pages was to offer a referral list of Web pages. We selected link pages among Web sites that listed pages related to similar topics. The Web communities were rebuilt using only those hyperlinks that were listed on their link pages.

6. Results

6.1. *Number of In-coming and Out-going Hyperlinks*

The first analysis focused on the distribution between the number of in-coming and out-going hyperlinks (i.e., between the in-degree and out-degree) in link pages. Figure 3 depicts the distribution of in-degrees and out-degrees on link pages for all Web community sites that are associated with environmental problems and atomic energy. The distribution showed that Web sites with large in-degrees did not have large out-degrees. This implies that more credible Web sites can be hyperlinked to fewer Web sites; therefore the first supposition is verified.

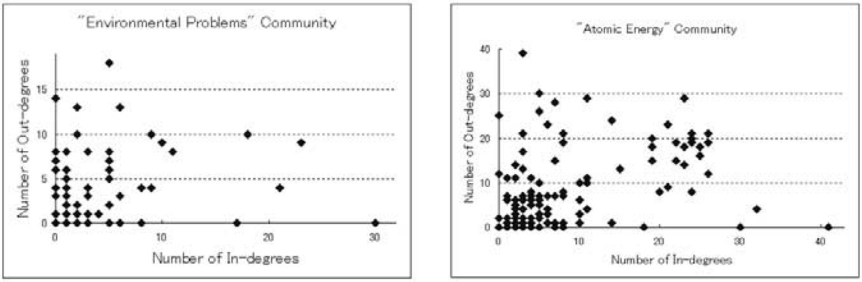


Figure 3. Distribution between in-degrees and out-degrees.

6.2. *In-degree Point*

The second analysis investigated the variation of the evaluated values for the number of in-coming hyperlinks on link pages. The evaluated value Q, which the authors call the in-degree point, was calculated as the averaged sum of the in-degrees of the linking Web sites as follows:
$Q(k) = (1 /Bk)\Sigma((Ai /Di)\Sigma LjiNj)$,

where $Ai = 1$ if the in-degree of the i-th site is equal to k, and $Ai = 0$ if else,
$Di = \Sigma Lji$, if $\Sigma Lji > 0$, $Di=1$ if $\Sigma Lji =0$,
$Lji = 1$ if the j-th site links to the i-th site in the link page, and $Lji=0$ if else,
Nj denotes the in-degree of the j-th site,
$Bk = \Sigma Ai$, if $\Sigma Ai > 0$, $Bk =1$ if $\Sigma Ai =0$.

$(1/Di)\Sigma LjiNj$ reflects the credibility of the i-th site. Di,Lji and Nj denote the number of the sites which have hyperlinks on the link page to the i-th site, the connectivity of the hyperlink from the j-th site to the i-th site, and the number of in-coming hyperlinks (the in-degree) of the j-th site, respectively. Since Nj indicates the number of sites that trust the entire content of the j-th site, $(1/Di)\Sigma LjiNj$ indicates the magnitude of the confidence of the sites that have hyperlinks on link pages to the i-th site. Thus, the value of $(1/Di)\Sigma LjiNj$ increases, if the number of credible sites making hyperlinks to the i-th site is increased. The averaged value Q(k) indicates the credibility of sites with an in-degree equal to k. An example of a network structure and the obtained in-degree point with the formula is illustrated in Figure 4.

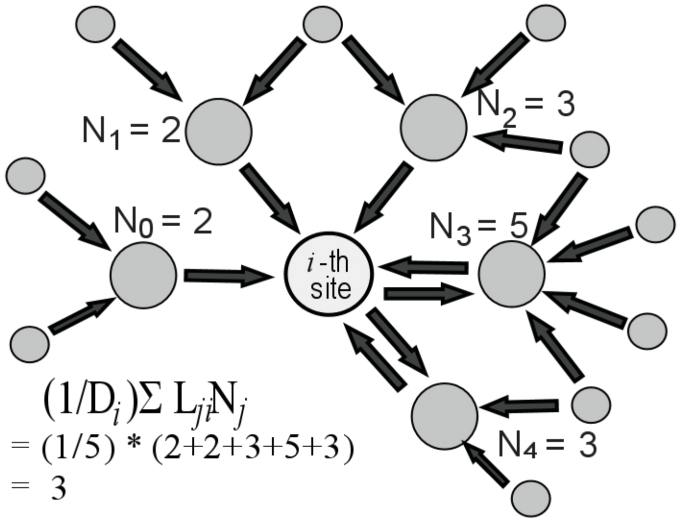


Figure 4. An example of a network.

Figure 5 shows the examples of the environmental problems and the atomic energy communities comprised of hyperlinks on link pages. The evaluated Q values were plotted for the in-degree values of the community Web sites. The values gradually increased with k. This inclination implies that the credible Web sites with large in-degrees create hyperlinks to the credible Web sites; further, the rate of credible sites increases with the in-degree of the sites. Thus, the second supposition is confirmed.

7. Discussion

The first analysis revealed that the Web sites with large in-degrees had small out-degrees in their link pages. The second analysis revealed that Web sites with large in-degrees created hyperlinks to other Web sites

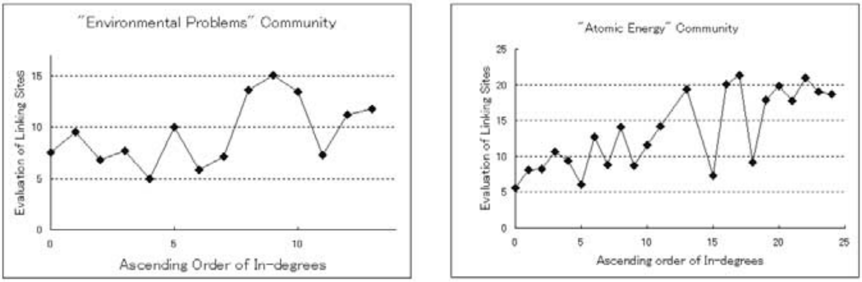


Figure 5. In-degree point for the in-degree.

with large in-degrees. These two results imply that a credible Web site exclusively offers links on link pages to highly credible Web sites. It is suggested that a credible Web site must carefully select the Web sites to which it offers a link. Thus, we can conclude that a credible Web site realizes its demeanor with regard to related sites and creates hyperlinks in accordance with its social position and its association with related sites or administrators. The administrator evaluates the contents and hyperlinks of other Web sites with reference to the concept of the social position of his or her own Web site. There after, he or she selects hyperlinks and edits the link page.

As mentioned earlier, a ranking algorithm like the Page Rank algorithm (Brin & Page, 1998) would generate the probability for a random surfer who is visiting pages or sites. This probability reflects the number of in-coming hyperlinks on both the link pages and other pages. Although the calculation process is basic and the number of in-coming hyperlinks most likely reflects the credibility of the pages within a certain range, the algorithm could easily become complicated with regard to the potentially great number of hyperlinks. The deference, including the reliability of the related sites, would be neglected. When we need to obtain authorized knowledge on an unfamiliar topic, we need to rely on dependable clues. We suggest, then, that the hyperlinks found on link pages of related sites should be improved to ensure consistency for the community.

In order to develop a good Internet society, we must uphold the "personality" of individuals on the Web. However, at this date it seems too categorical to enhance the consciousness and technical traits of the participants. As revealed by the "rich get richer" model, a uniform commitment of a newcomer could enlarge the social difference between the rich and the poor (Barabási, 2002). Perhaps we should be more conscious of the fact that each individual would necessarily stand in some social position and make some order in an Internet society. In order to analyze the Web in greater detail and to better utilize it, we have to look beyond the arguments about "personality" as Goffman (Goffman, 1967) did when examining Durkheim's theory (e.g., Durkheim, 1912). We must recognize increasing social differences and consider the peculiarity of the participants while referring to the Goffman theory. The authors' study has attempted to apply the sociological theory of Face Work to the analysis of the Web, thus informing users how to better utilize the knowledge available on the Web.

References

1. Barabási, A. L. (2002). LINKED: *The new science of networks.* New York: Perseus.
2. Brin, S. & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems,* 30, 107–117.
3. Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., & Wiener, J. (2000). Graph structure in the web. *Computer Networks and ISDN Systems,* 33(1-6), 309–320.
4. Durkheim, E. (1954). *The elementary forms of the religious life, transformed by Swain, J. W.* Glencoe, Illinois: Free Press.
5. Goffman, E. (1963a). *Behavior in public places.* New York: Free Press.
6. Goffman, E. (1963b). Stigma: *Notes on the management of spoiled identity.* Upper Saddle River, NJ: Prentice-Hall.
7. Goffman, E. (1967). *Interaction ritual.* New York: Pantheon.
8. Haveliwala, T. H. (2002). Topic-sensitive PageRank. In *Proceedings of the 11th International World Wide Web Conference* (pp. 517–526).
9. Jeh, G. & Widom, J. (2003). Scaling personalized Web search. In *Proceedings of the 12th International World Wide Web Conference,* (pp. 271–279).
10. Kleinberg, J. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM* 46(5), 604–632.
11. Kleinberg, J., Kumar, R., Raghavan, P., Rajagopalan, S., & Tomkins, A. (1999). *The web as a graph: measurements, models, and methods.* Paper presented at the Proceedings of the 5th Annual International Computing and Combinatorics Conference, (pp. 26–28).
12. Kumar, R., Raghavan, P., Rajagopalan, S., & Tomkins, A. (1999). Trawling the web for emerging cyber-communities. *Computer Networks,* 31 (11-16), 1481–1493.

Chapter 22

THE USE OF KEYPHRASES FOR SELECTING METADATA FROM TAXONOMIES

Rohana K. Rajapakse
Research and Innovation, University of Plymouth
Plymouth, PL4 8AA, U.K.
rrajapakse@plymouth.ac.uk

Brian Mushens
School of Computing Communications and Electronics, University of Plymouth
Plymouth, PL4 8AA, U.K.
bmushens@plymouth.ac.uk

Chris Johnson
School of Computing Communications and Electronics, University of Plymouth
Plymouth, PL4 8AA, U.K.
cjohnson@plymouth.ac.uk

This paper presents a meta-tag suggestion tool that selects topics from a given taxonomy of categories. It works by extracting keyphrases from the textual content of electronic data and using them as entry points to search for appropriate topics in a hierarchically-organized taxonomy. The keyphrase extraction module, based on the Kea system, has the functionality to use pre-built Kea models or create new Kea models on the fly, thereby providing flexibility in controlling keyphrase extraction. Terms in each keyphrase are stemmed and used as a query to search for appropriate topics in the taxonomy. All taxonomy topics retrieved by repeated searching are considered as candidate meta-tags. They are ranked based on the significance of keyphrase(s) matching the topic, the significance of the type of hit, and the specificity of the matching topic based on its position in the taxonomy.

1. Introduction

As the production and availability of massive collections of electronic data are growing at alarming proportions, dealing with this material is becoming increasingly difficult. Whereas in the past, it was possible to distinguish between content producers and consumers, this distinction is becoming increasingly blurred. Any businesses in any sector, be it commercial, government, academic, or administrative, and any member of the general public may well be both a content producer, as well as a consumer. These developments have caused at least two major problems in dealing with such vast electronic data collections: (1) producers are confronted with difficulties in the organization, management, and storage of their content; (2) consumers are faced with the difficulty of finding useful information efficiently from these information forests. The process of automated meta-tagging, described here, attempts to address both aspects of the problem.

Metadata is additional descriptive information about an object or resource, and serves as a condensed representation of the underlying resource. In this context, metadata comprises meta-tags that are representative of a document's informational content. In reality, assigning appropriate meta-tags to documents is not a trivial exercise because to execute it effectively, domain experts are needed to determine appropriate meta-tags.

Today's organizations often need to collaborate and share information with other organizations, as well as with individual customers. An organization must focus on the consumers' perspective of the content when organizing information and must categorize content by the familiar terms used by consumers to make decisions. These terms must be consistent and understood by everyone. In information organization and retrieval, this task is achieved by maintaining a consistent "controlled vocabulary" or "taxonomy" for content categorization and indexing. The emerging taxonomies designed to cater for different business areas help standardize the use of terminology. These taxonomies are gaining wider acceptance and are increasingly integrated into business applications.

In this paper we present a tool for suggesting metadata, the "metadata suggester" which extracts suitable topics from a given taxonomy as potential metadata for a given document. Section 2 reviews related research and existing products on the two related topics of keyphrase extraction and meta-tagging. In the sections 3 and 4, we introduce and describe our metadata suggester model. Section 5 provides an evaluation of the model followed by a discussion of the model, and planned future work is described in section 6.

2. Related Work

2.1. *Keyphrase Extraction*

A keyphrase-based meta-tagging tool requires extracting keyphrases from textual contents. Among the phrase extraction systems are: NPtool (Voutilainen, 1993), AZ Parser (Bennet, He, Powell, & Schatz 1999], Church's tagger (Church, 1988), LECTER (Bourigault, 1992), B&C (Baker & Cornacchia, 2000], Extractor (Turney, 1999), and Kea (Witten, Paynter, Frank, Gutwin, & Nevill, 1999).

In our work we use the Kea (Witten et al., 1999) keyphrase extraction system. In Kea, phrases constitute sequences of consecutive words with no intervening phrase boundary indicators. It uses the TF×IDF (term frequency × inverse document frequency) value and the position of the phrase in the document as features in its model. These feature values are calculated for all candidate phrases in the documents in a training corpus, and each candidate phrase is marked as a positive or negative example depending on whether or not it is among the author’s keyphrases. A Naïve Bayesian technique is used to assign weights to the features, based on the feature values of the positive and negative examples.

2.2. *Taxonomies*

A taxonomy is an ordered collection of (usually hierarchically organized) categories. Taxonomies are used for (1) organizing information by means of applying electronic labels or meta-tags, and (2) giving guidance to users looking for information (browsing). Where a number of departmental content management systems exist within an organization, a common taxonomy can enable the same content in different systems to be labelled and retrieved consistently.

Dewey Decimal Classification system (DDC) is a commonly used system in public and school libraries throughout the world. The Government Category List (GCL), the Local Government Category list (LGCL) and SeamlessUK are category lists that define the subject matter of government and local government-related community resources. The Integrated Public Sector Vocabulary "IPSV" (Integrated Public Sector Vocabulary [IPSV], 2005) is a new standard created by merging GCL, LGCL, and SeamlessUK, and is the recommended set of topics for tagging government and local government electronic content in the UK. The complete IPSV has 2,732 preferred terms and 4,230 non-preferred terms.

2.3. *Meta-tagging*

A popular mechanism to describe the content of Web pages so that search engines may make use of them to facilitate indexing is to embed meta-tags. Alternative encoding schemes are also being used mainly by smaller enterprise applications for providing internal search facilities. Although such meta-tagging practices have been in use for a long time, the processes of metadata selection and assignment have largely been performed manually. The massive collections of electronic material in which today's enterprise content management systems have to cope do not lend themselves to manual meta-tagging.

Dublin Core standards ([DCMI], 2006) provide sets of HTML, RDF, and XML elements/tags that support a broad range of meta-tagging purposes. Many of these tags are used in other metadata standardization efforts. "DCdot", Dublin Core's automatic metadata generation tool allows generation of Dublin Core metadata. *NetOwl* Extractor (NetOwl Extracter, 2006) provides assistance interacting with a human question answering function for tagging documents by ticker symbol, by name, or by other category. Wordmap Meta-Tagger (Bock, 2004) is a browser-based application that allows a content author to choose metadata labels from pick lists it displays. Meta-Tagger is integrated with Wordmap Enterprise Taxonomy Platform (ETP) and Wordmap Intelligent Text Classifier (ITC), and is pre-configured to the e-Government Metadata

Standard (e-GMS). Open Object's IPSV Recommender (KBroker IPSV Recommender, 2006) provides an automated means to determine suitable IPSV categories for public sector websites. Searchlight from Cintra (Cintra, 2006) is an e-GIF and e-GMS compliant modular Taxonomy Management solution that automates the process of generating searchable LGCL, GCL, and IPSV indexes and meta-tags for Web pages.

3. Keyphrases as Entry Points for Taxonomy Matching

Starting from the days of manual library catalogues, the use of keywords and keyphrases have been very well researched and used for document categorization, indexing, and retrieval. Although the keyword representations lack the rich semantics of natural language content, the ability to interpret keywords individually and independently of each other is seen as an advantage. This has resulted in the inception of historically dominant and successful statistical IR models. Using keyphrases can: 1) help users get a feel for the content of a collection, 2) provide sensible entry points into it, 3) provide guidance for query expansion, 4) facilitate document skimming by visually emphasizing important phrases; and 5) offer a powerful means of measuring document similarity. These features of keyphrases make them the most effective entry points for metadata selection.

In general, unguided keyphrase selection is inconsistent and error-prone. Taxonomy-guided manual indexing, on the other hand, is much more consistent. However, manual selection of appropriate topics from thousands of topics available in taxonomy is extremely tedious. Automating this process is potentially of great benefit in removing the burden of the manual process. We can use keyphrases extracted from document contents as a means of identifying relevant topics from a taxonomy. The simplest solution is to search an index of the taxonomy to find matching taxonomy topics for the keyphrases extracted. More complicated and fine-grained taxonomy topics could be found by combining multiple evidence that can be found based on the contextual meaning of the topics as indicated by their synonyms and broader, narrower, and related terms and any other information available in the taxonomy. Here we illustrate the main components of our model (Figure 1), and describe their functionality.

The input document first undergoes a text extraction and cleaning processes. During the text extraction process, text in the title and body of the document is extracted. The extracted text is then subject to a cleaning process for removing non-textual characters and other tags (e.g. HTML). The result is a stream of clean text. This stream is then passed to Kea to extract and identify significant keyphrases in the text. Each individual keyphrase is considered an entry point for discovering metadata and is used as a query to search an index of the IPSV taxonomy. Searching is performed by the Lucene search engine (Gospodnetic & Hatcher, 2005) on an index of the taxonomy created by Lucene. The individual terms in keyphrases are stemmed to remove morphological variances. The searching is performed on all fields available in the implementation of the taxonomy, and hits on taxonomy entries are retrieved. A hit on a taxonomy entry can take place on a stemmed version of the title of the taxonomy topic (direct hit) or on one or more other fields (indirect hit), e.g. synonyms. In addition to the taxonomy topics retrieved by direct or

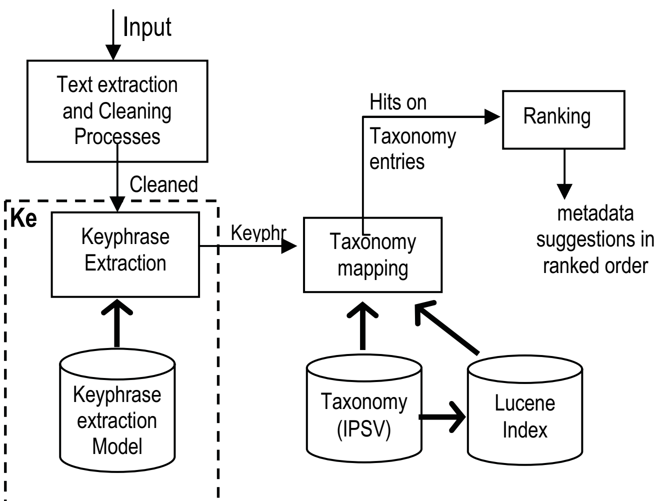


Figure 1. Schematic diagram of the model.

indirect hits, their immediate parents are also taken as candidates for metadata. (We call them parent hits.) This results in a list of taxonomy topics retrieved by one or more direct, indirect, and/or parent hits. Finally a ranked list of taxonomy topics is created and presented to the user as suggestions for meta-tagging.

4. Ranking Principle

There are a number of parameters in this retrieval process that may influence the relevance of a taxonomy topic to the document. Amongst these are: 1) how good a keyphrase (that makes a hit on a taxonomy entry) is at describing the content of the document, 2) the type of hit (direct, indirect or parent hit), and 3) how specific (or generic) the taxonomy topic is according to its position in the hierarchy.

We use the keyphrase scores given by Kea as a measure of how good a keyphrase is at describing the content of the document. We consider direct hits more important than indirect and parent hits to be the least important. Individual hits are weighted accordingly. The specificity of a topic in the taxonomy is measured based on how many siblings the parent of the topic possesses in the taxonomy hierarchy. The rationale behind this is the assumption that a topic with a large number of siblings of the topic's parent is less specific compared to a topic with fewer parent siblings. This is particularly relevant for taxonomies with unbalanced hierarchies where the level or position in the hierarchy does not necessarily relate to specificity. Alternatively, the siblings of the topic itself could have been considered. However this fails for topics that are leaf-nodes where there are no siblings.

Note that, although a given keyphrase can make only one direct hit on a taxonomy topic, more than one keyphrase from the same document can make direct, indirect, or parent hits on the same topic resulting in multiple hits on a given topic. Our ranking principle combines all such hits to compute a final score for each topic. This score decides its place in the rank list:

$$\textbf{\textit{Score}}_{\textbf{\textit{tag}}} =$$

$$\left(\sum_{all\ keyphrases} Phrase\ Weight \times \left(\alpha \times Direct\ Hits_{tag} + \beta \times Indirect\ Hits_{tag} + \gamma \times Parent\ Hits_{tag} \right) \right)$$

$$\Delta \left(4 \log_2 \left(\frac{No\ of\ siblings\ of\ parent\ topic}{Size\ of\ the\ biggest\ category} \right) \right)$$

where α, β and γ are weighting parameters for hit types.

Figure 2 shows some statistics collected on the topic of "*school types*" which was suggested by our model for the text "*One of the school types in Plymouth is Secondary Schools*". Two keyphrases, "***school types***" and "***secondary schools***", have been extracted by Kea from this piece of text of which the keyphrase "***secondary schools***" has a parent hit on the taxonomy topic "***school types***" and the keyphrase "***school types***" has a direct hit and seven parent hits. The seven parent hits correspond to direct hits by the "***school types***" on seven of the fifteen siblings of the topic category "***school types***" in the taxonomy. Figure 3a shows the siblings of "***School types***" in IPSV, and Figure 3b shows the suggestions made by our system.

Tag ID: 14128
Tag Title: school types
Level: 4
Items in the Category: 86
Direct Hits: 1
Parent Hits: 8
Indirect Hits: 0
Kea Phrase Hits to this Tag

    Kea Phrase: secondary schools
    Kea Phrase Rank: 3
    Kea Phrase Score: 0.3994953776974945
    Direct Hits: 0
    Parent Hits: 1
    Indirect Hits: 0

    Kea Phrase: school types
    Kea Phrase Rank: 0
    Kea Phrase Score: 0.3994953776974945
    Direct Hits: 1
    Parent Hits: 7
    Indirect Hits: 0

Figure 2. Statistics .collected on a taxonomy topic.

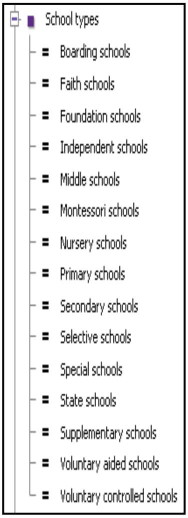


Figure 3a. Part of IPSV showing the siblings of "school types".

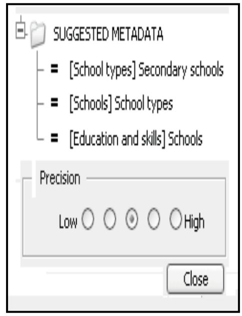


Figure 3b. Screen shot of the suggestions made.

5. Evaluation

There are at least two basic approaches to evaluating automatically generated metadata. One approach is to compare automatically generated metadata against a set of manually allocated metadata. This approach can be used to measure the comparative performance of different models using the traditional information retrieval metrics of precision and recall. In order to be valid, the evaluation must be made on the same collection using the same set of manual metadata sets. Although we are interested in performance comparison against other models, the lack of a benchmark collection of documents with manually allocated metadata, and the lack of published results on such a collection by other models (using the same taxonomy) restricted our desire to adopt this approach. The second approach is to gather subjective assessments from human subjects by providing them with a document and the corresponding set of metadata and asking them to assess in some way the relevance of individual metadata items to the given document (Jones & Paynter, 2001; Liddy et al., 2003). This approach requires, in addition to a collection of documents, domain experts who are able to make appropriate assessments. The difficulty of meeting these requirements and the time overheads of such an exercise compelled us to leave it for the future.

Instead, a basic evaluation was performed to determine the proportion of IPSV meta tags assigned to a collection of publicly available Web pages by our model compared to a commercial product.

We extracted a collection of 100 Web pages from a commercial website and selected a subset of 40 by carefully examining the contents to ensure that a wide range of topics was covered. These pages contained IPSV meta tags assigned to them by a commercial product. The average page length in the selected collection was 1,884 words, and there were 188 IPSV meta tags in total assigned to the collection, i.e. an average of 4.7 metadata items per page.

As detailed above, only the heading and body sections of each page were extracted in order to eliminate the influence of additional meta data and other content such as headers and footers. The extracted text was then passed to our model, and metadata suggestions obtained. These suggestions were then compared to the metadata that had already been assigned. We experimented with a combination of Kea model parameters to observe the effect on the recall and precision of our model to retrieve

Table 1. Examples for the functionality of Kea parameters.

Min Words in a Phrase	Max. Words in a phrase	Min occurrence frequency	Keyphrases Kea extract
2	2	1	Keyphrases having exactly two words and appearing one more times in the text
1	3	1	Keyphrases having 1, 2 or 3 words and appearing at least once in the text
2	2	2	Keyphrases having exactly two words and appearing two or more times in the text. Two-word keyphrases appearing at least once in the text

the already assigned metadata. These parameters are the minimumn umber of words in a keyphrase, the maximum number of words in a keyphrase, and the minimum number of times a keyphrase occurs in the text. These parameters give great flexibility to Kea in keyphrase extraction. The following table explains how the parameters affect keyphrase extraction.

It is important to note that our results below are based on the existing metadata assignments generated by an automated product rather than human domain experts. As a result, they do not reflect the true performance of our model, but offer a comparative view of how effective our model is at recalling assigned meta-tags.

5.1. *Measurement Metrics*

We have adapted the commonly used measures of recall and precision to suit the meta-tagging activity. Thus, we define recall as the proportion of assigned meta-tags that are identified by our model, i.e.

$$\boldsymbol{Recall} = \frac{No.of\ assigned\ metatags\ identified}{Total\ no.of\ metatags\ assigned\ to\ the\ document}$$

We define precision at a cut-off of 10, i.e. after the top 10 ranked meta-tags, as:

$$\boldsymbol{Precision@10} = \frac{No.\ of\ assigned\ metatags\ suggested\ within\ top\ 10}{10}$$

5.1.1. *The Effect of Keyphrase Lengt*

Table 2 compares the effect of keyphrase lengths (i.e. one-word, two-word or three-word phrases). The phrase occurrence frequency parameter of Kea was kept fixed (to one, i.e. any number of occurrences were considered) in this experiment. It shows that the use of lengthy keyphrases causes recall to drop (see Table2 and Fig. 4). As expected, the precision has been better with 2-word phrases than with 1-word phrases. This is mainly because the single-word phrases are less specific and nosier. However, despite the fact that 3-word phrases are more specific than one and two-word phrases, the precision has been very low with 3-word phrases. This is due to the fact that longer-phrase searches are prone to more mismatches. These results indicate the possibility of varying phrase length (between 1- and 2-word) to narrow or broaden the metadata suggestions without affecting precision significantly.

Table 2.

No of words in a keyphrase	Total no of tags Retrieved	Avg. Tags retrieved per page	Tags common to both products	Pages with no suggestions	Pages with no common tags (of > 0 suggestions)	% common tags (out of retrieved) within top 10	Average Recall	Precision at retrieval point 10 (average over all documents)
1	61854	1,546.35	173	0	0	26.59	0.9142	0.1150
2	765	19.13	73	0	11	76.71	0.3742	0.1400
3	362	9.05	31	13	13	77.42	0.1550	0.0600

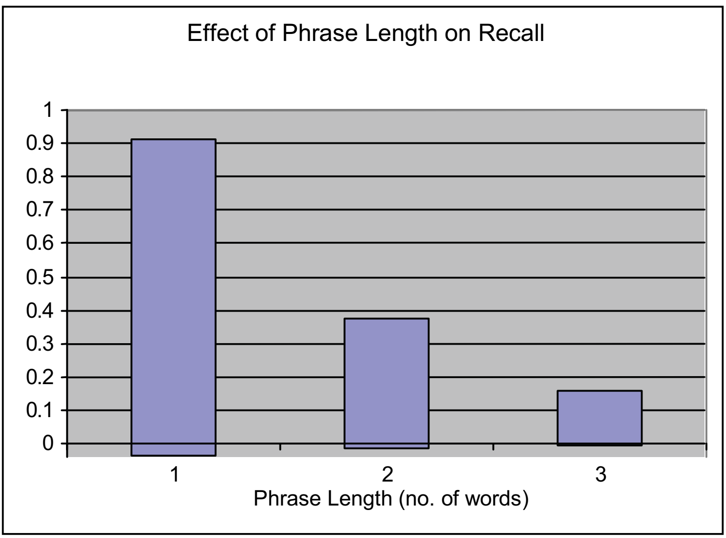


Figure 4. Effect of phrase length on average recall.

5.1.2. *Effect of keyphrase occurrence frequency*

Table 3 compares the effect of Keyphrase occurrence frequency. The phrase occurrence frequency was varied from 1 to 3 while the other two Kea model parameters (i.e. min and max number of words in a keyphrase) were kept fixed (1 and 3 respectively). The results suggest that the phrases that occur more frequently are slightly better in recalling metadata as more common tags are found for phrases that are more frequent (see Fig 5). However, occurrence frequency has shown only a

Table 3. Experimental results on the effect of phrase occurrence frequency.

Minimum no of times a keyphrase occurs	Total no of tags Retrieved	Avg. Tags retrieved per page	Tags common to both products	Pages with no Tags	Pages with no common tags (of > 0 suggestions)	% of common tags within top 10	Average Recall	Precision at retrieval point 10 (average over all documents)
1	28616	715.40	164	0	0	31.71	0.8646	0.1300
2	95146	2,378.65	184	0	0	20.65	0.9800	0.0950
3	104616	2615.4	186	0	0	20.97	0.9900	0.0970

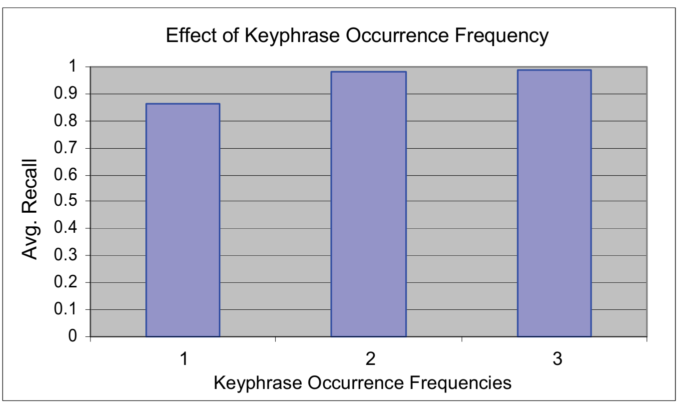


Figure 5. Effect of phrase occurrence on average recall.

marginal effect on precision. It is important to note that high frequency keyphrases are included in the set of low frequency keyphrases extracted, because the occurrence frequency parameter only sets the minimum phrase occurrence frequency (e.g. setting minimum occurrence to 2 extracts phrases that appear 2 or more times in the text).

5.1.3. *The Effect of Subphrases*

In this experiment, we loosened the strict condition on the number of words used in first experiment to include subphrases of multi-word phrases. The maximum number of words in a phrase was kept fixed (set to 3) and the minimum number of words in a phrase was varied between 1 and 3. The phrase occurrence frequency was kept fixed (set to one). The first data row shows results on 3-word phrases (as above, Table 2), the second row shows results on phrases with 2 and 3 words, and the third row on phrases with 1, 2 and 3 words. Both the Recall figures (column 4, Table 4 and Fig. 6) as well as Precision figures (last column in Table 4) have increased as the restriction on the number of words in a phrase is loosened. This result indicates that it is possible to expand (or narrow down) metadata suggestions based on inclusion (or exclusion) of subphrases in the set of keyphrases used.

Table 4. Experimental results on the effect of sub-phrases.

No of words in a keyphrase	Total no of tags Retrieved	Avg. Tags retrieved per page	Tags common to both products	Pages with no Tags	Pages with no common tags (of > 0 suggestions)	% common tags within top 10	Average Recall	Precision at retrieval point 10 (average over all documents)
3	362	9.05	31	13	13	77.42	0.1550	0.0600
2-3	535	13.38	52	6	13	86.54	0.2646	0.1125
1-3	28616	715.40	164	0	0	31.71	0.8646	0.1300

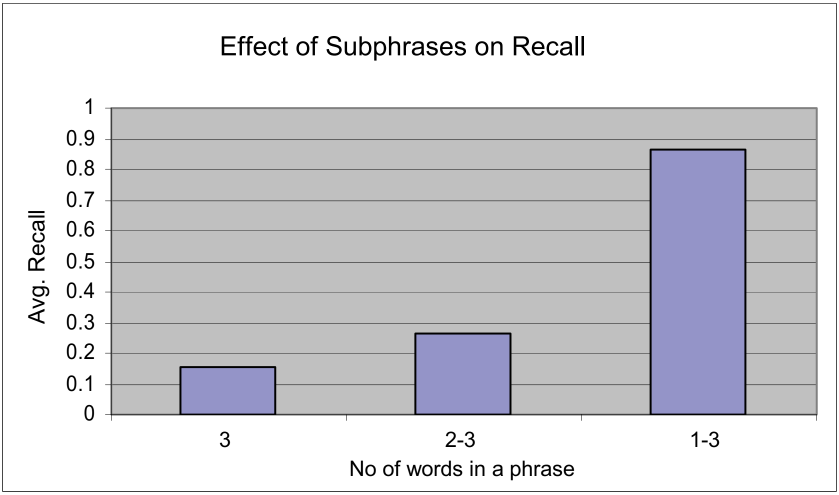


Figure 6. Effect of subphrases on recall.

6. Discussion and Conclusions

This paper reported our development work on an automated metadata selection system. Our approach was mainly to use keyphrases extracted from the content of documents as entry points to look up appropriate IPSV taxonomy entries for metadata. The model has been integrated into the iCM content management product of Goss Interactive Ltd. (*http://www.gossinteractive.com/*). A feature that users have found interesting and that other commercial products do not offer (to our knowledge) is the flexibility to broaden or narrow down the metadata suggestions. In addition to web content, the tool is able to deal with PDF, RDF, Word, Excel and PowerPoint file formats.

The experimental results given in section 5.0 suggest that the keyphrase length (in number of words) is a crucial parameter and by varying it, high or low recall can be achieved. The keyphrase occurrence frequency, on the other hand, has not affected our results as much as we expected.

Two significant results of our evaluation are: 1) the high recall of assigned metadata (near perfect for low number of words in a phrase) and 2) of the suggested items, a reasonably good percentage lies within the top 10 suggestions. Most importantly, these results were achieved in an evaluation scenario based on a set of machine assigned metadata by a software tool rather than a set of metadata manually chosen by domain experts. Machine assigned metadata is not accurate and is error-prone. A careful examination of metadata suggestions reveals our model to have more appropriate ones than those previously assigned.

Based on the initial success, we are now working towards further improving the accuracy of the model, as well as extending its flexibility to deal with customized metadata. Firstly, we intend to incorporate learning in the model so that associations between taxonomy topics and keyphrases can be established their significance can be learned interactively. Secondly, we want to add functionality to enable user-defined customized metadata item lists to be created and maintained so that users can make use of their own subjective metadata (folksonomies) if they wish. Finally, we will be conducting a more rigorous evaluation to measure the true performance of the model quantitatively (using precision- and recall-based metrics) as well as qualitatively with subjective, user-centred evaluation.

Acknowledgments

Thanks are due to Robert McCarthy and Nick Gant of Goss Interactive Ltd. for their support and contributions, and Patrick Kimber (of Goss) for his contribution in customizing and integrating Lucene software. Finally, appreciation goes to the Kea and Lucene groups for releasing their software under the GNU General Public License.

References

1. Barker, K. & Cornacchia, N. (2000). Using noun phrase heads to extract document keyphrases. *Proceedings of the Thirteenth Canadian Conference on Artificial Intelligence* (LNAI 1822) (pp. 40-52). Montreal, Canada.
2. Bennet, N.A., He, Q., Powell, K. & Schatz, B. R. (1999). Extracting noun phrases for all of Mediline. *Proceedings of the AMIA 99*.
3. Bock, G. E. (2004). Mapping words for an enterprise taxonomy: How Wordmap manages the terminology of an organization. White paper prepared for Wordmap Limited. Retrieved May, 2006, from http://www.wordmap.com/downloads/whitepapers/wordmap-seybold-paper.pdf.

4. Bourigault, D. (1992). Surface grammatical analysis for the extraction of terminological noun phrases. *Proceedings of the Fifteenth International Conference on Computational Linguistics*. COLING-g2, Vol. IIL, (pp. 977-981). Nantes, France.
5. Church, K. (1988). A stochastic parts program and noun phrase parser for unrestricted text. *Proceedings of the Second Conference on Applied Natural Language Processing*, (pp.136-143). ACL.
6. CINTRA UK. (n.d.). Searchlight Taxonomy Management. Retrieved May, 2006, from http://www.cintra.com/uk/uk_taxo.html.
7. Dublin Core Metadata Initiative. (n.d.). Retrieved May, 2006, from http://dublincore.org/.
8. Gospodnetic, O. & Hatcher, E. (2005). *Lucene in action*. Greenwich, CT: Manning Publications Co.
9. Greenberg, J., Pattuelli, M. C., Parsia, B. & Robertson, W. D. (2001). Author-generated dublin core metadata for web resources: A baseline study in an organization. *Journal of Digital Information (JoDI)*, 2(2). Retrieved May, 2006, from http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Greenberg/.
10. Integrated Public Sector Vocabulary. (n.d.). Guidance notes for local authority and other users. Retrieved May, 2006, from www.esd.org.uk/IPSVGuidanceNotes.pdf.
11. Jones, S. & Human, P. G. (2001). Human evaluation of Kea, an automatic keyphrase system. *Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries*, (pp. 148–156). Roanoke, Virginia, United States.
12. Liddy, E. D., Allen, E. E., Finneran, C. M., Gay, G., Hembrooke, H. & Granka, L.A. (2003). MetaTest: Evaluation of metadata from Generation to use. In proceedings of the *3rd ACM/IEEE-CS joint conference on Digital libraries* (pp.398 – 398).
13. NetOwl Extracter. (n.d.). Retrieved May, 2006, from http://www.netowl.com/products/extractor.html.
14. KBroker IPSV Recommender. (n.d.). Retrieved May, 2006, from http://www.openobjects.com/solutions/kbrokeripsv.jsp.
15. Turney, P. D. (1999). *Learning to extract keyphrases from text.* Technical Report ERB-1057 (NRC #41622). Canadian National Research Council, Institute for Information Technology.
16. Voutilainen, A. (1993). NPtool: A detector of English noun phrases. *Proceedings of the workshop on very large corpora*, Columbus, OH.
17. Witten, I. H., Paynter, G. W., Frank, E., Gutwin, C. & Nevill, C. G. (1999). KEA: Practical automatic keyphrase extraction. *Proceedings of ACM Digital Libraries Conference*, pp. 254-255.

This page intentionally left blank

INDEX

The Third International Conference on Knowledge Management, University of Greenwich, London, 31st July−2nd August 2006

ORGANIZING COMMITTEE

General Chair	Professor Suliman Hawamdeh
Provisional Conference Chair	Professor Les Johnson
Conference Co-Chairs	Martyn Laycock Dr. Bruce Cronin Professor Janet McDonell Matthew Khoe
Event Support Manager	Diane Barnett