FRAMEWORK FOR ANALYZING ONLINE ASYNCHRONOUS DISCUSSION BY INTEGRATING CONTENT ANALYSIS AND SOCIAL NETWORK ANALYSIS

ERLIN

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Doctor of Philosophy (Computer Science)

Faculty of Computer Science and Information Systems
Universiti Teknologi Malaysia

APRIL 2011

ALHAMDULILLAH ...

I wish to express my great gratitude to my beloved parents for everything that they have done for me.

I thank all my brothers and sister for their support.

And I am especially would like to extend my deepest gratitude to my husband, *Bayu Buana*, who always support and encourage me in the good times, as well as the bad times. Also to my beloved children, *Farhan Bayu Putra*, *Aunaya Bayu Putri* and *Farrel Bayu Putra*, who have given me unconditional love and purity that lights up my days. They are the most important people in my life. Thanks for giving me a happy family. With no doubt, the endless love of my family kept me on track and helped me survive this adventure.

ACKNOWLEDGEMENTS

All praise and thanks are due to Allah the Almighty in whom I put my trust. I extend also my full respect to my beloved Prophet Mohammed peace and blessing upon him.

I would like to express my gratitude to all those who have given me the possibility to complete this thesis. In particular, I wish to express my greatest sincere appreciation to my main supervisor, Assoc. Prof. Dr. Norazah Yusof, for encouragement, guidance, critics and willingness to let me explore. I am also very thankful to my co-supervisor, Assoc. Prof. Dr. Azizah Abdul Rahman, for guidance, priceless advices, and motivation throughout this research. They have awarded me with generous support, and their assistances are invaluable in my research process. Their direction leads to the accomplishment of my thesis.

A great gratitude also goes to Pemerintah Provinsi Riau for financial support at the first year and to Direktorat Jenderal Pendidikan Tinggi Departemen Pendidikan Nasional for financial support at the last two years of my study. I am also indebted to Ketua Yayasan Komputasi Riau; Prof. Dr. Dadang Iskandar, M.Sc., for encouragement and funding my study time by time. My sincere appreciation also extends to Ketua STMIK-AMIK Riau; Drs. Suprasman, MM, my inspirator; Drs. H. Chan Sirdi, MLs, and my colleagues who have provided assistance at various occasions.

I also wish to express my gratitude to the faculty members and staff of FSKSM-UTM for their help and support. Last but certainly not least, many thanks go to all the wonderful friends and colleagues that I got to know throughout stay in Malaysia. Thanks for your support, whenever I needed to vent my stress, setbacks, more stress and frustrations.

ABSTRACT

Online Asynchronous Discussion (OAD) is a powerful way to conduct online conversation and a significant component of online learning. Unfortunately, existing Learning Management System (LMS) that generally provides online discussion cannot afford a comprehensive evaluation on the content of the transcripts and the level of interaction among participants. Therefore, this research explores the analysis process of OAD qualitatively and quantitatively. The work focuses on Content Analysis (CA) and Social Network Analysis (SNA), two popular methods employed by educators and researchers to analyze online discussion in e-learning environment. Although these two methods are well established, the techniques remain manual. Furthermore, presently, these two methods of analysis are conducted and studied independently. Hence, this research proposes a new framework integrating CA with SNA called CASNA, which provides comprehensive information of the result, and automation of the processes. CASNA is applied and embedded in LMS (Moodle) to validate the proposed framework. This research also introduces sentence as the unit of interaction instead of message to assess the level of participation among students. In addition, in order to qualitatively analyze the online discussion, two text classifiers; the Support Vector Machine (SVM) and the Back-propagation Neural Network (BPNN) approaches are employed to categorize the sentences based on Soller's model and the results are compared. The evaluation of these two classifiers is done based on precision, accuracy, recall and F-Measure. The result shows that SVM outperform BPNN in terms of precision and accuracy; falls behind BPNN in terms of recall and F-Measure. This research also discusses the use of network indicators of SNA. Adjacency matrix, graph theory and network analysis techniques are applied to quantitatively define the network interactions among participants. This framework takes advantage of the strength of each method and offers dynamic analysis of the textual messages. It is expected to be more informative to educators as well as researchers in measuring the quality and quantity of OAD.

ABSTRAK

Perbincangan Dalam Talian Tak Segerak (PDTTS) adalah langkah terbaik dalam menjalankan perbincangan dalam talian dan merupakan komponen penting dalam pembelajaran dalam talian. Namun begitu, Sistem Pengurusan Pembelajaran (SPP) sedia ada, yang umumnya menyediakan kemudahan perbincangan dalam talian, tidak mampu untuk melakukan penilaian menyeluruh terhadap kandungan transkrip dan aras interaksi antara peserta. Oleh kerana itu, penyelidikan ini mengkaji proses penganalisaan PDTTS secara kualitatif dan kuantitatif. Kajian ini memfokuskan kepada analisis kandungan dan analisis rangkaian sosial, yang kedua-keduanya adalah kaedah terkenal yang digunakan oleh pengajar dan penyelidik untuk menganalisis perbincangan dalam talian pada persekitaran e-pembelajaran. Walaupun kedua-dua kaedah ini telah mantap, namun kaedah ini masih dalam bentuk manual. Malah, kedua-dua kaedah analisis ini masih digunakan berasingan. Oleh yang demikian, penyelidikan ini mengusulkan rangka kerja baru yang menyepadukan analisis kandungan dan analisis rangkaian sosial (CASNA) yang meliputi penggabungan maklumat keputusan yang menyeluruh, dan pengautomasian proses. Untuk mengesahkan rangka kerja yang dicadangkan, CASNA telah diaplikasikan dan dibenamkan ke dalam SPP (Moodle). Kajian ini juga memperkenalkan ayat sebagai unit interaksi untuk menggantikan mesej dalam menilai aras penyertaan di kalangan pelajar. Di samping itu, untuk menganalisa perbincangan dalam talian secara kualitatif, dua pendekatan pengklasifikasian teks; *Mesin Bantuan Vector (MBV)* dan *Rangkaian Neural* (RN) telah digunakan untuk mengkategori ayat berdasarkan model Soller dan hasilnya telah dibandingkan. Penilaian terhadap kedua-dua pengklasifikasian ini dijalankan berdasarkan kepersisan, kejituan, perolehan-kembali, dan ukuran-F. Penyelidikan ini juga membincangkan tentang penggunaan indikasi analisis rangkaian sosial. Matrik kesebelahan, teori graf dan teknik analisis rangkaian telah digunakan supaya dapat menerangkan secara kuantitatif rangkaian interaksi antara peserta. Rangka kerja ini mengambil kelebihan dari kekuatan kaedah-kaedah yang dicadangkan serta mampu menawarkan analisa secara dinamik bagi mesej teks. Ia diharapkan memberi maklumat yang lebih bermakna kepada pengajar dan penyelidik dalam mengukur kualiti dan kuantiti PDTTS.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

ICT	Information And Communications Technology
CMC	Computer Mediated Communication
WBE	Web-Based Learning
CBI	Computer Based Instruction
CLE	Cybernetic Learning Environment
ILS	Integrated Learning System
CBT	Computer Based Training
CSCL	Computer Supported Collaborative Learning
CA	Content Analysis
SNA	Social Network Analysis
LMS	Learning Management System
OAD	Online Asynchronous Discussion
CASNA	Content Analysis And Social Network Analysis
NN	Neural Network
SVM	Support Vector Machine
BPNN	Back-propagation Neural Network
MLP	Multi Layer Perceptron
TSR	Terms Space Reduction
DF	Document Frequency
CS-DF	Category Frequency-Document Frequency
TF	Terms Occurrence Frequency
IDF	Inverse Document Frequency
PCA	Principal Component Analysis
WFS	Whole Feature Set

CHAPTER 1

INTRODUCTION

The use of information and communications technology (ICT) in teaching and learning processes is increasing rapidly mainly due to the many assumed benefits of computer mediated communication (CMC). The increasing popularity of internet and its capability to offer transparent communication between the diverse computing platforms has been simplified mainly on the processes of providing the learning opportunities to distantly located learners.

The most recent development is the so-called cyber teaching or teaching via cyberspace where, the teaching processes are done using the internet. Educators can now provide services without having a face-to-face classroom environment with students. Interactions between teachers and students can be carried out using communication media such as the computers, internet, e-mails, etc. Students can also obtain information from a wide range of different sources through cyber space or virtual space via the internet.

Another term that is increasingly popular is the e-learning. In e-learning, the development, distribution and enhancement of learning resources are done by using ICT, especially the internet. Nowadays, e-learning has evolved into various learning models based on ICT such as Web-Based Learning (WBE), Computer Based Instruction (CBI), Distance Learning, Distance Education, Cybernetic Learning Environment (CLE), Integrated Learning System (ILS) and Computer Based

Training (CBT). These models offer easier means for collaboration between participants who are physically far from each other.

Communications in e-learning can either be asynchronous (communications are sent and received at different times) or synchronous (communication are sent and received at virtually the same times). Examples of asynchronous communication tools are e-mail, mailing list, discussion board, survey and pools, calendar, and newsgroup. Examples of synchronous communication tools are chat, whiteboard, instant messaging, and audio-video conference.

Compared to synchronous discussions, asynchronous communications present several advantages in more on allowing students to interact with each other at their chosen time. Students are also afforded with more time to reflect their understanding of the course content, to think and to develop substantive comments and also to search for extra information before contributing to the discussions (De Wever et al., 2005; Pena-Shaff and Nicholls, 2004).

These online asynchronous discussions are now commonly utilized as the means of promoting interactions between distance learning course members in tertiary learning (Spatariu et al., 2004). Numerous web-based courses depends on online asynchronous discussions as the computer mediated communication (CMC) tool to increase learning due to the asynchronous discussions' capability to sustain high levels of thinking and to offer a convenient and flexible communication forum to engage students actively (Berge, 2002). Educators are also able to regularly arrange their online class courses into weekly sessions or tutorials and talk about the course theories and models in the discussion forums. The discussions may comprise replies to the educator, questions, arguments, discussions, debates, or to deliberate ideas and thoughts.

Thus, online asynchronous discussions have the potential to make collaboration efforts more rewarding and productive for students by enabling them to communicate at anytime, anywhere, and from many networked locations. Learning in asynchronous discussions is the loosest form of collaborative learning. This

opinion is generally explained as the domain of Computer Supported Collaborative Learning (CSCL). De Wever et al. (2005) found that asynchronous discussion forum takes a central place in CSCL environment. A participant in asynchronous discussions communicates a common interest, exchange information, share ideas, and assist each other. Participants call on each other when they have a problem to solve or something to offer. In an online asynchronous discussion, participants participate freely and have enormous deal of personal choices. By providing prolonged opportunities for students to articulate and read another's thought, such communication networks may lead to effective online asynchronous discussions focused on achieving high order of critical thinking.

Since online asynchronous discussions are crucial to the success of CSCL or CMC, it is important to develop the assessment methodologies to analyze and rate the discussions. A rising number of researchers have undertaken the challenge of generating the techniques to determine and investigate the quality of asynchronous discussions. Spatariu et al. (2004), having evaluated existing literatures, recommended that the mainstream of studies be freely grouped into one of the following four categories; "construct being measured; levels of disagreement, argument structure analysis, interaction based and content analysis". However, there have not been much explorations and evaluations done to develop a framework for integrating some of these existing techniques.

To address these shortcomings, this research proposes a framework for analyzing online asynchronous discussions by integrating content analysis (CA) and social network analysis (SNA). Both CA and SNA are the most known and acceptable methods to analyze online discussions in e-learning environment. This framework takes advantage of the strengths of each method and offers dynamic analysis of the textual messages.

1.1 Problem Background

Content analysis and social network are two popular methods to analyze asynchronous discussion forum (Neuendorf, 2002; Anderson et al., 2001; de Laat, 2002 and Willging, 2005). The purpose of content analysis is to reveal information that is not assigned at the superficial of transcript (De Wever et al., 2005) and SNA has been successful in uncovering relationships not seen with any other conventional method (Willging, 2005). However, there has not been much exploration and evaluation of framework to integrate these methods, although this integration offers advantages in which the transcripts and structural network of interaction are able to be analyzed. It is apparent that the field of CMC or CSCL lack of established methodologies to analyze asynchronous discussion forum.

Many researchers in the field argue for using the content analysis as a vehicle for classifying, analyzing and determining communication transcripts (De Wever et al., 2005; Hendri, 1992). Unfortunately, analysis in CSCL or CMC research focused on surface level characteristic of the communication (Strijbos et al., 2006). The reliability of content analysis of communication transcripts is not, or only briefly, reported in much of the literature arguing for and using content analysis (Ratfell, 2007).

Synthesis of communication transcripts in a virtual learning normally involves analysis of social interaction or using communication tools to observe what participants are talking about, and the motivation that lead them to socially connect with others in virtual spaces. SNA method paired with current developments in software for visualizations which could present a clearer picture of what is happening in the online discussion (Willging, 2005). On the other hand, SNA is typically used to learn the way people participate and interact with each other, particularly into the relationship among participants rather than the discussion content (Wang et al., 2007); discussion content of each participants of the group is not provided.

Furthermore, existing Learning Management System (LMS) that generally provide online discussion cannot afford a comprehensive evaluation on the content of

the transcripts and the level of interaction among participants. The instructors of online classes are not offered with those structural indicators that would let them to assess the participation and interaction among students in their classes. In many instances, only statistical information, such as frequency of postings is encompassed, but this is not a very helpful measurement for the interaction activity. Therefore, this research explores the analysis process of online asynchronous discussion (OAD) qualitatively and quantitatively. We focus upon Content Analysis (CA) and Social Network Analysis (SNA), which are two popular methods employed by educators and researchers to analyze online discussion in e-learning environment. Although these two methods are well established, the techniques remain commonly manual, requiring human skilled specialists. However, human categorization is not an effective way for number of reasons; time-consuming, labour intensive, lack of consistency in category and costly.

Presently, these two methods of analysis are conducted and studied independently and not embedded into any LMS. However, online discussion will greatly benefit if these two methods can be merged to form one comprehensive and coherent method for studying OAD. Hence, this research proposes a framework for integrating CA with SNA (CASNA). CASNA can be applied and embedded into Moodle which is the mostly known LMS.

Usually, SNA determines the level of participation in online asynchronous discussion based on a number of messages, even though this message consists of only one word or one sentence. However, it is not fair for participants who sent or posted a message that consist of many sentences that represent many ideas but only being considered as one contribution of participation. Participants that posted or replied many sentences and posted or replied few sentences should not be treated equivalent level of participation. Those participants who posted long message that represent many ideas should be recognized to contribute more in the participation. Hence, this research also introduces sentence as the unit interaction instead of message to assess the level of participation among students.

1.2 Problem Statements

This research proposed a framework to integrate content analysis and social network for analyzing asynchronous discussion forum within collaborative learning environment. Integration of these methods can analyze not only the transcript but also the structural network of interaction.

The main research question that was posed for exploration in this study: *"How to build a framework for integrating content analysis and social network analysis in order to analyze online asynchronous discussion in collaborative learning that can give more reliable and meaningful analysis?"*

The sub-questions of the main research question are as follows:

(i) Why existing online asynchronous discussion forum cannot afford a comprehensive evaluation regarding the communication transcripts and communication structures of the participant?

(ii) What is the appropriate unit analysis for analyzing online asynchronous discussion?

(iii) What is the effective way to analyze online asynchronous discussion to support Learning Management System in e-learning?

(iv) How to develop a framework to generate a tool for integrating content analysis and social network analysis?

1.3 Research Objectives

Based on the problem statements mentioned above, this research encompasses a set of objectives that are associated with milestones of the research process. The research objectives are mentioned below.

(i) To investigate the current content analysis and social network analysis method in supporting online asynchronous discussion.

(ii) To identify the unit of analysis for analyzing online asynchronous discussion.
(iii) To propose a new framework for integrating content analysis and social network analysis that can be embedded into Learning Management System.
(iv) To demonstrate the applicability of the proposed framework through a development of supporting tools.

1.4 Research Scopes

Scope of the study is divided into three items, i.e. scope of process that is performed, scope of framework in which the method is applied in, and scope of tool supports and case studies. Each of them is described below.

Scope of process: Framework has been proposed particularly to analyze online discussion in learning management system. Specifically, this analyze includes:

(i) Automating content analysis of discussion transcripts that previously do not embedded into LMS;
(ii) Automating social network analysis of discussion transcripts which provides the structural indicators;
(iii) Creating adjacency matrix, centrality, density, map and graph of students' interactions;
(iv) Categorizing the text and determining the role of each participant in online discussion.

Scope of framework: CASNA has been proposed as a new framework for analyzing online asynchronous discussion. The discussion structures follow the Moodle's structure version 1.9.5 and focused on text-based analysis. The framework was designed for educational setting especially for online asynchronous discussion and focuses on student to student interaction and student to teacher interaction or in simple word participant to participant interaction.

Scope of tool support and case studies: This research has also built a supporting prototype tool to implement the framework in simplifying the analysis process efficiently and comprehensively. The supporting prototype tool is called CASNA tools. The tools consist of four components. The first component is text analyzer tool for segmenting the message into sentences due to this research has been selected sentence as unit analysis. The second component is human content analysis to analyze and code the list of text manually before being analyzed by artificial intelligence. The third component is content analysis using artificial intelligence to categorize the list of text of online discussion automatically. The fourth component is social network analysis tool which is activated and run in order to detect and capture the changes that has been performed during online discussion.

The case study of online discussion forum held on Moodle in e-learning system of Universiti Teknologi Malaysia (UTM) have been applied in CASNA tools to demonstrate the applicability of the framework through the tool itself as well as to evaluate the results obtained by the tool.

1.5 Research Significance

Discussion forums are a powerful way to conduct conversations online and an important part of online learning. Participants give credence on these asynchronous forums to connect one another in ways that potentially promote knowledge sharing. In spite of the significance of these forums, the framework that contains methods and techniques to analyze the quality and quantity of OAD is needed.

Insight of a collaborative learning within a certain analysis transcripts alone is not sufficient. The analysis of network structure of the discussion must also be taken into account. Therefore, this research employs content analysis and social network technique to analyze OAD for students participating in a course. Content analysis was used to evaluate the quality of communication transcripts and social network analysis was employed to analyze network structures of the response relations between participants during discussions. This framework takes advantages of the

strengths of each method and offers a dynamic analysis of the textual messages. This allows researchers to examine both the learning conversation skill and interaction structure, simultaneously. In the past, the majority of study in CMC or CSCL used these methods separately. It is expected to be more informative to educators as well as researchers in measuring the quality (i.e. content) and quantity (i.e. interaction) of OAD by providing the best of both methods.

1.6 Thesis Organization

This thesis is divided into eight chapters that discuss on specific issues associated to content analysis and social network analysis for analyzing online asynchronous discussion. It also describes the proposed framework for integrating these two methods in details. The thesis is organized in the following outline.

Chapter 1: Presents an introduction describing the role of ICT in education setting, especially in e-learning. It also discusses the important of online asynchronous discussions in CMC or CSCL, followed by problem background, problem statements, the objective, scope, thesis outline and the significance of the research.

Chapter 2: Discusses the literature review of content analysis and social network analysis. First, it discusses the fundamentals issues which will be used to realize the working of existing methods. Next, explains the evaluation of the method and related issues. It begins with evaluation on content analysis, followed by social network analysis and evaluation on several researches that used both of these methods. The evaluation focus is primarily on their capability to support the analysis of online discussion.

Chapter 3: Discusses the research methodology that describes research design and formulation of research problems and validation considerations followed by the procedure that is utilized in this research work. It also describes the

operational frameworks, data collections and data analysis techniques that are used to conduct the research.

Chapter 4: Explains the framework for integrating content analysis and social network analysis. It begins with the design of conceptual framework that represents the whole of the processes. This chapter also describes content analysis of discussion transcripts as well as social network analysis. At the end of the chapter, integrated methods in the proposed framework are discussed.

Chapter 5: Describes the approach for text categorization of collaborative learning skill in online discussion. Two popular text classifiers; neural network and support vector machine were employed and compared by experimental result. The procedure on how the text of online discussion transform into a suitable data format that can be processed by neural network and support vector machine also will be discussed.

Chapter 6: Presents the design and implementation of social network analysis. First section describes the design of network analysis followed by creating and representing indicators of SNA. The algorithm for creating map and graph to visualize the students' interaction will be discussed in more detail. Second section demonstrates the experimentation, including the result and ended by summary.

Chapter 7: Portrays the evaluation and discussion of the proposed framework through the tool. For measuring the capability of the framework, testing and evaluating on several scenarios are conducted. Investigating the collaborative learning role and category will be discussed more detail. For evaluating the dynamically of social network analysis, three scenarios are applied; based on phase, based on message versus based on sentence and based on number of link versus based on sum of weight. At the end of this chapter will be closed by summary.

Chapter 8: Provides the statements on research achievements, research contributions and conclusion of this thesis. This is followed by the research summary and suggestions for future works.

REFERENCES

Alvaro, F., and Joanne, L. (2007). Interaction Visualization in Web-Based Learning using iGraphs. *Proceedings of Hypertext*. Manchester, UK.

Anderson, T., Rourke, L., Garrison, D. R., and Archer, W. (2001). Assessing teaching presence in a computer conference context. *Journal of Asynchronous Learning Networks*, 5(2).

Aviv, R., Erlich, Z., Ravid, G., and Geva, A. (2003). Network Analysis of Knowledge Construction in Asynchronous Learning Networks. *Journal of Asynchronous Learning Networks*, 7 (3).

Benkhalifa, M., Bensaid, A. and Mouradi, A. (1999). Text Categorization using The Semi-Supervised Fuzzy C-Means Algorithm. *18th International conference of the north american fuzzy information processing society* – NAFIPS, 561-568.

Berelson, B. (1952). *Content Analysis in Communication Research*. Illinois: Free Press.

Berge, Z.L. (2002). Teaching How to Learn Through Online Discussion. In E. Biech (Ed.) Pfeiffer HRD Annual. Volume 1: Training. San Francisco, CA: Jossey-Bass/Pfeiffer, 307-313.

Biddle, B. J., and Thomas, E. J. (1996). *Role Theory: Concepts and Research*, John Willey & Sons, Inc.

Butts, C. T. (2006). Exact Bounds for Degree Centralization, *Social Networks* vol. 28, 283–292.

Chen, Z., Ni, C., and Murphey, Y. L. (2006). Neural Network Approaches for Text Document Categorization. International Joint Conference on Neural Networks, Sheraton Vancouver Wall Centre Hotel, 16-21.

Cruz, I. F. (2009). Graph Drawing Tutorial. From http://www.graphdrawing.org/literature/gd-constraints.pdf

Daradounis, T., Martinez-Monas, A., and Xhafa, F. (2004). An Integrated Approach for Analyzing and Assessing the Performance of Virtual Learning Groups. In G. de Vreede, L. A. Guerrero and G. M. Raventos (Eds), *Lectures Notes in Computer Science* (Inc 3198), 289-304, Heidelberg: Springer-Verlag.

de Laat, M. (2002). Network and Content Analysis in an Online Community Discourse, In: G Stahl (Ed.) Proceedings of *Computer Support for Collaborative Learning* (CSCL) Conference. 626-626.

de Laat, M., Lally, V., Lipponen, L., Simon, R., J. (2007). Investigating Patterns of Interaction in Networked Learning and Computer-Supported Collaborative Learning: A Role for Social Network Analysis. *Journal of Computer Supported Collaborative Learning.*

De Wever, B., Schellens, T., Valcke, M. and Van Keer, H. (2005). Content Analysis Schemes to Analyze Transcripts of Online Asynchronous Discussion Groups: A Review, *Computers & Education* 46 (2005) (1), 6–28.

Dillenbourg, P. (1999). What do You Mean by 'Collaborative Learning'? In P. Dillenbourg (Ed.), *Collaborative Learning. Cognitive and Computational Approaches*. Oxford: Elsevier Science 1-19.

Dumais, S. (1998). Using SVMs for Text Categorization. In IEEE Intelligent Systems, M. A. Hearst, B. Schölkopf, S. Dumais, E. Osuna, and J. Platt, Eds: *Trends and Controversies—Support Vector Machines*, vol. 13.

Erlin, Norazah, Y., and Azizah, A. R. (2009). Students' Interaction in Online Asynchronous Discussion Forum: A Social Network Analysis, International *Conference on Education Technology and Computer* (ICETC), 25-29.

Erlin, Norazah, Y., and Azizah, A. R. (2010). Messages Segmentation of Asynchronous Text-Based Discussion in Social Network Analysis. *2nd International Conference on Education Technology and Computer* (ICETC).

Fahy, P. J., Crawford, G., Ally, M., Cookson, P., Keller, V. and Prosser, F. (2001). The Development and Testing of A Tool for Analysis of Computer Mediated Conferencing Transcripts. *Alberta Journal of Educational Research, 46*(1), *Spring,* 85-88.

Farkas, J. (1994). Generating Document Clusters using Thesauri and Neural Networks. *Canadian conference on Electrical and Computer Engineering*, Vol.2, 710-713.

Fleiss, J. L. (1971). Measuring Nominal Scale Agreement among Many Raters. *Psychological Bulletin,* 76,378-382.

Freelon, D. (2009). Worked Examples for Nominal Intercoder Reliability. From http://dfreelon.org/recal/recal-worked-examples.pdf

Freeman, L. C. (2002). *The Study of Social Networks* From: http://www.insna.org/INSNA/na_inf.html

Gabrilovich, E. and Markovitch, S. (2004). Text Categorization with Many Redundant Features: Using Aggressive Feature Selection to Make SVM Competitive with C4.5. *Proceedings of the 21st International conference on machine learning*, 321-328.

Galleta, D. F., and Heckman, R. (1990). A Role Theory Perspective on End-User Development. *Journal of Information Systems Research*, IV, N2, 168-187.

Garrison, D. R. (1992). Critical Thinking and Self-Directed Learning in Adult Education: An Analysis of Responsibility and Control Issues. *Adult Education Quarterly, 42*(3), 136-148.

Garrison, D. R., Anderson, T., and Archer, W. (2000). Critical Inquiry in A Text-Based Environment: Computer Conferencing in Higher Education. *The Internet and Higher Education, 2*(2-3), 87-105.

Garrison, D. R., Anderson, T., and Archer, W. (2001). Critical Thinking, Cognitive Presence, and Computer Conferencing in Distance Education. *American Journal of Distance Education*, 15(1), 7-23.

Gunawardena, C. N., Lowe, C. A., and Anderson, T. (1997). Analysis of Global Online Debate and The Development of An Interaction Analysis Model for Examining Social Construction of Knowledge in Computer Conferencing. *Journal of Educational Computing Research*, 17(4), 397–431.

Hakkinen, P., Jarvela, S., and Makitalo, K. (2003). Sharing Perspectives in Virtual Interaction: Review of Methods of Analysis. In B. Wason, S. Ludvigson and U. Hoppe (Eds.), Designing for change in networked learning. *Proceedings of The International Conference on Computer Support for Collaborative Learning*. 395–404.

Hamza. A. A. (2008). Back Propagation Neural Network Arabic Characters Classification Module Utilizing Microsoft Word. *Journal of Computer Science* 4 (9): pp. 744-751.

Hanneman, Robert A. and Mark Riddle. (2005). *Introduction to Social Network Methods.* Riverside, CA: University of California, Riverside (published in digital form http://faculty.urc.edu/~hanneman/)

Hara, N., Bonk, C. J., and Angeli, C. (2000). Content Analyses of On-Line Discussion in An Applied Educational Psychology Course. *Instructional Science*, 28(2),115–152.

Hara, N., Bonk, C., and Angeli, C. (1998). *Content Analysis of Online Discussion In Educational Psychology Courses.* Paper presented at the meeting of the Society for Information Technology and Teacher Education 98, Washington, DC. Retrieved January 17, 2002, from http://www.coe.uh.edu/insite/elec_pub/HTML1998/re_hara.htm

Haythornthwaite, C. (2002). Building Social Networks via Computer Networks: Creating and Sustaining Distributed Learning Communities. In K. A. Renninger and W. Shumar (Eds.), *Building Virtual Communities: Learning and Change in Cyberspace,* 159–190. Cambridge: Cambridge University Press.

Henri, F. (1992). Computer Conferencing and Content Analysis. In A. R. Kaye (Ed.), *Collaborative Learning Through Computer Conferencing*. London: Springer.

Hidalgo, J.M.G. (2003). Text Representation for Automatic Text Categorization. Eleventh Conference of the *European Chapter of the Association for Computational Linguisitics*, EACL,from http://www.conferences.hu/EACL03/

Howell Richardson, C., and Mellar, H. (1996). A Methodology for The Analysis of Patterns of Participation Within Computer Mediated Communication Courses. *Instructional Science, 24,* 47-69.

http://dfreelon.org/utils/recalfront/

http://svmlight.joachims.org/svm_multiclass.html

Hua Li, C. and Park, S.C. (2006). A Novel Algorithm for Text Categorization using Improved Back-Propagation Neural Network. In L. Wang et al (Eds): FSKD, 2006, LNAI 4223, 452-460.

Inch, E. S., and Warnick, B. (2002). *Critical thinking and communication*. Boston: Allyn and Bacon.

Jarvela, S., and Hakkinen, P. (2002). The Levels of Web-Based Discussion: Using Perspective-Taking Theory As An Analysis Tool. In H. Van Oostendorp (Ed.), *Cognition in A Digital World*, 77-95. Mahwah, NJ: Erlbaum.

Joachims, T. (1998). Text Categorization with Support Vector Machines: Learning with Many Relevant Features. *10th European Conference on Machine Learning*, 137-142.

Kadushin, C. (2005). "Who Benefits from Network Analysis: Ethics of Social Network Research", *Social Networks*, vol. 27, 139.

Kazama. J and Tsujii, J. (2005). Maximum Entropy Models with Inequality Constrains: A Case Study on Text Categorization. *Machine Learning*. *60*(1-3), 159-194.

Kecman. V. (2005). Support Vector Machines – An Introduction. In ed. L. Wang (Wang Ed.), *Support Vector Machines: Theory and Applications*, Springer – Verlag Berlin Heidelberg.

Krebs, V. (2003). Data Mining E-mail to Discover Social Networks and Communities of Practice, from http://www.orgnet.com/ email.htm,

Krebs, V. (2006), How to do Social Network Analysis, From http://www.orgnet.com/sna.html

Krippendorff, K. (1980). *Content Analysis, An Introduction to Its Methodology.* Thousand Oaks, CA: Sage Publications.

Krippendorff, K. (2004). *Quantitative Content Analysis: An Introduction to Its Method.* Beverly Hills: Sage Publications.

Lam, W. and Low, K.F. (1997). Automatic Document Classification based on Probabilistic Reasoning: Model and Performance Analysis. Proceedings of the *IEEE International Conference on Systems, Man and Cybernetics*, Vol.3, 2719-2723.

Li, C. H., and Park, S. C. (2006). A Novel Algorithm for Text Categorization using Improved Back-Propagation Neural Network. In L. Wang et al. (Eds.): FSKD, LNAI 4223, 452-460. Springer-Verlag Berlin Heidelberg.

Lockhorst, D., Admiraal, W., Pilot, A., and Veen, W. (2003). *Analysis of Electronic Communication using 5 Different Perspectives*. Paper presented at *ORD*.

Lombard, M., Snyder-Duch, J., and Bracken, C. C. (2010). Practical Resources for Assessing and Reporting Intercoder Reliability in Content Analysis Research Projects. From http://astro.temple.edu/~lombard/reliability/

Lui, A.K.F., Li, S.C. and Choy, S. O. (2007). An Evaluation of Automatic Text Categorization in Online Discussion Analysis, *Seventh IEEE International Conference on Advanced Learning Technologies,* 205-209.

Ma, L, Shepherd, J and Zhang, Y. (2003). Enhancing Text Clasification using Synopses Extraction. *The fourth International Conference on Web Information System Engineering.* 115-124.

Mason, R. (1992). Analyzing Computer Conferencing Interactions. *Computers in Adult Education and Training, 2*(3)*,* 161-173.

McManus, M, and Aiken, R. (1995). Monitoring Computer-Based Problem Solving. *Journal of Artificial Intelligence in Education*, *6*(4), 307-336.

Murphy, E., Ciszewska-Carr and Maria A. Rodriguez Manzanares, M. A. R. (2006). Methodological Issues in the Content Analysis of Online Asynchronous Discussions: Unitizing, Reliability, and Latent Content. *Journal of The Research Center for Educational Technology.*

Muukkonen, H., Lakkala, M., and Hakkarainen, K. (2001). Characteristics of University Students' Inquiry in Individual and Computer-Supported Collaborative Study Process. In P. Dillenbourg, A. Eurelings, and K. Hakkarainen (Eds.), *European Perspectives on Computer-Supported Collaborative Learning*, 462–469.

Neuendorf, K. A. (2002). *The Content Analysis Guidebook*. Thousand Oaks, CA: Sage Publications.

Neural Network Toolbox for MATLAB. From http://www.mathworks.com/products/neural-net/

Newman, G., Webb, B. and Cochrane, C. (1995). A Content Analysis Method to Measure Critical Thinking in Face-To-Face and Computer Supported Group Learning. *Interpersonal Computing and Technology, 3*(2), *56-77*

Newman, M. E. J. (2005). A Measurement of Betweenness Centrality Based on Random Walks. *Social Networks*, vol. 27, 39.

Ng, H.T., Goh, W.B. and Low, K.L. (1997). Feature Selection, Perception Learning, and Usability Case Study for Text Categorization. 20th Annual international ACM-SIGIR *Conference on Research and Development in Information Retrieval*, 67-73.

Nunamaker, J.F., Chen, M. and Purdin, T.D.M. (1991). System Development in Information Systems Research. *Journal of Management Information Systems,* 7 (3), 1991, 89-106.

Nussbaum, E. M., Hartley, K., Sinatra, G. M., Reynolds, R. E., and Bendixen, L. (2002). Enhancing The Quality of On- Line Discussions. *The annual meeting of the American Educational Research Association*, New Orleans, LA.

Oriogun, P. K. (2006). Content Analysis of Online Transcripts: Measuring Quality of Interaction, Participation and Cognitive Engagement within CMC Groups by Cleaning of Transcripts, from http://www.itdl.org/Journal/Mar_06/article03.htm

Paul, S., Roberts, M. C., Andrew, E. S., and Janet, W. (2009). Use of Automatic Content Analysis Tool: A Technique for Seeing Both Local and Global Scope. *Int. J. Human-Computer Studies* 67, 424-436.

Pena-Shaff, J. B., and Nicholls, C. (2004). Analyzing Student Interactions and Meaning Construction in Computer Bulletin Board Discussions. *Computers & Education*, 42, 243–265.

Peter, S., et al., (2010). A re-Examination of The Community of Inquiry Framework: Social Network and Content Analysis. *Internet and Higher Education* 13 (2010) 10–21.

Poon, N. and Daniel, B. K. (2006). Social Network and Content Analysis of Interactions in A Video-Mediated Virtual Community. Proceedings of the *Sixth International Conference on Advanced Learning Technologies* (ICALT'06).

Porter, M.F. (1997). An Algorithm for Suffix Stripping. In Morgan Kaufmann, *Readings in Information Retrieval*, 313-316. Morgan Kaufmann Publishers Inc.

Ratfell (2007). The Reliability of Content Analysis of Computer Conference Communication. *Journal of Computer & Education, Elsevier*. 49.230-242.

Reffay, C., and Chanier, T. (2003). Social Network Analysis Used for Modelling Collaboration in Distance Learning Groups. *Intelligent Tutoring Systems: 6th International Conference,* ITS.

Rourke, L., Anderson, T., Garrison, D. R., and Archer, W. (1999). Assessing Social Presence in Asynchronous Text-Based Computer Conferencing. *The American Journal of Distance Education, 14*(2). Retrieved March 7, 2001, from http://cade.athabascau.ca/vol14.2/rourke_et_al.html

Rourke, L., Anderson, T., Garrison, D. R., and Archer, W. (2001). Methodological Issues in the Content Analysis of Computer Conference Transcripts, *International Journal of Artificial Intelligence in Education* 12, 8–22.

Schaeffer, E. L., McGrady, J. A., Bhargava, T., and Engel, C. (2002). Online Debate to Encourage Peer Interactions in The Large Lecture Setting: Coding and Analysis of Forum Activity. *The annual meeting of the American Educational Research Association,* New Orleans, LA.

Scott, J. (2002). *Social Network Analysis: A Handboo*k, 2nd ed., London: Sage.

Sebastiani, F. (2002). Machine Learning in Automated Text Categorization. *ACM Computing Survey*, 34(1), 1-47.

Sinclair, P. A. (2009). Network Centralization with The Gil Schmidt Power Centrality Index. *Social Networks*, vol.31, 214–219.

Soller A. (2004). Computational Modeling and Analysis of Knowledge Sharing in Collaborative Distance Learning. *User Modeling and User-Adapted Interaction* 14: 351–381.

Soller. A. (2001). Supporting Social Interaction in An Intelligent Collaborative Learning System, *Journal of Artificial Intelligence in Education*, 12, 40-62.

Soucy, P and Mineau, G.W. (2001). A Simple k-NN Program for Text Categorization. *The first IEEE International Conference on Data Mining* (ICDM_01), Vol. 28, 647-648.

Spatariu, A., Hartley, K. and Bendixen, L.D. (2004). Defining and Measuring Quality in On-line Discussion. *Journal of Interactive Online Learning*. Vol 2(4).

Strijbos, J. W., Martens, R., Prins, J., and Jochems, W. (2006). Content Analysis: What are They Talking About? *Journal of Computer & Education*, Elsevier. 46. 29-48.

Suh, H. J., and Lee, S. W. (2006). Collaborative Learning Agent for Promoting Group Interaction, *ETRI Journal,* Vol 28, No.4, 461-474.

Susan, L., Lin, P., and Wang, Y. (2007). Studying the Effectiveness of the Discussion Forum in Online Professional Development Courses. *Journal of Interactive Online Learning,* vol.6, number 3.

Trochim, W. M. K. (2006). Unit of Analysis, retrieved October 20, 2006, from http://www.socialresearchmethods.net/kb/unitanal.htm

Veldhuis-Diermanse, A. E. (2002). Csclearning, Participation, Learning Activities and Knowledge Construction in Computer-Supported Collaborative Learning in Higher Education. Wageningen: Grafisch Service Centrum Van Gils.

Wang, Y., and Li, X. (2007). Social Network Analysis of Interaction in Online Learning Communities. *Seventh IEEE International Conference on Advanced Learning Technologies* (ICALT 2007)

Wang, T. Y., and Chiang, H. M. (2007). Fuzzy Support Vector Machine for Multi-class Text Categorization. Journal of Information Processing and Management, Elsevier, 43, 914-929.

Wasserman, S., and Faust, K. (1997). *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.

Weber, R. P. (1990). *Basic Content Analysis*. 2nd ed., Newbury Park, CA: Sage.

Weinberger, A. and Fischer, F. A Framework to Analyze Argumentative Knowledge Construction in Computer-Supported Collaborative Learning, *Computers & Education* 46 (2005) (1), 71–95.

Willging, P. A. (2004). Using Social Network Analysis to Study Teamwork Dynamics. In T. N. Garavan, E. Collins, M. J. Morley, R. Carbery, C. Gubbins and L. Prendeville (Eds.), *Proceedings of the Fifth UFHRD/AHRD Conference*, University of Limerick, Ireland.

Willging, P. A. (2005). Using Social Network Analysis Techniques to Examine Online Interactions. *Journal of US-China Education Review*, Vol 2, No.9 (Serial No.10).

Yang, Y. and Pedersen, J. O. (1997). A Comparative Study on Feature Selection in Text Categorization. Proceedings of ICML-97, *14th International Conference on Machine Learning*, 412-420.

Yu, B., ben Xu, Z., hua Li, C. (2008). Latent Semantic Analysis for Text Categorization using Neural Network. *Journal of Knowledge-Based Systems*, 21, 900-904.

Zigur, I., and Kozar, K. (1994). An Exploratory Study of Roles in Computer Supported Groups, MIS Quarterly, 18(3), 277-297.