A CONCEPT SPACE APPROACH TO SEMANTIC EXCHANGE

by

Tobun Dorbin Ng

A Dissertation Submitted to the

COMMITTEE on BUSINESS ADMINISTRATION

In Partial Fulfilment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY
WITH A MAJOR IN MANAGEMENT

In the Graduate College

THE UNIVERSITY OF ARIZONA

2000

STATEMENT BY AUTHOR

SIGNED: ______________________________

ACKNOWLEDGMENTS

It has been an exceptional journey - one that I had never dreamed of. My research career was started not long after I became an MIS undergraduate student in 1989, when I attended Dr. Hsinchun Chen's first class at The University of Arizona. I am deeply indebted to Dr. Chen for his encouragement, advice, mentoring, and research support throughout my undergraduate, master's, and doctoral studies. I also truly appreciate his patience and tolerance during my numerous mishaps. This dissertation is part of the research carried out through his vision throughout last ten years.

I am fortunate to have the opportunity to work with a group of energetic people in Dr. Chen's AI Lab. I have enjoyed every moment that we have worked together including all those late night lab activities. All former and present AI Lab members have taught me many things about life. I appreciate all their friendships and their collective encouragement to finish this dissertation. I want especially to thank Chienting Lin, Thian-Huat Ong, Marshall Ramsey, Yohanes Santoso, Harry Li, Samuel Yim, and Chris Schuffels for sharing their technical wisdom, and to thank Andrea Houston, Joanne Martinez, Kris Tolle, and Rosie Hauck for sharing their research ideas.

Finally, it is impossible to have my research career without my parents' love and support, as well as my family and friends' encouragement. This dissertation is dedicated to them.

To all of you, thank you.

DEDICATION

This dissertation is a result of collective efforts from my parents, family, and friends. They often give me valuable advice, reminders, and wisdom about life, when I had a hard time understanding. However, they always give me room to explore and make mistakes.

This dissertation is dedicated to my parents, who have given me all their love and support and let me freely do whatever I want. Without them, there is no way I could possibly have accomplished this. Their understanding on the value of education is truly beyond my comprehension. I am just a lucky beneficiary. Along the way, my sister Karen and brother Gavin have shared their caring thoughts.

I have also been very lucky to have my uncle and aunt, Bob and Hannah Ng, and my cousins, Belinda and Gigi, throughout my stay in Arizona. They have helped me learn about life and live a good one. They have also given me comfort and advice whenever I needed them.

Throughout my ten plus years in Arizona and Dr. Hsinchun Chen's AI Lab, I have seen many people come and go; however, I am very grateful that many of them have taken me as their friend. They have taught me how to live, love, and feel; these, in turn, become the catalyst to my desire to finish this dissertation.

I truly thank all of them from the bottom of my heart.

We are loved beyond our capacity to comprehend. - Jewel Kilcher

TABLE OF CONTENTS

TABLE OF CONTENTS – *Continued*

TABLE OF CONTENTS – *Continued*

TABLE OF CONTENTS – *Continued*

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

This dissertation work investigates the use of information technologies that clarify semantic meaning to help users elaborate their information needs by providing library-specific knowledge to the information seeking process. The research involved two interdependent semantic technologies: concept space consultation and library-specific, domain-specific, automatically generated concept spaces.

The concept space consultation phase used spreading activation algorithms - branch-and-bound and Hopfield net algorithms - to explore knowledge sources in specific domains. This research demonstrated the comparable effectiveness of exploration of a library database using a man-made classification scheme and thesaurus as opposed to an automatically generated concept space. The results showed that the use of spreading activation algorithms identified more relevant concepts than the use of the manual browsing method.

The concept space technique automatically identifies and extracts concept from a library collection while at the same time computing the strength of associations between concepts. This research demonstrated that the concept space technique was able to create human-recognizable concepts and their associations. In addition, the technique could be scaled to generate very large library-specific concept spaces for a very large underlying library collection.

Moreover, the interdependent use of both semantic technologies creates a semantic medium for users and library-specific knowledge sources to exchange content with context - context in user information need and that in corporeal knowledge.

CHAPTER 1

INTRODUCTION

This dissertation work investigated the use of information technologies that clarify semantic meaning to help users elaborate their information needs by specifying their library-specific knowledge during the information seeking process. That process starts with expression of a user information need, which makes it dynamic, cognitive, and user-oriented. However, existing information retrieval processes take only the end result of the cognitive aspect as query input and implicitly assume that the query input is the true (or almost true) representation of the user information need. Is this assumption always true?

The research in knowledge discovery, data mining, and machine learning has demonstrated the ability to generate knowledge of underlying data collections. In addition to automatic computational methods, extensive human effort has been invested in building domain-specific thesauri and classification schemes. However, all these semantic-bearing entities are targeted to represent knowledge coverage in particular domains and to capture semantic relationships between identified concepts. Can these knowledge sources adequately serve users' information needs?

When the information seeking process reaches the stage of retrieving information from search engines, users must perform two tasks - making a query and evaluating a list of retrieved documents. If users are not satisfied with the results, they will repeat the same two tasks with different queries or they will give up their searches. This "blackbox" information retrieval approach has existed since the beginning of the information retrieval field in the 1960s and has since been popularized by the Internet. Nowadays, users are very familiar with what to expect from search engines - a possibly long ranked list of documents that they may never go through. In addition, in information retrieval practices, a common belief has been created that "relevant" documents appear only at the top 10 or 20 of ranked retrieved result.

The role of users in this "blackbox" approach focuses on evaluating retrieved documents for their relevancy. This approach downplays, if it does not ignore, the necessity of an information need having been fully expressed. This dissertation work adopted a user-centric and interactive approach to helping users elaborate their information needs with library-specific knowledge and simultaneously gain insight into a library's offerings related to their information needs. Through semantic communication between users, users are enabled to express their information needs in the context of the library-specific knowledge of the target information source. Shared contextual information also gives users knowledge of the potential value of retrieved information. Under this approach, the information seeking process becomes a semantic journey from the initial expression of a user's information need to acquisition of knowledge of relevant information rather than a purely computational operation.

In order to explore semantic exchange between users and collections, this dissertation research focused on two areas: interactive consultation with knowledge sources and automatic generation of semantic-bearing knowledge sources from corresponding libraries. Concept space consultation demonstrates a semantic exchange between users and knowledge sources during query expression. Concept space generation shows how large-scale semantic-bearing knowledge sources can be automatically generated.