CHAPTER 6

CONCLUSIONS

This dissertation composes of two tracks of information technology research and development to help users elaborate their information needs when seeking information. The first track is *concept space consultation*, which emphasizes the algorithmic and iterative use of both man-made and machine-generated knowledge sources to present related concepts to users' queries. The second track is *concept space generation*, which shifts the focus from the user aspect to how such knowledge sources can be automatically generated from large library collections.

6.1 Contributions

6.1.1 Dynamic Use of Knowledge Sources

Knowledge sources are static once generated as described in Chapter 2. However, the concept space consultation process dynamically taps into knowledge sources and returns related concepts to users. In addition, users do not need to comprehend the overall knowledge and its structure of any knowledge source. They only need to concentrate on evaluating what is being revealed.

1. *Bringing knowledge to users*

Even though knowledge sources are widely available to users, users often do not use them because of unfamiliarity with different knowledge representations and structures. The concept space consultation provides a transparent access to multiple knowledge sources and brings related concepts to users.

2. *Alleviating from time-consuming and cognitively demanding browsing process*

   Compared with manual browsing, the algorithmic consultation process derives a relatively small set of closely related concepts from multiple knowledge sources. Therefore, the time and cognitive demand for evaluating the usefulness of each concept are less.

6.1.2 Complementary to Man-made Thesauri

Concept associations resemble human-recognizable semantic associations. Moreover, the concept space technique captures concepts and their relationships with respect to underlying library-specific and domain-specific collection as effectively as do man-made thesauri. However, concept spaces and man-made thesauri have two major differences.

1. A concept space provides a more extensive list of associated concepts than does a man-made thesaurus.

2. A concept space can be generated within days while a man-made thesaurus requires years to build.

In a particular domain, an automatically generated concept space and man-made thesaurus compliment each other to capture associative concepts.

6.1.3 Context-specific Concept Spaces

A concept space provides context-specific information with respect to the underlying library database. For example, Figure 6.1 shows two lists of concepts related to "Artificial Intelligence" in the *computer science* and *cancer research* domains, respectively. Each is a ranked list, with the most closely related concepts placed at the top. Both lists have overlapping concepts such as "expert systems" and "neural networks". However, the list in the computer science domain veers toward the context of "learning systems" and "cognitive systems" while the one in the cancer research domain more closely approaches the context of "decision support system" and "decision trees."

The advantage of having context-specific concept spaces is that users can more readily detect whether knowledge sources contain the information they need. This is especially important for information retrieval. If users sense that an appropriate context exists in a knowledge source, it will be very likely that the relevant information is going to be in the underlying library and eventually can be retrieved. If users do not sense any relevant context in a given knowledge source, they will have sufficient information to decide against perusing the underlying library for information.

```
''Artificial Intelligence''
    in Computer Science Domain          in Cancer Research domain
    ----------------------------------- ---------------------------------------
 1  EXPERT SYSTEMS                      Expert Systems
 2  NEURAL NETS                         Algorithms
 3  KNOWLEDGE BASED SYSTEMS             Artifical Neural Networks
 4  INFERENCE MECHANISMS                Breast Neoplasms
 5  LEARNING SYSTEMS                    Computer Systems
 6  KNOWLEDGE REPRESENTATION            DECISION SUPPORT SYSTEM
 7  NEURAL NETWORKS                     Decision Trees
 8  COGNITIVE SYSTEMS                   Discriminant Analysis
 9  PROBLEM SOLVING                     Fuzzy Logic
10  ARTIFICIAL NEURAL NET               Information Systems
11  FORMAL LOGIC                        Neural Networks
12  ARTIFICIAL INTELLIGENCE TECHNIQUES  Pattern Recognition
13  LOGIC PROGRAMMING                   Reproducibility of Results
14  PATTERN RECOGNITION                 SYSTEMS
15  INTELLIGENCE                        Software Design
16  NATURAL LANGUAGES                   Structure-Activity Relationship
17  ROBOTS                              artificial intelligence techniques
18  EXPERT SYSTEM                       knowledge bases
19  FUZZY LOGIC                         knowledge-based systems
20  IMAGE PROCESSING                    Ann Arbor
21  KNOWLEDGE ACQUISITION               Artificial Intelligence research
22  COMPUTER VISION                     BONE MARROW MORPHOLOGY
23  COGNITIVE SCIENCE                   Bayes Theorem
24  KNOWLEDGE ENGINEERING               Brain Neoplasms
25  SEARCH PROBLEMS                     CANCER PATIENTS
26  ARCHITECTURE                        CARCINOGENICITY IN RODENTS
27  SOFTWARE ENGINEER                   CLINICAL DATABASES
28  SOFTWARE ENGINEERING                COMPUTER EXPERT SYSTEM
29  PHILOSOPHICAL ASPECTS               COMPUTER PROGRAM
30  KNOWLEDGE-BASED SYSTEM              COMPUTER SYSTEM
31  COMPUTER SCIENCE                    COMPUTER VISION
32  DISTRIBUTED ARTIFICIAL INTELLIGENCE CONTROL THEORY
33  AI TECHNIQUE                        Carcinogenicity Tests
34  MACHINE LEARNING                    Cervix Uteri
35  INTELLIGENT SYSTEMS                 Chamness GC
36  GENETIC ALGORITHMS                  Clark GM
37  INTEGRATION                         Clinical Protocols
38  MOBILE ROBOTS                       Comparative Study
39  BUILDING                            Computer Automated Structure Evaluation
40  MODELLING                           Computer Graphics
```

Figure 6.1: Context-specific concept space consultation

In addition, users can use contextual information to supplement their background knowledge and shape an expectation of the content of retrieved documents, simply because the contextual information revealed by a concept space comes from a subset of the library - the documents selected for potential retrieval.

6.1.4 Concept Space As Semantic Exchange Medium

The synergy of applying concept space consultation over library-specific concept space goes beyond the function of eliciting user information need. Noting from its experiments and usage, the concept space consultation virtually provides a semantic exchange medium for communicating ideas between users and knowledge sources. Specifically, the exchanging information is the context about the main idea of the "discussion." However, the semantic exchange process in human computer interaction is different from that of human communication.

Contextual information is an essential element for meaningful human communications. Oftentimes, contextual information is hidden in the environment, the background knowledge of communicating parties, and the expectation to the outcome. Nonetheless, our ability to perceive context is often good. Otherwise, we will ask for the context to clarify the main concern. The omission of explicitly expressing contextual (or supportive) information has its pros and cons. The advantage is the simplicity and clarity of staying the main idea of current interest. The disadvantage is that meaningful communication may not be possible because of insufficient or inaccurate understanding of information need - what is really wanted cannot be comprehended by other parties. However, it is not the case in the human-computer communication.

When submitting a query to a knowledge source such as a search engine, a user does not expect the knowledge source to ask about the context of the query. Nevertheless, the use of contextual information is beneficial to help user elaborate information need by having the knowledge source to present its context of user's main idea. The insight of the contextual communication in human-computer communication can be described in the following four steps after a user initiate the communication by querying knowledge source with user's main idea.

1. The knowledge source first acknowledges the recognition of the main concept raised by the user.

2. The knowledge source then presents its knowledge about the context of the main concept.

3. The user is able to perceive how the knowledge source knows about user's main concept.

4. The user may

    (a) stay on the main concept and extend it with other concepts presented by the knowledge source to further explore,

(b) change focus to other suitable concepts because some associated concepts may better describe user information need, or

(c) terminate the exploration because the user does not perceive the knowledge source being able to provide needed information.

Essentially, such a human-computer interaction is a semantic exchange process, which adds the semantic dimension to the user interface to help user find relevant information.

6.2 Lessons Learned

The major lesson learned is that both concept space consultation and generation work. Strictly speaking, machine-generated knowledge sources (concept spaces) are not able to tell what semantics that they are carrying or algorithmic spreading activation process over knowledge sources conveys any semantics. Nonetheless, the computational process in concept space generation captures the knowledge of concepts and co-occurring strengths between concepts from the wisdom in the underlying library collection. Such knowledge does resemble human's knowledge about concepts and their relatively relationships. The computational process in concept space consultation makes use of such human-knowledge-resembling concept associations to further harvest from the derived human wisdom. Both processes are mechanical; however, the overall consultation result is semantical. Concept space consultation provides a semantic exchange medium to help a user engage in expressing an information need - a cognitive process. Observing related concepts found by spreading activation algorithms serves as a memory jogging tool to refine users' articulation of an information need.

From this, a second lesson learned is derived. It is not necessary to present the discovered knowledge in its entirety to users. In fact, doing so may actually scare away potential users because of the tremendous volume of knowledge. Different knowledge sources have their own knowledge representations, with which users are hardly familiar. Besides, the enormous volume of a knowledge source cannot be easily visualized on limited screen real estate. Nonetheless, users can effortlessly benefit from a "strategic" use of knowledge sources - automatic exploration and minimum amount of easy-to-understanding and helpful information.

The third lesson learned is that the automatic concept space technique is scalable conceptually and computationally. The technique works equally well for a wide spectrum of domain areas such as computer science and engineering, medical and health sciences, geosciences and law enforcement, as well as for different sizes of library ranging from hundreds to millions. The technique can be executed on a wide range of platforms including various kinds of Unix and Windows operating systems as well as configurations ranging from mobile laptop computing to high-performance parallel supercomputing.

The final lesson learned is the ability of the consultation process to give an overview of potentially retrieved documents. Such an overview is virtually a summarization method utilizing both context revealed in a concept space and a user's background knowledge. Depending on the information need, such an overview may fit the need, without its being necessary to go through the retrieval process. This lesson is a surprising finding of this dissertation.

6.3 **Future Directions**

In regards to this final finding, I believe there is a promising potential of using concept-oriented other than document-oriented (which clustering often employs) approach to performing summarization over a set of previously conceptually retrieved documents. To the least form of summarization, each document processed by the concept space technique has a ranked list of concepts (term phrases), which can be used as a quick summary for survey. To quantify the value of those *summary lists*, user studies will be needed.

This research demonstrated the feasibility of providing semantic communication between users and knowledge sources or systems. It is logical to extend this semantic communication to creation of a semantic protocol for machine communication. Agents or robots can use the concept space approach to exchange information with context, enabling individual agent or robot to identify whether it is beneficial to communicate with others of its kind, based on the exchanged contextual information. The key to achieve this goal can be inferred from what relevance means in information retrieval. Finding relevant information relies on human judgment on retrieved documents. By default, no agent is able to find relevant information from others using the familiar information retrieval search. The missing piece is the ability to evaluate the relevance of retrieved information. I believe a new search mechanism - contextual search - needs to be developed. The basic idea of contextual search is to represent information need in two categories: the main idea (focus) and contextual (supporting) information.

In this study, concepts are presented in textual format. However, concepts can be in multimedia format. Computationally, concept association works regardless of medium. After concepts can be clearly identified and extracted from multimedia documents, concepts to the process of concept association are merely references. The key of performing multimedia concept association relies on how concepts can be identified and what similarity measure is meaningful in their corresponding media. However, the semantic value that the concept association over multimedia information tries to resemble needs to be determined.

The last item worth noticing is the use of two other types of context analysis. So far, the concept space technique has used *association clusters*, which is the co-occurrence analysis. It may be fruitful to research in the use of *metric clusters* and *scalar clusters* (Baeza-Yates and Ribeiro-Neto, 1999). Metric clusters add the notion of *distance* between co-occurring pair of concepts. However, how *distance* is being measured is an open question. Scalar clusters use the co-occurring (neighboring) concepts of two targeting concepts to compute their similarity. That is, two concepts can be related without being co-occurred in any document in a library.