

CHAPTER 1

INTRODUCTION

This research investigates the feasibility of techniques in Case-based Reasoning (CBR) and metadata. These techniques allow knowledge sharing and reuse to be completed in Knowledge Management Systems (KMS).

This chapter presents an overview and general introduction to the thesis. This chapter is organized as follow. Section 1 presents a general introduction to the thesis. Section 2 identifies the research problem. Section 3 presents an overview of the research. And research objectives are presented in Section 4. Section 5 gives an overview of the proposed approach. Section 6 presents the organization of this thesis.

1.1 Introduction

Cataloging is the method used to create an index to library collection. It presents the details of the books and library resources. English and most European libraries, traditional cataloging is done according to an accepted set of rules, the most recent version of which is the Anglo-American Cataloging Rules 2 known as AACR2.

These rules govern exactly what information should be included, the order it should be included, and how each line should be punctuated. In fact, AACR2 governs the entire design of the card. AACR2 insures that cataloging is done correctly and

uniformly in all libraries. This is good, but it makes cataloging a specialty in Library Science since there are so many rules and it's complicated; good cataloging is an art. Once the item description is created, then librarians assign the correct subject headings and classify the item by assigning it a shelf location. All of this information is found on the card which is known as the main entry card. But for each item a card set needs to be created. In addition to a main entry card, a card set includes a subject card for each assigned subject, added entry cards for each added entry and a shelf list card which includes the copy information and price. Each of these cards has the appropriate identifying line on the top so it can be filed appropriately. Even if you purchase a card set, you may need to edit each card, to put the subject headings on the top and the shelf location in the corner of each card. Cataloging is a special area of professional librarianship, needing serious training and expertise to do it right.

But in the digital age, everything purchased for library are cataloged already. The correct descriptive and subject heading cataloging for its online can be found at other libraries. These cataloging records that are available online can be electronically downloaded directly into an automated system. Library suppliers will provide the cataloging, electronically too. The only thing librarians will definitely have to catalog from scratch is material that is produced in house such as videos of lectures, or programs. With an electronic catalog, librarian enter each piece of information once with an accompanying identification code, and then the machine manipulates the information allowing to search by all the traditional methods, author, title and subject and also by a new search method called keyword searching which searches every word in the record including subtitles and contents notes.

Unlike English and most European libraries, Thai library cataloging depends on manual indexing. There is no automated tool to help Thai librarians in Thai cataloging because the difficulty of Thai language characteristic. The processes to catalog Thai resources in Thai library are heavy work for Thai librarian because Thai librarians have limitation in discourse analysis for different domain ontology. The key idea to solve this problem is to transform old system to the new system. The old system is manually cataloging Thai resources by Thai librarians without Thai domain expert involvement. The new system is automated cataloging Thai resources by computer system with expert librarian and domain expert involvement via the internet. The famous method used to build knowledge sharing and reuse and involvement in organization is Knowledge Management.

Knowledge Management (KM) is an emerging discipline that focuses on efforts leading to the rational allocation of organizational knowledge assets. Typical KM solutions are described in terms of a knowledge cycle that entails knowledge tasks such as capture, distribution, and reuse. Knowledge cycles are strongly correlated with the case-based reasoning (CBR) cycle, which includes retrieve, reuse, revise, and retain steps (Aamodt & Plaza, 1994). The strong association between the CBR cycle and KM's knowledge cycles justify the consistent use of CBR to guide the design of KMS. The affinity between KM and CBR goes beyond their cycles. At the research level, the KM literature recommends that effective KM solutions target people, processes, and technology. From a CBR perspective, Aamodt and Nygaard (1995) have long ago suggested that CBR research has to consider practical applications and focus on optimizing not the CBR system alone but the combination of a CBR system

and its user. This represented an important starting point for viewing CBR as an approach contributing to KM. As a result, there have been many research activities on CBR and KM. The relationship between these fields is illustrated, for instance, by a number of CBR and KM-related events. In 1999, the AAAI Workshop Exploring Synergies of Knowledge Management and Case-Based Reasoning (Aha et al., 1999) focused on requirements for the effective contribution of CBR to KM. In 2000, the AAAI Workshop on Intelligent Lessons learned Systems (Aha & Weber, 2000) targeted a broader scope but its intelligent component relied mainly on CBR. In 2001, the program committee of the traditional German Workshop on CBR decided to change the name of their well-known annual CBR event to German Workshop on Experience Management. The close ties between KM and CBR is also evidenced in books. Tautz (2000) describes how to customize experience management systems to organizational needs especially from a software engineering point of view. Bergmann (2002) represents an encompassing textbook on experience management, presenting all aspects of real-life CBR applications. Watson (2003) presents corporate memories from a CBR perspective. Dubitsky et al. (1999) refer to this as KM/CBR synergy and it explained why CBR was used successfully in many Knowledge Management System (KMS). However, it is not a one-sided relationship. KM offers much success to CBR as vice versa. In particular, KM researchers and practitioners are recognized the important of organizational issues in the success of KMS and there are many CBR practitioners who can learn from the KM community. The method to represent the relationship between KM and CBR in practical is to implement KMS which is integrated CBR techniques.

The purpose of building KMS can be seen as sharing corporate knowledge in the organization to provide support for knowledge sharing within Communities-Of-Practice (COP). However, due to downsizing and outsourcing functions of organization in the recent years, organizations have begun to lose knowledge as people leave and take the knowledge and skills with them. Moreover it is difficult to keep distributed knowledge up-to-date and to share the knowledge in the rapid changing environment. Globalization of organizations also means there is a need for knowledge to be shared in different locations. In general, successful KMS gives motivation for sharing the knowledge within an organization. The development of KMS involves the following steps: create, capture, refine, store, manage and disseminate knowledge. One of the major problems identified in developing KMS is the knowledge acquisition phase, in particular the “create” and “capture” knowledge steps identified above. This research aims to investigate the feasibility of using techniques in CBR and metadata to improve knowledge acquisition process that knowledge sharing and reuse can be achieved in KMS.

1.2 Research Problems

Two problems that are worth to mention are difficulty on Thai language acquisition process and a lack of good tool for automated document cataloging to allow knowledge sharing and reuse between Thai domain experts and Thai librarians.

The reason of difficulty on Thai language acquisition is characteristic of Thai language as sentences are written as a long series of characters without word or

sentence markers. The Thai alphabet consists of 44 consonants 32 vowels 4 tone marks and there is no capital letter. There is no change in word form or word inflection as an expression of tense, case or gender; word ordering plays an important part in determining the syntactic role of word. The same form of words in different positions contains different syntactic properties and therefore conveys different meanings. To express tense and case, additional words often are inserted to clarify the meaning. Thai grammar does not follow the extended projection principle, as found in English, where a sentence must have an overt subject. The subject can be omitted even if it is pronominal; this characteristic is referred to as null subject parameter. Thai contains relatively few headwords. Many Thai words are formed from a combination of different nouns, verbs and auxiliaries to form compound nouns.

Regarding to the problem on a lack of good tool for automated document cataloging to allow knowledge sharing and reuse between Thai domain experts and Thai librarians. Knowledge sharing within an organization is often focused on skills and facts that can be written down and taught to others. This type of knowledge is often classified as explicit knowledge. On the other side, tacit knowledge is knowledge and skill that is hard to describe such as experience or native talent in people. In general, it is difficult to code tacit knowledge. Knowledge acquisition is generally considered to be one of the most important steps in KMS development. Knowledge acquisition method is designed to construct explicit knowledge from tacit knowledge. This is often done by using manual methods such as interviewing, tracking the reasoning process, and observing documented and undocumented

knowledge. The aim of this process is to find what information or knowledge is being used and how it is being used. One of the problems identified in the development of KMS process is the difficulty in knowledge acquisition phase. Current manual methods described above can be slow, costly and erroneous. In addition, it is difficult to describe and represent knowledge in a particular domain of interest in the traditional development style. Another problem that has been identified is the difficulty in sharing and reuse knowledge in KMS. Inefficient knowledge representation also results in the knowledge stored in KMS difficult to change and update, in particular if one wishes to extend the knowledge gained through new processes or new learning. Then this research aims to investigate ways to allow knowledge to be shared and reuse in KMS.

1.3 Overview of Research

Handling knowledge within organizations has become a vital day-to-day challenge. More and more knowledge is placed in vast computer system in the form of digital documents. Digital documents can be based on individual stand-alone document files or on internal document types within the computer system. Usage of digital documents has introduced many new sharing and efficiency possibilities for spreading knowledge. Though, usage of such systems can easily limit knowledge sharing if the “correct” documents are difficult to locate. With a fast increasing collection of documents, locating the correct document becomes ever more challenging.

Metadata can be used to give each document a description presenting key properties of the document. These metadata are based on a pre-determined and standardized metadata schema which present the possible description types (elements) and the valid content of these elements. The metadata descriptions can be a part of the data foundation used for document querying and retrieval by presenting the recourse and its most central characteristics in query results. A major challenge is to create metadata descriptions due to high user knowledge requirements, timely metadata registration processes, high human costs and the ongoing challenge of ever more documents being published. These issues can be reduced or even avoided entirely by enabling computer software to generate metadata instead of or as a supplement to manual metadata actions

KMS, as a linkage tool between librarian and domain expert, is considerably weak in terms of facilitating the concept of knowledge sharing and reuse. To ensure knowledge sharing and knowledge reuse can be achieved in KMS, especially in a networked environment such as the World Wide Web (WWW), this research proposes to use CBR techniques to provide an opportunity to allow new knowledge to be updated, stored and retrieved in the KMS. The new knowledge can be generated based on past knowledge and experiences. CBR is used to resolve new problems through the process of retrieve, reuse and refine processes of past solutions. The case in CBR is a conceptualized piece of knowledge representing past experience. Case representation includes a detailed problem description and a detailed solution. In general, when new problem arises, the retrieval process identifies the case with the most similar problem description from the past cases and applies the solution or

adapts the solution to the new problem. To achieve this and to support rich knowledge representation in the distributed networked environment such as the WWW, there is a need to present a well-defined set of domain of interest in Community-Of-Practice (COP). This research also proposes the use of metadata as a feasible approach to conceptualize a set of terms in the COP. Metadata is described as “data about data”. Metadata has been widely used in Artificial Intelligence, particularly in knowledge representation. It aims to capture domain knowledge in a generic way by providing a commonly agreed understanding of a domain. In this research, metadata is used to capture common interest of knowledge in the domain. It is applied to explicitly formalize the specification of a shared conceptualization. It provides new opportunities to prevent ambiguities in knowledge representation by supporting well- agreed terms or vocabularies. We aim to apply metadata as a form of meta- knowledge to allow consistent conceptualization to be referenced in the KMS.

In summary, this research focuses on using CBR techniques to improve the knowledge acquisition process that knowledge sharing and reuse can be achieved. The concept of knowledge sharing and reuse is important to allow up-to-date new knowledge to be added or extended in KMS. The research will also investigate the application of metadata that a well-defined set of domain of interest in community can be presented in a way that it provides commonly agreed understandable terms to be shared and reused easily. The research will investigate the application software in the web-based platform that a machine-understandable knowledge representation can be implemented to capture and update new knowledge in the KMS.

1.4 Research Objectives

The objectives of this research are as follows:

To design techniques in CBR that allows knowledge sharing and reuse to be achieved.

To implement metadata that provides a mean to standardize conceptualization of knowledge and to facilitate well-defined domain knowledge in community of practice.

To propose Natural Language Processing (NLP) and Information Extraction (IE) techniques that allows knowledge sharing between librarians and domain experts.

To construct the web based KMS that allow application software to be implemented to facilitate the process of knowledge sharing and reuse through the process of representing knowledge in machine-readable and machine-understandable form.

1.5 A Brief Overview of Research Approach

A framework that integrates CBR techniques to the KMS development cycle is proposed in this research. This aim is to improve knowledge sharing and reuse. The main feature of this approach is to apply the four phases of CBR cycle in KMS.

Briefly, the approach allows knowledge that is stored to be retrieved from the knowledge database. Then the knowledge is reused, refined and retained to allow the knowledge adaptation process to occur in KMS. By refining and reviewing the knowledge in the refine phase of CBR, new knowledge can be kept updated and retained and finally stored in the knowledge database.

A prototype will be developed based on the metadata extraction system from theses which content related to sufficient economy and Thai folk wisdom. The domain of knowledge is provided by expert librarian. Knowledge is stored as cases in the form of reasoning rules. The proposed architecture is developed in a distributed networked environment of the WWW. Metadata is used to provide standardized structure and conceptualization of knowledge domain. The aim is to minimize inconsistency and ambiguity in knowledge representation. The prototype is developed as a web-based system. An evaluation of the prototype is carried out.

1.6 Organization of Thesis

The rest of the thesis is organized as follows. Chapter 2 presents relevant literature review related to this research. The literature review will discuss literature in document technology, background of metadata, approaches for metadata generation, Information Extraction (IE), Case-based Reasoning (CBR) and Knowledge Management System (KMS). Chapter 3 presents the methodology used in the research. Chapter 4 presents experimental results. Chapter 5 concludes the thesis.