Chapter II

LITERATURE REVIEW

This study is mainly focused on using ontology to organize and develop Thai herbal medicine knowledge terminology. Ontology method helps to define the terms, concepts and hierarchical relationships among them. Ontology and it application are described below.

1 Definitions of ontology

2 General ontology development processes

3 An ontology-based health information system design

4 Ontology evaluation approaches

5 Ontology development tool

6 Review of related literatures

7 Database perspective and roles of domain ontologies in database design

8 Semantic search system and evaluation

1. Definitions of ontology

Ontology is the concept which is separately identified by domain users, and used in a self-contained way to communicate information . It is a knowledge model that represents a set of concepts within a domain and the relationships among these concepts (Riano, Real, Lopez-Vallerdu, Campana, & Recline, 2012). Ontology help to define a common vocabulary set for researchers who need to share information in a particular domain. It includes machine-interpretable definitions of basic concepts in the domain and relations among them (Vadivu & Hopper, 2012). Ontology is recognized as an important tool for coping with very great compound and/or multiple sources of information and concepts . In other words, ontology refers to the formalization of the knowledge in a domain. It is an extremely important tool for the organization and contextualization of knowledge, particularly in well-bound contexts such as scientific research, or within individual organizations (Brewster & O'Hara, 2007). It also provides understanding of the structure of information, and, interoperability between different applications or database systems. In some cases, ontology is used to refer simply to controlled terminologies (A. C. Yu, 2006).

In general, ontologies can be used for various purposes according to (Noy & McGuinness, 2001):

1) To share common understanding of the structure of information among people or software agents. This is one of the most common goals in developing ontologies

2) To enable the reuse of domain knowledge

3) To make explicit domain assumptions

4) To make domain knowledge from the operational knowledge

5) To analyze domain knowledge

6) To do ontology-based search system

According to (A. C. Yu, 2006) biomedical ontology is useful for:

Terminology management

Because of the complexity of medical or biological information, it is difficult to manage a healthcare information system for effective communication among healthcare professionals. It requires deep analysis and formal representation of the meaning of terms. Adoption of the ontology approach helps to facilitate this task and assures that information will be in a form which is usable by computers.

Integration, interoperability and sharing

In biomedicine, these are important for purposes of the facilitation continuity of healthcare; in biological research, it facilitates the sharing of experimental data among researchers.

Knowledge reuse and decision support

Ontology is an application independent, which helps to create knowledge bases and can be reused in new systems without additional system development

Basic building components of ontology include: concepts or classes, properties of each concept describing various features and attributes of concepts. Details of all these ontology components are presented as below.

- *Concepts/classes* can refer to what is generally in reality or what we are interested in
- *Properties* are details to describe each concept
- *Relationships* explain how two terms or more are related. If the relation describes a relationship among two terms, it is called a slot. If the relationship describes more than two, it is called function.
- *Axioms* are logical statements that assumed to be true without proof
- *Instances* represent specific entities from the domain knowledge base

Example of ontology component in Fig. 1

- Concept is Car, Price and Brand
- Properties of Price is a number represent price of cars and Brand is a name of car's brand

- Relationship of Car to Price and Brand are hasPrice and hasBrand respectively

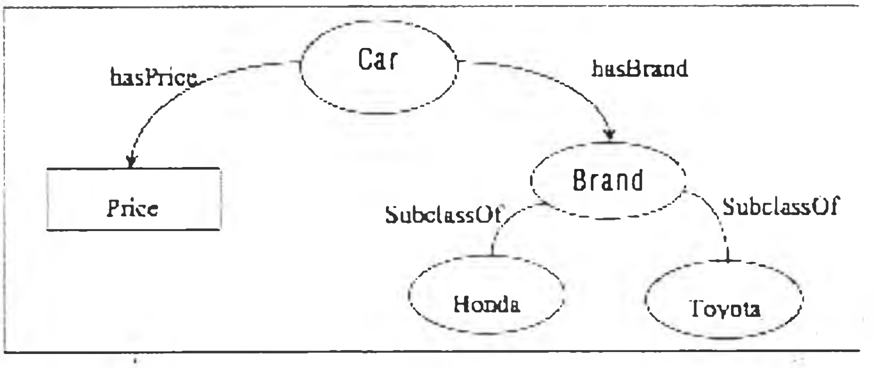

Figure 1Example of ontology components (Exerpted from Prakittikornchai, 2007)

2. Ontology development processes

There are various methodologies for developing an ontology. The most common methodology is described in defining classes/concepts in ontology, arranging the class/concepts in taxonomic hierarchy, defining slots an describing allowed values for these slots and filling in the value for slots of instances. Details are presented as below:

2.1 Determine the domain and scope of the ontology

Ontology developments start by determining scope by answering several basic questions:

- What is the domain that the ontology covers?
- For what purpose are we going to use the ontology?
- For what type of questions should the information in the ontology provide answers?
- Who will use and maintain the ontology?

All these questions will help to limit the scope of the model ontology and answers can be changed during the development process.

2.2 Consider reusing existing ontologies

This is a step designed to explore and check whether someone has done the same kind of domain knowledge. This might be necessary for the system to interact with other applications that have already been committed to particular ontologies or controlled vocabularies. Ontology is in electronic form and can be import easily into the system users need.

2.3 Enumerate important terms in the ontology

This step list related terms which will be used in the ontology, what are the properties of all terms are or what terms need to be explained to users. These depend on the scope to the ontology. Comprehensive list of terms is very important in order to represent all domain knowledge and their relations.

2.4 Define the concepts and the concepts hierarchy

This is the step in which to define terms, concepts, properties, axioms and instances. There are several possible approaches.

A top-down development process starts with the definition of the most general concepts in the domain and subsequent specialization of the concepts.

A bottom-up development process starts with the definition of the most specific classes, the leaves of the hierarchy, with subsequent grouping of these classes into more general concepts

A combination development process is a combination of the top-down and bottom-up approaches. It starts with defining more salient concepts first and then generalizes and specializes them appropriately.

2.5 Define the properties of concepts

To provide more information explaining each concept, the defining properties of concepts is necessary. In step 2.3, the concepts have been chosen. This step is the step to determine what is the properties of each concept are.

3. An ontology-base health information system design

Another ontology development process, *A Four Stage Approach for Ontology-based Health Information System Design* (Kuziemsky & Lau, 2010) describes and illustrates methodological stages to capture user knowledge in a biomedicine area. This methodology will be applied in this study because the methodology is more specifically suited to the ontology development for Thai herbal medicine which is concluded as one of the biomedicine area. Indeed, the method support ontology based Health Information System (HIS) development which can be conform with the ontology-based search engine in this study. The four stages as shown in Fig.2

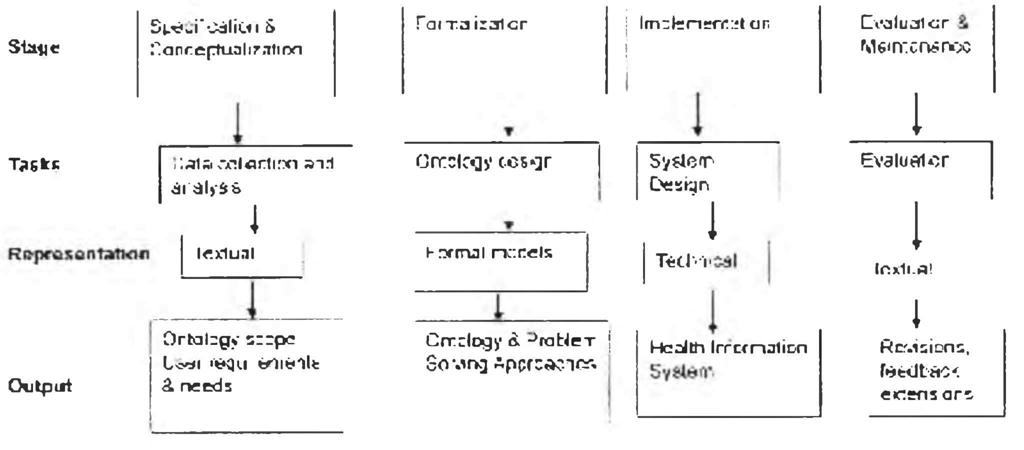

Figure 2 Four stage approach for ontology base HIS design

Stage 1-specification and conceptualization

Specification identifies the purpose and scope of ontology while conceptualization provides the concepts, vocabulary and relationships for ontology design. There are two tasks at this stage: Data collection and data analysis

Data collection

The specific data sources used for a project will vary according to the study context. To identify ontology scope and purpose there are three goals the data sources need to accomplish.

- To be able to validate concepts and processes

- To incorporate conceptual models and other relevant research literature as a meal of liking research and practice

- The use of historical data such as retrospective patient cases, to understand current data collection practices and how that data formalized into information and knowledge to be returned to end users.

speaks of a method called Participatory Design (PD). PD helps to design a product and ensures the usability and utility of the product by engaging end users in the design process. It is the way to acquire understanding and to make sense of the traditional, tacit and often visible ways people perform their everyday activities. In the HIS, it was used to involve users in information system design: interface, data entry and retrieval tools. The main strength of PD is that it provides a means of user engagement to obtain a rich perspective on clinical practice. It enables an accurate domain knowledge representation in ontology development.

Data analysis

In the literature, it used Grounded theory (GT) methodology for data analysis development. The coding cycle contains three coding: open, axial and selective coding.

- Open coding involves the initial examination, comparison, conceptualization and categorization of data to establish concepts and categories

- Axial coding refines the concepts and categories and establishes connections between them

- Selective coding establishes the core categories and links the different concepts and categories to it

The strength of GT is to code data into concepts and categories to develop a theoretical based understanding about the data.

Hybrid grounded theory-participatory design method (Figure 3) was conceived to obtain the data and code it for ontology.

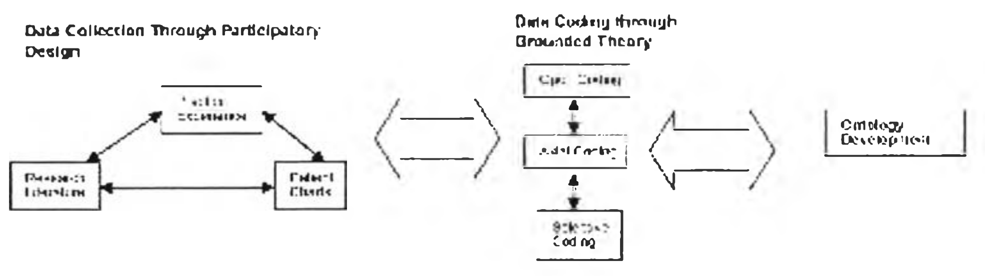

Figure 3 Participatory design and ground theory (excerpt from (Kuziemsky & Lau, 2010))

The end product of this stage is textual data that has been coded. The selective codes from GT represent the scope of ontology and become the starting point for the formalization stage.

Stage 2- Formalization

The formalization extends the selective codes by developing hierarchies and relationships such as IS_A and PART_OF relationships for the ontology. The formalized ontological models were developed using the Protégé 2000 ontology development tool (Musen, 2000) with frames as the representation language.

This stage involves the development of the domain ontology, sub-ontologies and problem-solving approaches. Domain ontology and sub-ontologies represent the structure and relationships of the ontology while problem-solving approaches provide specific solutions to domain problems.

Stage3-Implementation

Implementation of the ontology as a computer-based tool involved conversion of the ontology and problem-solving approaches into a technical architecture. The domain knowledge and sub-ontologies were implemented as Health Information System components i.e. the database schema, decision support or information retrieval tool. Microsoft Access was used to develop prototypes with the Visual Basic to program the added functionality. Tasks in the stage involve converting the domain and sub-ontologies into database tables. The ontology concepts become tables in the database tables. The relationships between the ontology concepts are implemented as relations between the various datable tables. Another task is to develop user views of the computer based tools. This consists of developing the forms, rules, reports and interfaces that enable users to interact with the system. For examples, the database relationships become the basis for access data elements and the problem-solving approaches become rules, queries, screens and the means of navigating through the database content.

An example of the implement stage is shown in figure 4. The current case sub-ontologies were developed (a) from current case of patient pain and converted into tables in the database as seen in (b). The user interface was developed by using the problem-solving approach, in this case a decision support tool for pain management. It is a computer-based tool that provides data entry and the display of results.

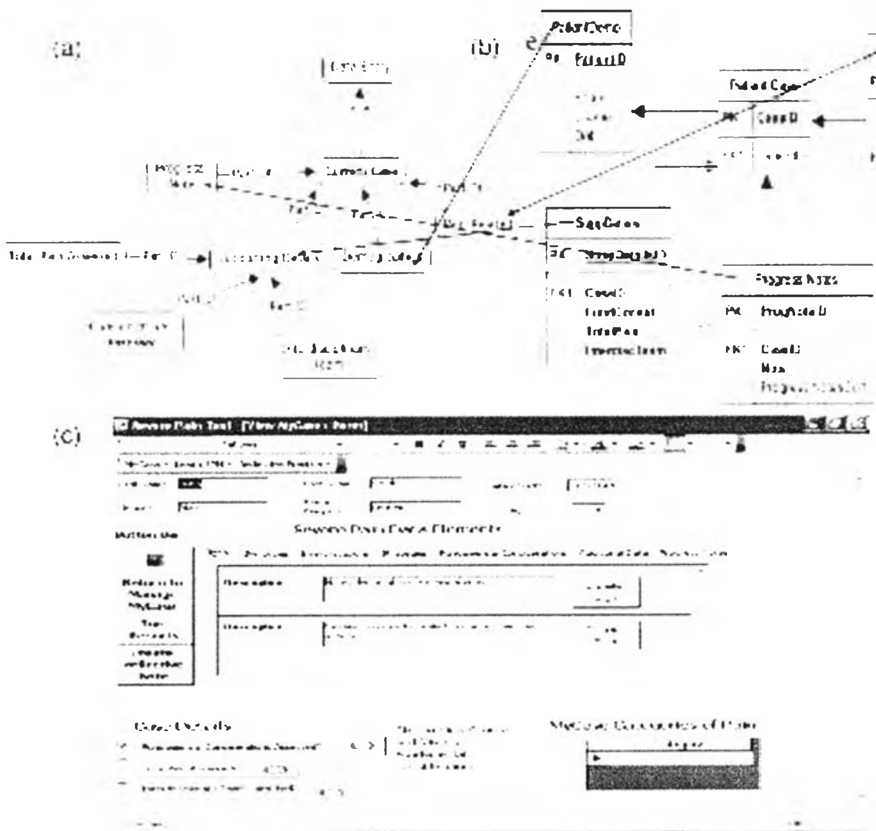

Figure 4 (a) sub-ontologies development, (b) database tables implementing the ontology, user interface from computer-based tool.(excerpt from (Kuziemsky & Lau, 2010))

Stage 4-ontology evaluation

Ontology evaluation can be divided into two methods: technical evaluation and user evaluation.

Technical evaluation

Technical evaluation involves verification and validation of the ontology concepts and vocabulary. This can be performed by using qualitative or quantitative methods or combinations of both. In the literature, the qualitative evaluation method was completed as part of establishing the GT-PD method. When derived ontology concept, GT-PD process allow users to verify ontology concepts.

User evaluation

Usability test was employed. This was not a direct evaluation of the ontology, it more likely to test of the implementation. The implementation testing could be considered as a more relevant evaluation form a HIS perspective. The reasons are clinicians do no concern on correctness or other aspects of ontological quality, but rather they want the tools which can provide value for their care delivery. The usability testing processes evaluate both usefulness and usability of the computer-based tool. The testing use both qualitative and quantitative method.

The usability use five subjects do three cases of patient presenting with severe pain. They were asked to do task and subtasks. After finished their tasks the usability were coded and quantified. The most common issues were system understandability and navigation. Usefulness of the computer-based tool was assessed by asking open-ended question following each of the three testing cases about different features of the computer-based tool. Specific questions were asked about each of the domain ontology concepts to assess the contribution each concept made to the design of the computer-based tool. The interview questions included queries on specific feature (i.e. how helpful was the ability to solve the problem?), and general questions (i.e. does the computer tool fit with your normal practice work flow?). Quantitative method was used by asking subjects to give a score on how helpful of the tool was.

4. Another Ontology Evaluation Approaches

In (J. Yu, Thom, & Tam, 2009), the study suggested three main approaches to ontology evaluation: gold standard evaluation, criteria-based evaluation and task-based evaluation.

- *Gold standard evaluation*: This approach compares ontology with ontology. It determines the accuracy of discovered relations generated form certain proposed ontology with the existing ontology. It seem not so useful outside the domain of ontology learning because if a known gold standard ontology exists then there is no reason to evaluate other ontologies.

- *Criteria-based evaluation*: This approach is an evaluation base on criteria such as consistency, completeness, conciseness, expandability and sensitivity. This kind of evaluation can be performed only by human. It is difficult to perform automate test. The criteria focus on the characteristics to the ontology in isolation from the application.

- *Task-based evaluation*: This evaluation based on competence of ontology in completing tasks. This can judge whether ontology is suitable for the application or task in a quantitative manner by measuring its performance within the context of the application.

This study will use only technical evaluation to verify and validate ontology by applying criteria-based and task-based evaluation approach. Criteria-based evaluation is a qualitative approach, using experts to evaluate ontology in criteria of correctness, consistency, completeness and conciseness. Task-based evaluation is a quantitative approach to evaluate correctness and coverage of information retrieve from ontology search system.

5. Ontology development tool

5.1 Protégé

For ontology development, there are several tools for constructing the hierarchy of domain knowledge or ontology knowledge base. One of the most popular tools is Protégé (http://protege.stanford.edu/overview/). It was developed by Standford Center for Biomedical Informatics Research, Standford University School of medicine. The functions of the tool allow a rich set of knowledge-modeling structures creation and provide actions to support the creation, visualization, and manipulation of ontologies in various representation formats. This tool supports two main ways of modeling ontologies: the Protégé-Frames and Protégé-OWL.

- The Protégé-Frames editor enables users to build and populate ontologies that are frame-based, in accordance with the Open Knowledge Base Connectivity protocol (OKBC). In this model, an ontology consists of a set of classes organized in a subsumption hierarchy to represent a domain's salient concepts, a set of slots associated to classes to describe their properties and relationships, and a set of instances of those classes-individual exemplars of the concepts that hold specific values for their properties.

- The Protégé-OWL editor enables users to build ontologies for the Web Ontology Language (OWL). "An OWL ontology may include descriptions of classes, properties and their instances. In ontology, the OWL formal semantics specifies how to derive its logical consequences, i.e. facts not literally present in the ontology, but entailed by the semantics. These entailments may be based on a single document or multiple distributed documents that have been combined using defined OWL mechanisms"

5.2 Hozo

Hozo is an ontology editing tool which has been developed by the Institute of Scientific and Industrial Research (ISRI), Osaka University (Kozaki, Sunagawa, Kitamura, & Mizoguchi, 2007). It was developed based on both of a fundamental consideration of an ontological theory and a methodology of building an ontology. The functions provide users with a graphical interface which and browse and modify an ontology locally. The features of Hozo describe as below:

- Support role of representation of ontology
- Visualization of ontologies

- Distributed development based on management of dependencies between ontologies

Hozo is not support the OWL editor as Protégé, but it allows an exporting of ontology to OWL and RDF language in order to further integration with other applications.

6. Review of Related Literatures

Ontology has become the knowledge representation medium of choice in recent years in a range of computer sciences and biomedicine. Several researches have applied ontology to their work.

6.1 Ontology on herbal or traditional medicine

In traditional medicines, traditional Chinese medicine (TCM) has become the most popular alternative medicines. Ontology has been applied to manage and organize basic terminologies and assertions representative of TCM (Gu, 2010). The study extract terms and concepts of TCM focused on five element theory, TCM disease principle, diagnosis which mainly focused on syndrome, rules of treatment and therapies. The result of the study defined 821 classes or concepts in TCM, 26 properties, 134 axioms and 78 properties of axioms in their ontology.

In India, ontology was designed and developed for mapping of Indian medicinal plants with standardized medical terms (Vadivu & Hopper, 2012). Because of the Information about medicinal plants is scattered in text form and unconstructed, search engines are not efficient and manual processing is required.

Because it was difficult to browse for medicinal uses of herbs, ontology mapping was used to organize and standardize terms and relationships.

In African traditional medicine (ATM), ontology was used to define the taxonomy of selected herbs in ATM and their botanical hierarchy, it also was used in defining relationships among herbs and treatment methods (Oladosu, Align, & Mbarika, 2012). In (Kamsu-Foguem, Diallo, & Forum, 2012), ontology was used as an important tool to extract formal model describing ATM. In this case, ATM ontologies include symptoms, medical sign, disease, activity (treatment or diagnosis) and dosing devices. This aim was to facilitate knowledge sharing and reusing.

In Thai Traditional medicine (TTM), Thai herb ontology prototyping has been develop by defining herb concepts of Thai traditional medicines (Prakittikornchai, 2007) in the form of ontology classes and properties as shown on figure 5. The scope of the study cover 27 sets of Thai herbs derived from the data provided by Royal Thai Government Ministry of Public Health's. The study developed user interface to easily find details about Thai herbs, relations of herbs to Thai Traditional medicine, and symptoms and herbs to treat symptoms. This study defined classes of Thai herb which are Thai Herb, Medicine (household remedies), part of use (part of herb which has medicinal property) and symptoms. In medicine class, it can be divided into 6 subclasses according to therapeutic properties of medicine which are blood, carminative, clod, cough, diarrhea and heart condition. In 2004, Thai medicinal plant ontology was developed to classify and provide standard vocabulary in knowledge area related to medicinal plants. The study developed an ontology-based system for querying multiple medicinal plant data sources. The ontology classes included phytochemistry, plant taxonomy, economic botany and therapeutic categories of herbs. Another application which used an ontology approach can be seen in (Nantiruj, Maneerat, Varakulsiripunth, Izumi, & Shiratori, 2008), in which an e-health advice system with Thai herbs developed with an ontology. The system provided three functions: the Thai herb search, treatment and advice functions. The first two functions were designed using the database query method while the advice function used ontology to define relationships among classes. The ontology classes in this study included: symptoms, personal information, anamnesis, provinces and Thai herbs. The system recommended appropriate herbs to treat symptoms based on users and inference rules. Another e-health system with Thai herb recommendation outputs used ontology to express domain knowledge of Thai herbs. The ontology classes displayed were person, Thai herbs, taste (of Thai herb), symptom, chronic disease and living places. The system provides recommended Thai herbs based on information received.

All of Thai herbal medicine or Thai traditional medicine literature reviewed was developed by computer science experts/engineers. While, evaluation focused only on technical or function of the ontology-based system. The scopes of the ontology developments, from literatures above are not cover all classes and properties of Thai herbal medicine domain knowledge. Diagnosis, treatment and properties of herb based on Thai Traditional Medicine has not been developed. Although, one study included taste of Thai herb as an ontology class but it was not referred to as a basic properties of herbs or appropriate for herbs for personal symptoms. It was used only for recommending herbs which had taste preferred by users. There has been no study that describes the ontology of basic Thai traditional medicine knowledge or focuses on extracting Thai traditional medicine terminology. It is for this reason that this study will be conducted.

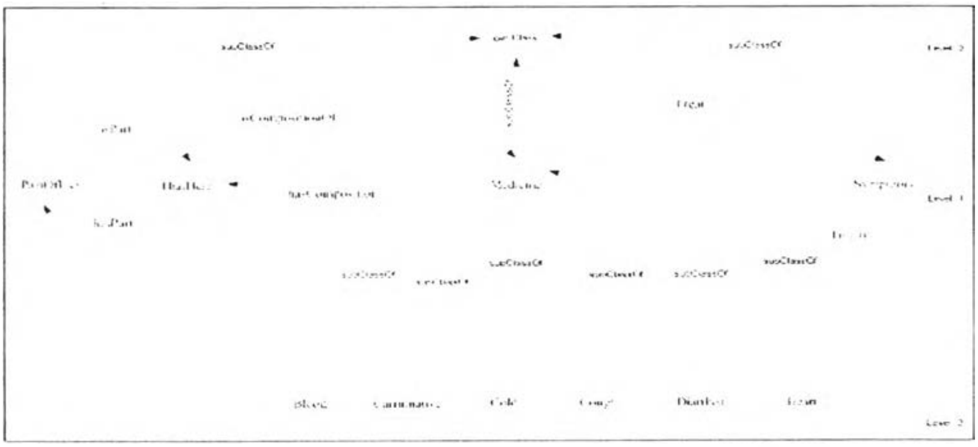

Figure 5 Example of Thai herbal medicine ontology (exempt from (Prakittikornchai, 2007))

6.2 Ontology in search engine development

Ontology can also apply in database and/or search engine development as shown in the following two reviews.

Design and Development of an Ontology Based Personal Web Search Engine (Sakthi Priya, Revathy, Pradeesh, & Rene Robin, 2012)

In this study, an ontology editing tool such as Protégé, an open source ontology editor and knowledge-based framework were used to develop a personal web search engine. This study was used to construct concepts/classes gathered from the World Knowledge Base (WKB) like Wikipedia, WordNet and other sources. Concepts/classes were converted into an ontology database. In this study, MS-Access was use as a database tool, to store all classes and subclasses of ontology.

Search engine was evaluated by using 20 questions to receive feedback in terms of accessibility, navigation, security and efficiency.

Technological Resource Search Engine Based on Ontology (Wang, Mao, Dai, & Sun, 2010)

Traditional search engines based on keyword retrieval are low in precision, because of the lack of user understanding and received huge volume of result which can lead to information confusion. To solve these problems, this study examined the technological resource search engine based on ontology. In the whole system, the core data layer was developed by using ontology techniques to identify concepts, sub-classes, instances and relations. The system established a technological resource indexing according to the ontology base and then establishes technological resource indexing database. The result showed that it helped users to get more precision results and satisfied users better than traditional search engine.

7. Database Perspective and Roles of domain ontologies in database design

7.1 The Database Perspective

Database system concerned with the storage, maintenance and retrieval of the data which is available in the system in explicit form. In database environment, each item or record is separated into several fields which contain the valued for a specific characteristic or attribute identifying the corresponding record which it is linked. The information retrieved will consist of all records or items which are exactly match with the keyword search. In this case, it is difficult to formulate precise to search keyword and the result may include items that may or may not match the information requests exactly. In ontology employed system, its objective is to achieve better procession and recall in the text retrieval system by query expansion through the use of semantically related-terms (Khan, 2000). Some ontology, it was shown to be potentially relevant to enhanced recall and get more precision result compare to general keyword search.

7.2 Role of Domain Ontologies in Database Design

The database design process generally follows five steps: Planning and analysis, conceptual design, logical design, physical design and implementation. The developer collect information about requirements by review exist information and ended-users interview. Database tends to manage only some for some purpose or part of domain knowledge. In contrast, ontology are intended to give details and explain the world, to capture all domain knowledge and provide understanding of real world because it made of concepts, instances. properties, axioms and relationships. In case of the domain knowledge has various concepts or sources of information. Ontology is advantageous for data modeling and conceptualization according to the four characters of ontology: explicit, formalization, sharing and conceptualization (Wang, 2010). Ontology can be used to help people and machine to communicate concisely by facilitating the information exchange based on semantics rather than just on systax (Calegari & Pasi, 2010).

In searching for interested information and receive more precise result for user, personalized ontology can be applied. Personalized ontology is defined by performing a mapping between the relevant concepts extracted from the visited pages and the concepts of interest extracted from search queries. In

(Calegari & Pasi, 2010), personal ontology is built by considering the relevant concepts extracted from the preferred web page after web search. the user's queries and user's preferred document. Result in extracting useful concepts defined in ontology and used to expand the user's query for better located relevant information, and ontology can be used to re-ranking the results produced by search engine according to the user interests. This methods can be applied to this study in extracting useful concepts from FAQs of TTM-IS which they can refer to ended-user prefer terms. This can improve user's search

8. Semantic search system and evaluation

An evaluation technique of ontology to determine the efficiency of semantic search system can be measured in data retrieving model as precision and coverage. An experimental approach will be apply for this stage by calculation the retrieved objects from the search system result compare to the amount of whole identified objects and relevant objects (in ontology). Correctness and coverage can be conceptualized as below.

Precision refers to the ability of ontology in capturing concepts, classes and relationships correctly according to the frame knowledge. This can be calculated from equation describe below:

$$\text{Precision} = \frac{\text{Number of relevant objects that are retrieved}}{\text{Number of retrieved objects}}$$

Coverage means all relevant concepts; classes and properties in the ontology adequately cover the concepts of the frame of knowledge.

$$\text{Coverage} = \frac{\text{Number of relevant objects that are retrieved}}{\text{Number of relevant objects}}$$

Precision and coverage ratio have range between 0-1. To measure overall precision and coverage of ontology, the frequently asked questions of TTM-IS were used to design the scenario for information retrieving in order to test the efficiency of ontology application.

Table 1 Precesion and coverage score measurement

Scale of measure	**Score**	**Description**
precision	1	ontology-based search engine provide correct result according to the reference
	0	ontology-based cannot provide correct result according to the reference
coverage	1	the result cover all relevant objects
	0	the result not cover any relevant objects