A Methodology to Develop Ontologies for Emerging Domains

Chai Meenorngwar

Author's Note

The author would like to thank the Royal Thai Government and Valaya Alongkorn Rajabhat University.

Abstract

The characteristic of complex, dynamic domains, such as an emerging domain, is that the information necessary to describe them is not fully established. Standards are not yet established for these domains, and hence they are difficult to describe and present, and methods are needed that will reflect the changes that will occur as the domains develop and mature. This research proposes the Liverpool Metadata or LiMe methodology to develop an ontology and organise the knowledge that is necessary for developing the domain environment descriptions. Its aim is to capture Knowledge Information (KI) from research articles and translate this into semantic information with web description languages such as XML(s), RDF(s), and OWL. LiMe represents an Ontological Framework, which provides the concept characteristics, represented as a concept framework that specifies conceptualisations of the knowledge. LiMe supports the Semantic Web development. "e-Learning" has been chosen as an example of an emerging domain in this research. The characteristics of e-Learning concepts will be extracted from research articles of journal websites such ScienceDirect, Springer, etc and represented as knowledge. LiMe also explicitly represents how these concepts are developed and evolve to represent the domain.

Keywords: E-learning domain; emerging domain; knowledge information; ontology; semantic Web

Introduction

Overview

The range and quantity of information available via the Internet today has created well-known problems of information overload, including difficulty of access and problems of selecting information that is appropriate and reliable. To address these problems, ways were required to categorise and organise information for access by users. The idea of using multiple sources can facilitate the reliability of knowledge, but increases the need for effective knowledge management.

A domain of knowledge can typically be seen from different perspectives. Also, information about them is diverse and possibly contradictory. Think for example of the huge mass of information contributed every day on the Internet. Therefore, methods are needed to classify and identify information to find reliable sources to construct the knowledge.

In addition, information can change and be flexible, based on time and need. For example, complex domains such as software development have a lot of platforms and standards. Knowledge or concepts in the domain have been defined or represented in different ways. Therefore, users find it difficult to choose the suitable system or concepts for their own environmental needs.

This shapes a complex and unstructured environment where unstable concepts and information are contributed all the time in a domain. The representation itself of the domain is also difficult. It needs methods to capture new concepts, organise existing concepts, and translate into well-formed information that could be shared and reused.

Objectives

The work in this research sits broadly in the field of Knowledge Management (KM). KM (Eriksson, H., 2004) is identified as the capabilities and communication that include: (1) converting individual to group-available knowledge; (2) converting data to knowledge; (3) converting text to knowledge; (4) connecting people to knowledge; (5) connecting knowledge to knowledge; (6) connecting people to people; and (7) connecting knowledge to people. It is represented as the combination of documents and ontology that can help organisations describe, store, catalogue, and retrieve information in a systematic manner.

This research introduces an approach that can help the users to classify their information and to represent it with a well-formed structure. The approach provides an ontological framework to structure one individual existing domain. This work focuses in particular on the problem of information management in an organisation. Information within an organisation needs to be accessed for different purposes. Experiences from individuals in the organisation help forming the common understanding, which could used or reused to develop new information, therefore it needs to be made shareable and reusable. In fact, individual experiences are a very important source of knowledge. For examples, researchers use the educational experiences to find the information about their experiences, governors used the working experiences to organise their daily information, and teachers collected the information from books, experiments, and so on to prepare their courses.

The aim of the research described in this research is to investigate issues involved in the representation and management of knowledge arising in an emerging domain. A number of techniques have been used for representing domain knowledge. In most cases, these methods assume the existence of a well-defined body of knowledge that can be assumed to be reliable and definitive, and needs only to be organised appropriately. In the case of emerging domains of knowledge, however, these assumptions are not valid. In this case, the "body of knowledge" is incomplete and constantly changing, and may include significant errors, inconsistencies, and instances of different assumptions, conclusions and terminology. Only when the domain reaches a state of relative maturity can these issues be resolved definitively. Meanwhile, however, there remains a need for researchers and practitioners to make use of the knowledge while it is in this state of evolution.

This research proposes a framework, Liverpool Metadata (LiMe), as the way to transform the individual experiences into relevant information for a particular domain by applying an ontology approach, structuring these experiences in terms of concepts and the relation between concepts. Concepts are defined from different perspectives under the same domain. These could be redefined, reused and described as specification of the particular domain. The development processes of the LiMe methodology is described in the following section. LiMe provides techniques to measure the relation between the concepts in the ontological framework. This allows to store and access with the other. The relation between the concepts presents as knowledge to improve the framework from new information. It is described with the well-defined descriptions such the formal language such XML(s) and RDF(s).

An ontology is a shared description of concepts and relationships in domain knowledge. It consists of terms, their definitions, axioms relating to them, and a taxonomy organizing them. The main objective of an ontology is to enable communication and knowledge sharing by capturing a shared understanding of terms that can be used by humans and programs. It has been argued for the use of knowledge representation techniques capable of reflecting the situated nature of human cognition (Gahegan, W.P.a.M., 2007). It also facilitates the sharing and reuse of information and can reduce the analysis, design, and development time of complex systems.

Within the body of knowledge to represent, a distinction can be made between information coming from referential sources and information coming from practical sources.

Referential sources use documents such as a research paper which provides reliability to the domain. Practical sources use the working experiences such tasks, activities, etc. In case study section, the approach will be exemplified by means of two case studies, one in the educational field (e-learning case study) and one in the governance field (e-inspection case study).

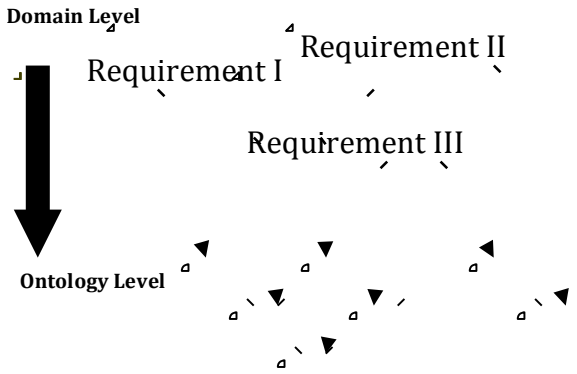

Figure 1: This description of LiMe approach.

For both the above cases, information was developed with the cyclical processes. Firstly, new concepts were defined from individual experiences and formed the structure of knowledge. This was represented as a tree of concepts. Secondly, each concept was linked with the other concepts forming relationships. Users are helped define and arrange these concepts by the LiMe environment. LiMe introduces similarity of concepts in the ontological framework and provides the user with directions for descriptions: generalisation and specification. Therefore, the users can define the appropriate descriptions for each individual information environment.

Figure 1 illustrates the main spirit of LiMe approach: users characterise their own requirements on the domain and are helped to represent them in an ontology. Organisations typically have to deal with lots of information which is unstructured and difficult to reuse and share.

Case Studies Outline

In this study, two different organisations, educational and government environment will be used to show how an ontological approach can help classify information in complex scenarios.

The e-Learning case study demonstrates the use of referential sources to capture online learning concepts from research papers to shape the domain knowledge for the Valaya Alongkon Rajabhat University. E-learning is a good example of an "emerging domain", that is a domain which has the particular additional difficulty that the current body of knowledge is not stable (O'Hara, C.B.a.K., 2007). Research into e-Learning is currently very active (M del Puerto P., 2008), and the concepts involved are constantly also being redefined and introduced in different ways. In the case study described in this research, human researchers deal with an emerging domain by a process of continuous review of published literature, from which the current consensus emerges. In the same way, published research papers will be used as the input resources of this research.

The second case study will demonstrate the use of practical sources to represent the information that is used for describing the problems in the project inspections task for the Royal Thai Government. It will later describe these case studies in detail.

Research problem

In the Internet era, people are using the information from the websites or place that they connect to. Technology provides a convenient living style. However, there are some problems for information developers in case of complex and ever changing, emerging domains, such as in the government sector or e-learning. The increasing amount of information, especially internal information such documents, projects, tasks, requests, etc, contributes to the unstructured nature of information.

The obstacle of accessing the appropriated information needs much more time and high development costs. This research will provide processes of information classification by using an ontology approach. This is the beginning of this research problem. The research question and problem is described below:

- Research question: Is it possible to organise the information of an emerging domain by using an ontology approach?
- Hypothesis I: The experiences or information from the different people could be represented with the ontology. These come from the individual person of the organisation.
- Hypothesis II: Information from an emerging domain could be used and shared the information by using the existing ontological framework.
- Hypothesis III: Semantic Web could be developed from the existing ontological framework as input.

In this research, an attempt will be made to define ontologies to facilitate environment description and represent a complex, frequently changing domain. A tool, LiMe has been implemented to capture the relevant information from a particular domain. The

objective is to transfer information and data from paper or oral communication to a representation of the knowledge in a computer system.

Literature review

In this research, an ontology development approach is proposed for capturing information and knowledge in complex domains, such as an emerging domain or a domain involving flexible information, various approaches and methods that change constantly. E-Learning systems will be used to illustrate a domain of the former kind, while a government setting will be used to illustrate a domain of the latter. In the implementation of the research, languages such as XML(s), RDF(s), and OWL are used to describe the domain environment. This chapter reviews the literatures to support the research approach. The section has the following four main sections:

- Knowledge: problems such as using knowledge in various platforms, describing knowledge with different approaches, time to develop knowledge in the organisation, etc. Knowledge development is introduced to facilitate and solve these kinds of problem.
- Ontology: ontology technology could be used to organise the knowledge.
- An e-Learning system: the e-Learning systems has introduced as domain example.
- Semantic Web: it has been used to represent the flexible knowledge information in the domain.

Knowledge and Information

Knowledge characteristic has been classified as degree of articulation and aggregation (Cooper, 2007). It is information in the context of other information, such as the relationship between data, information, knowledge and wisdom represented in it.

Knowledge is different from information when it has been used or introduced as problem solutions. The knowledge definitions are concerning on the goal of the problem. For example, the knowledge (in term of learning/teaching of online environment) is the information about the courses in the pedagogical curriculum. Knowledge is the information which solved the particular problem.

Information is derived from raw data in the events. For example, the registration data such as student information, courses registration details, are contributed when the students choose the online courses. Information could be constructed from these data such as registration table, numbers of the courses that open for selecting, instructor/teachers/allocating to the courses.

Knowledge has been defined as classification, without the classification human could be thought, action, or organisation such example of Dewey Decimal Classification (DDC) which is a method that uses in US Library of Congress classification (Wingyan, 2007). Knowledge is unstructured information provided by different sources such research papers or working experiences. The next section will present some techniques to manage knowledge.

Ontology

This research is concerned with building an ontological e-Learning requirements framework to facilitate the users or the developer to understand and use it for referencing, describing, searching, retrieving their own environment from the academic research methods or article as knowledge resources. An ontology specifies a common conceptualisation, independent of data model, and this may be presented as Semantic Web. It extracts data user contributions, and captures data as people share their knowledge in terms of classes and relations between classes. It represents existing things by illustrating and structuring the knowledge from important vocabularies. Basically, people adopt their vocabularies to the ontologies. Then, description languages such XML(s), RDF(s), OWL have been introduced to encode the structured data and tie it with common vocabularies as classes, properties, and relations with well-maintenance namespaces.

The domain will be represented as a common framework and helps to integrate or exchange data from multiple resources. The consistent knowledge of a specific domain environment is captured and combined with different information sources. Then, a reasoning approach is needed to support to interpret this framework as semantic knowledge.

In ontology, the characteristics of an interest domain have been described as concepts or entities, properties of the concepts, and relations between concepts that include the constraints (Patil, 2005). Thus, it will be used as value-mapping (support the various format or data), and scalability (depending on the context of data) (N.Huhns, K.M.a.M., 1997).

An ontology is a specification of a concept or property as knowledge (Sheng, 2004) or a concept framework (Zhang, 2006) and content management that consists of five primitives (Wang, 2006): class, relation, function, axiom, and instance. It specifies a conceptualisation of a domain in a term of concepts, attributes, relations, instances, and theories. A concept is a set of individuals or objects in a domain. An attribute is used to depict an intrinsic feature of objects. In addition, the domain scopes or objectives of the domain will be described with concepts and relations. Semantic translation determines the similarity between terms as instances of different domains and maps instances from one to other.

In practical terms, Semantic Web technology uses Ontology abilities to communicate between human and computer by providing an explicit specification for the conceptualisation of the existing domain. The classic Web will be extended with the meaning of concepts on Semantic Web which could also be shared and reused.

Next, it will explain examples of the research areas that used ontologies to describe their domain environment.

In Information Retrieval systems (Hwang, M.K., 2007), ontology is used to create, query, inference, and management information that help users to edit, delete, and modify the existing knowledge in the domain. In order to retrieve the information from the ontology, the reasoning and processing will be used in the query engine.

For example in the tourism domain (Dai, B.A.W., 2005), it is not only information such as the accommodation profiles (details, facilities, etc) that is annotated with RDF

metadata which could be retrieved but also tourism information such as water quality, places, etc could be annotated as semantic data and used for intelligent search (Sebastian Hübner, R.S., 2004).

Wingyan (2007) proposed Web directories to use ontologies to organise voluminous information into hierarchical structures, and help users to quickly locate relevant information and to support decision-making.

E-Learning

An e-Learning system is an education system that is provided in an online environment, usually via the Internet. Various related terms include virtual classroom, online learning, web-based learning, computer based learning, web instructions, etc.

The use of the Internet in education has the potential to motivate students and teachers, increase student participation and interaction in the classroom, and provide students with a more active role in their learning and increased autonomy in the educational process. While teachers are requested to use the capability of the new high technology to facilitate learning processes, students are encouraged to improve their learning through computer and networked-based activities.

For example, the Ubiquitous e-Learning (Norm Friesen, R.M., 2005) is a formal education which not only outside the classroom but also outside the education environment such as workplace, street, home.

In addition, an e-learning environment (Norm Friesen, 2005) regards teaching as a continuous process transferring knowledge with delivery in different forms such as offline and online learning; self paced and live learning; structural and unstructured learning; formal and informal learning.

With LiMe, a learning environment to be developed as e-Learning will be designed and organised, and the environment based on individual requirements. These requirements will be transformed as a common understanding framework which available to be modified by each user. The different facilities such human or knowledge experiences, technologies, learning materials, etc could be solved by using this common understanding framework.

Semantic Web

Semantic Web has been used to produce a semantic context-aware knowledge management framework that enables to integrate knowledge discovery, retrieval, and reuse (Norm Friesen, 2003).

Semantic Web technologies use smart tools to assist the system administrators to manage and control various kinds of problem. The requirements of the domain environments could be represented without misunderstanding by extracting and modelling the knowledge from the various documents and using Ontology to access and manage knowledge. Consequently, the common understanding of concepts is presented as semantic knowledge.

One of the most important aspects of the Semantic Web is searching knowledge from ontologies. Rules of representation have been designed in machine understandable

form (Nenad, 2002) facilitates to achieve the semantic information. However, it needs mechanisms and background knowledge about the domain for processing on ontologies such as updating or adopting their knowledge and reasoning strategies.

Liverpool Metadata

This chapter will illustrate the methodology to build and share ontologies for representing an Emerging Domain such as the e-Learning requirements domain, and will introduce 'LiMe' (Liverpool Metadata), as a means to facilitate the description of the Knowledge Information.

The idea of LiMe is to provide the descriptors or concepts which represented knowledge that obtained from research papers. In an emerging domain, the research papers provide the only effective knowledge resources, and using an ontology enables to describe this as knowledge information from them.

LiMe presents the knowledge specification of domain environment and provides the ability to share and reuse knowledge, providing a common understanding among different perspectives. People often give different names or definitions for the same thing, or different things can be described with the same definitions. An ontology aims to help this kind of problem.

Aim

The characteristic of an Emerging Domain (ED) is that the information necessary to describe it is not fully established, and hence it is difficult to describe and present, and needs methods that will reflect the changes which will occur as the domain develops and matures.

Knowledge/Information Ontologies provide the knowledge or descriptions that are necessary for developing the domain environment descriptions. The aim is to capture knowledge information from the research papers and convert to Web Description Language such as XML(s), RDF(s), and OWL.

It represents Knowledge/Information as Ontological framework, in which Concept Ontologies provide the concept characteristics which are represented as a concept framework that specifies conceptualisations of the knowledge. Representation Ontologies use the Semantic Web to illustrate the domain environment based on an ontological framework.

Characteristics

Methodologies used to develop an ontology have five different techniques: frames and first order logic, description logic, software engineering, and databases (Gomez-Perez, 2004).

LiMe uses the database technique and presents the domain with hierarchy of concepts as tree in the figure 2. It has been designed to store knowledge from information or paragraphs of the research articles. Both information and paragraphs are called

Knowledge/Information (KI) which is a consensual knowledge used to extract the concepts and their properties as Object Oriented modelling.

Environment will be organised and represent the characteristic of the domain. The particular environment is the subsystem or sub-organisation that represents the functions in the domain. For example, in e-Learning domain, it consists of learning, teaching, and management.

Knowledge/Information in the particular environment is used as referencing resources that defined concept, properties, and instances. This information also facilitates to define the relation between concepts. Relation is the relationship between two or more concepts. LiMe classifies the relation in two different relation categories: specification and generalisation. It also presents the semantic meaning direction. Specification is the top-down approach and generalisation is the bottom-up approach. Both approaches are used to develop trees or taxonomies that are called ontologies in the domain. A circumstance of domain uses ontologies to exchange the common understanding and give as a structure framework.

LiMe methodology organises domains as a five-level taxonomy. For example:

- Domain: e-Learning
- Environment: Learning, Teaching, Administration, Infrastructure, etc.
- Knowledge/Information: definitions or meaning, functions, Examples, etc.
- Metadata: Learning Material, Student, Teacher, Learner, Instructor, etc.
- Properties: Learning process, tasks, etc.

LiMe Resources

LiMe illustrates a hierarchy of research papers as Web resources and Knowledge for developing the Ontology of a particular domain environment. A research paper is organised with two parts: Reference Resource and Knowledge. It is introduced as Web Resources which contain reference information and knowledge. Knowledge will be classified as Information that is captured from the research paper or the individual experience which is contributed by the developers.

A research paper contributes information such as research problems, research methods, objectives, research results, and conclusion, represented using text, tables, or diagrams. This information is used as Knowledge /Information (KI) for developing an ontology. LiMe captures KI from the research papers using the individual experience and background knowledge of the (human) reader. Moreover, LiMe uses KI to extract or define the concepts that are related to the domain. A concept may be a general concept or class, a specific concept or property/instance, or a relation concept that represent the relationship between concepts, instances, or properties.

Instead of searching the knowledge based on keywords from the journal, LiMe organises knowledge that facilitates to reduce the retrieval time. The unnecessary article will not be listed. However, the appropriate concepts that facilitate to identify or describe the knowledge are important, costly, and time consuming process.

LiMe's Development Cycle

To capture the Emerging Domain (ED), flexible or new concepts are extracted from the research domain. LiMe presents these concepts knowledge as Metadata and uses

to develop ontology. LiMe proposes the development cycle (Figure 2) with four basic methods: KI identification, Concepts extraction, Ontology development, and Requirements representation.

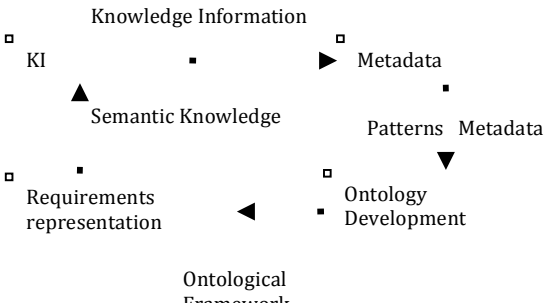

Figure 2: This figure illustrates LiMe development cycle.

The information relating to the ED is gathered from research papers. Our aim focuses on transforming the Domain specification to Semantic Knowledge.

Domain specification > Semantic Knowledge

In LiMe, domain specifications will be represented as the requirements from various researchers that contributed KI included both approaches and results in this research domain area. This knowledge will be organised with concepts that extracted from this knowledge information. Concepts also represent the patterns of knowledge which is used to classify the knowledge categories such as meanings, definitions, specifications, functions, tasks, etc.

An ontology in our research is the knowledge classification. It describes the domain specification. It translates the KI in each particular environment to ontological framework. This framework is the place for interchanging the knowledge in the environment and will be interpreted as semantic knowledge with Semantic Web.

Knowledge identification

In order to extract the knowledge from the research paper, LiMe imports the Knowledge/Information by using the academic journal search engine which the keywords to gather the domain specification from search engines of the academic journal websites such as Springer, ScienceDirect, IEEE Xplorer, etc. However, this phase does not an automated mechanism, human still have to choose and find the related papers. This process could take a lot of time especially for non-expert knowledge domain developer within huge related domain articles are listing.

LiMe describes KI as the crucial information or context information that help the users (developers, researchers, etc) to understand about the domain where could locate on paragraphs of paper articles such abstracts terms, definitions, notation, abbreviations, examples, approach, results, experiences, discussion, related topics, and so on.

In addition, LiMe also introduces the patterns of knowledge such as meaningful/definitions (descriptions), components (properties, instances), restrictions (relations, condition, constraints), etc. which could be added and improved. LiMe will store these patterns as KI categories and use them to reduce the time of capturing in the future. In order to understand, the tasks of Knowledge identification have been represented as follows:

1. Define the scope or particular environment of domain of interest, which is the objective for developing an ontology. For example, this research concerns on describing the e-Learning requirements domain, therefore, the objective is to

develop an ontology to annotate requirement in e-Learning domain, to help e-Learning researchers or organisation developers.

2. Define the keywords that related to the domain or scope, such as topics, title, instances, etc. Instead of only generate keywords from background experiences consideration, keywords could be found in the LiMe's thesaurus within the existing environment framework.
3. Use the keywords to find the related articles from the academic journal websites. With the large number of online papers, existing keyword-based searches retrieve many irrelevant papers that may use a certain word in different contexts; they might also miss papers when different words about the desired content are used.
4. Find the crucial information related to the domain, based on the previous patterns or categories. A pattern is a kind of context that identifies the relation to the scope, environment, or domain which is not easy to identify. Especially, different researchers express their knowledge in different ways. Background experiences of the domain will help to identify the knowledge context from the general information.
5. Capture the KI from articles and store it to the LiMe system. This KI will also translate to the formal language XML(s). LiMe also captures the article profiles such as title, author(s), journal, volume, issue, page, and URL. This information is a reference resource to refer during developing an ontology. Note that LiMe does not upload the file resource.
6. Update and improve the pattern identification. All the tasks are repetitive tasks. LiMe enables the developer to define the patterns which are the contextual criteria of KI.

Metadata extraction

LiMe produces sets of Metadata of the domain environment which are extracted from KI obtained from the research domain articles. LiMe presents Metadata in the term of "concept". In order to extract the concept from the KI, the follow steps are followed:

1. Find the general or specific topic such as subject or object in the statement. LiMe is concerned with capturing the definitions, components, or functions from the KI. There are various ways to find the concept in the paragraphs: find the specific concept, find the general or abstract concept, and use experiences to define the concept.
2. Given the type of concept, LiMe has four different concept types: class, properties, relations, and instances. Class is the entity or the existing things of the environment in the domain. Properties are the specification details of the concept. Relations are the relationship between concepts, which are properties of a concept. Instances are the example objects for concepts. Some concepts could be both class and properties. LiMe presents concepts as the Object Oriented model in Class, attributes and objects.
3. Define an explicit relation hierarchy between the concepts in the same KI. In addition, properties, and instances are used to specify characteristic of concept. LiMe uses taxonomies to organise concepts, properties, relations, and instances in the ontology. LiMe has relation based on type of the concepts.
4. Compare this topic with LiMe's thesauri that provide semantic between concepts such as synonym relationships. Then, update the new concept to the thesauri. A concept might take different assumptions from different perspectives and be used in different areas. In order to clarify the definition, LiMe proposes the existing

concepts with an ontological framework that could be specified the definition for creating the new concept in the thesauri.

5. Generate tree or taxonomy of the concepts to represent knowledge and also translate this taxonomy to formal description language such as XML(s), and RDF(s). Therefore knowledge is represented with one or more taxonomies from a particular KI as independent descriptions.

At this step, LiMe produces Metadata that will be used to describe the knowledge from KI. LiMe has classified Metadata based on the three basis functional types described from the statements in KI. Descriptive Metadata is a concept that describes the information such as meaning, definitions, etc of the knowledge. Structural Metadata is a concept that classifies or structures information of the knowledge. Finally, Administrative Metadata is a concept that describes information such as constraints, conditions, rules, etc of the knowledge. A set of concepts extracted from KI will be represented in this task.

Ontology development

LiMe proposes to develop an ontology for representing the Emerging domain. The development process is mainly integrating the taxonomies constructed from KI of research articles. An Ontological Framework (OF) is the result of this method. It enables developers to communicate and interact to the Emerging Domain by contributing the common understanding of concepts. This is a structure information that objective, accessible reusability, and flexible accomplishment.

In order to develop an ontology, LiMe proposes two basic processes: Similarity measurement, and Taxonomies integration as following.

Similarity measurement

LiMe uses the similarity between two concepts to reduce the redundancy and presents consistent concepts. Similarity could be easily detected by humans, whereas computers need to evaluate parameters to identify the similarity.

Currently, there are some methods that contribute to similarity algorithms such as Information-based similarity (Al-Mubaid, 2006), functional and textual based method (Ganjisaffar, 2006), and similarity graph (Andreasen, 2003). And, the relationships between concepts could be described in the terms of Synonymies, Hyponymies, and Overlapping (Maria Ruiz-Casado, 2005). Synonymies denote that two or more concepts have the same meaning. Hyponymies denote that a concept has more than one meaning. Overlapping indicate that concepts are neither synonymies nor one hyponymy of each other, but represent to some extent the same reality.

LiMe proposes to use the combinations of two techniques to compute the similarity between concepts: first, using the weight of the concepts, and secondly, using the distance between concepts.

Weight-based technique. This similarity method described in (Ganjisaffar, 2006) is based on functional and textual information. The concept similarity function calculates, from a pair of concepts, a real number between 0 and 1, expressing the degree of similarity between two concepts, based on two characteristics: Taxonomy

based concepts and weight Information Content. The "1" value indicates that a pair of concepts are strongly similar whereas "0" indicates that they are different.

Edge-based technique. This similarity technique counts the edges between concept c1, and c2. For example, (Zhumin, 2006) describes Wu and Palmer algorithm that calculated the similarity between concepts as following.

Sim(c1,c2) = 2 x / (x + 2 x)

Where and are the length of the path from c1 and c2 to their most specific common super-concept c3, and is the length of the path from c3 to the root of the hierarchy.

Taxonomy integration. In order to integrate taxonomies from a domain, LiMe focus on a similarity measurement. The relation between common concepts will be defined with "is-A" and "part-Of" relationships. The "is-A" relation is used to express that a pair of concepts have fully similar characteristics. The "part-Of" relation is used to express that there are partly similar characteristics between two concepts.

LiMe uses both a Top-down and a Bottom-up approach to integrate and express the knowledge direction in taxonomies. Top-down approach is used to annotate a more abstract concept with the specific existing concepts. This could be extracted not only from the research papers but also provided by domain experts as their background experiences.

To optimise taxonomies, LiMe uses Term matching technique patterns (Asanee, 2004) that integrate the similarities concepts to structure, form, or extend the taxonomies with four different cases.

A term matching technique is used to integrate a taxonomy which has a concept that could be expressed with relations to a different taxonomy. In additional, a consistency concept is the similarity between concepts in different taxonomies and could be expressed with the"is-a" or "part-of" relation. LiMe calls an existing taxonomy that is extended with a consistency concept as a core hierarchy.

Requirements representation. LiMe is concerned with representing Knowledge that is constructed from individual user perspectives. LiMe defines knowledge provided from the individual research article as requirements.

In order to provide a dynamic or flexible representation, a distinction can be made of the source contributing knowledge in two basic types: the reference requirements and the user-defined requirements. This will allow the users or developers to have a flexible opportunity to define their knowledge as knowledge template.

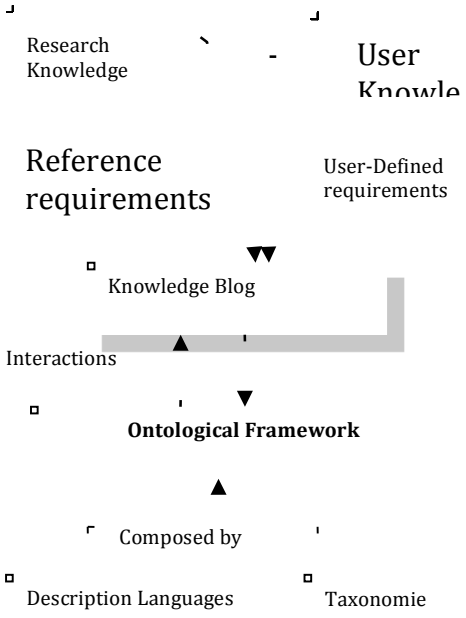

Figure 3: Knowledge Blog in LiMe.

The reference requirements are perspectives on the knowledge, the information from the research paper or articles. The user-defined requirements are contributed by the users or developer in the organisation and will be used to understand the background knowledge about the domain. It is possible that they do not have the knowledge or understanding about the domain environment.

With the various different knowledge perceptions, a flexible representation approach is required to handle the various information formats. This work proposes a semantic blog, the Knowledge Blog (KBlog), to organise and describe the different understandings of the Ontological Framework in the domain (Figure 3).

The main contribution of the Knowledge Blog is the idea of using a Blog to present the conceptual knowledge. With Blog technology, the knowledge contents are gathered from the individual requirements and research contributions on the web. The KBlog provides the interface to the knowledge of the domain as mechanism of knowledge annotation and facilitates the users to look and find across the blog comparing their knowledge with the others.

Case Studies

In order to illustrate how to use the ontological Framework, this section will present the results from two case studies: an educational and a government based case. LiMe helped produce the ontological framework which was used to develop Semantic Web solutions in each domain.

The following evaluation has been carried out, and feedback was obtained on the development. These web based applications are fully implemented and have been used in a real environment.

Teacher environment

LiMe has been introduced to support the web development for Valaya Alongkorn Rajabhat University (VRU). This university had attempted to implement an e-Learning policy but this did not work in their environment. LiMe was employed to to help improve the teaching and learning environment.

In order to do this, LiMe started from the teaching environment. It mainly supported the teachers in the grading system. In this environment, the activities of students and teachers are homework submission, class attendance, online exam, and grading.

The web developer used LiMe to find useful concepts of the teaching environment from the ontological framework and improved the framework with their teacher working experiences. This provided suitable design requirements and a clear picture before developing the software. Teachers had the opportunity to contribute their requirements.

LiMe was especially useful to reduce the time of the requirement collection. The, it was used to present the environment structure. This structure could be modified or improved to accommodate the individual requirements. The following figure (Figure 4) represents the screen snapshot of the application that used LiMe to design and organize information from the ontological framework.

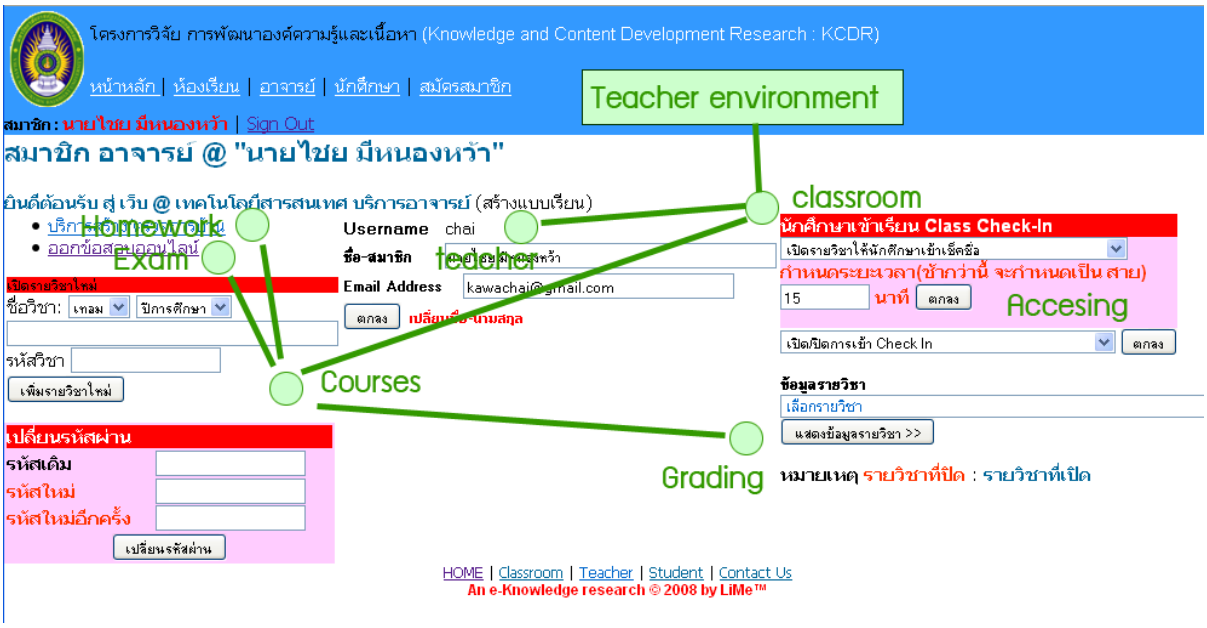

Figure 4: web-based application in the teacher ontological framework.

Faculty environment 1

In this experiment, LiMe has been used to develop the information management for the faculty environment in Thailand. It provided the ontological framework that represented information about faculty. Figure 5 shows the Faculty of Science and Technology at VRU. Faculty improved or modified this framework from their requirements. LiMe improved the budget management framework for every section in the faculty.

LiMe also provided useful features in the operation patterns. It helped define services for each member in the section as service framework. With LiMe, level of services not only is classified but also related to the relevant information from the type of member of staff using the system.

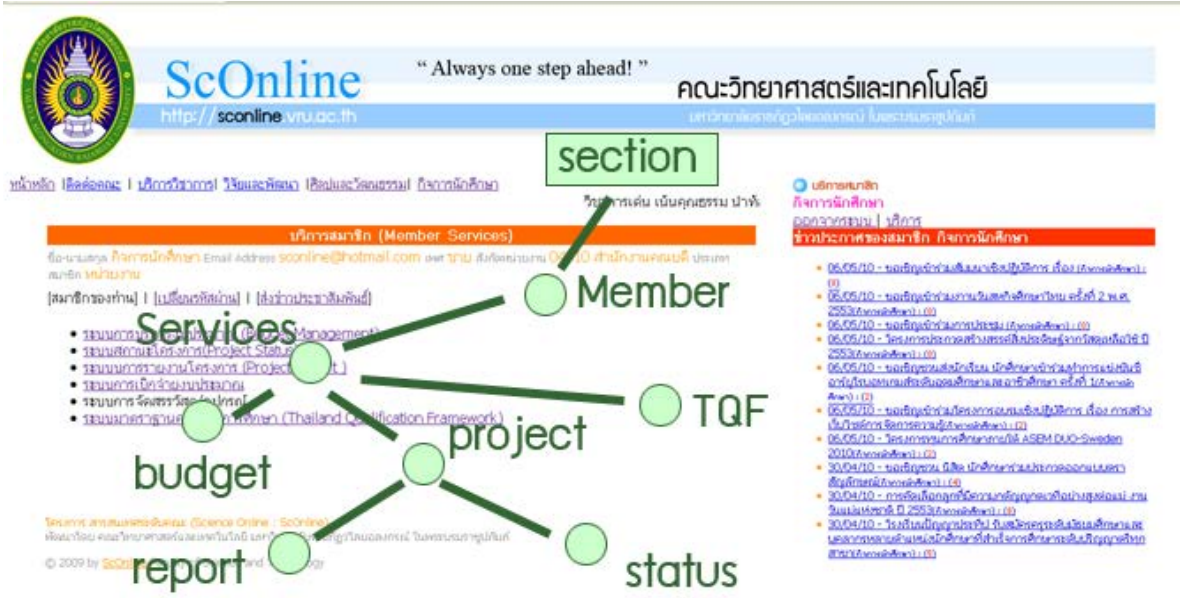

Figure 5: web-based application in the Faculty environment

In addition, the users described their projects within the ontological framework developed from the university framework. This helped establish the required interoperability between the faculties in the university. The process of implementing university strategies and monitoring project quality assurance was also improved.

Faculty environment 2

In this experiment, an ontological framework has been applied to develop and organise the information for the inspection system of the Royal Thai Government (Figure 6).

In Thailand, projects are created from the organisation of Ministries. Many projects contribute to a budget plan. Projects need to be tracked to make sure they are implemented correctly also in remote project areas such villages, and provinces. The inspection serves not only to monitor the processes but also to provide relevant information to the project owners.

LiMe has been used to collect the requirements and design the model of the government inspection. This model will be deployed in the real environment. Therefore, these requirements are very important and need a suitable structure of information to support in the inspection process.

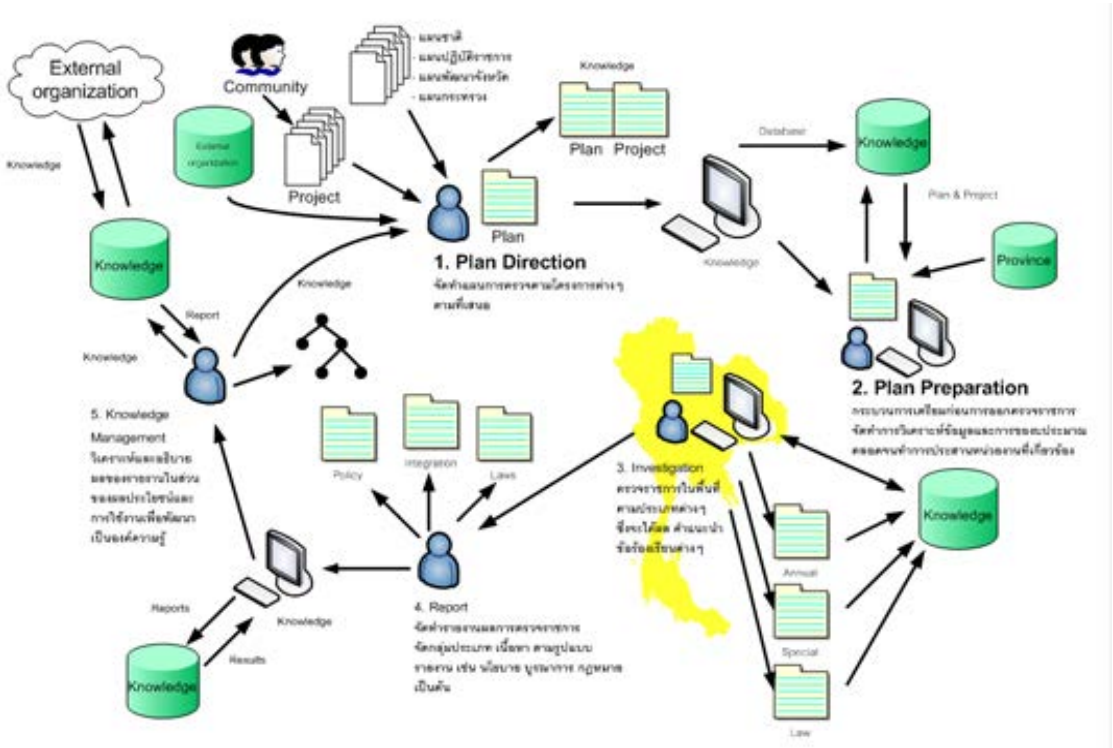

Figure 6: The inspection in Thailand.

Practically, the government inspection has been designed in five processes: plan direction, plan preparation, investigation, report, and knowledge management. An ontological framework was developed for organising and retrieving the information need in the system. It also helped to classify the project problems collected from

various areas. The common understanding of the projects was provided to the inspectors. This information was stored as the central information sources.

The Plan Direction process is the defining process. Problems coming from previous projects will be addressed and used to find solutions or improvements. It involves document classification, risk analysis, and inspection background information. This is the useful information to support the inspectors from the remote area.

The Plan Preparation process is the plan creation. The inspectors will design the tasks, problems, and schedules required by the projects. The relevant information about the remote area, such as contact information, activities, requesting, and others, will be organized to support the task. The topics that needed further information will be developed. The most important information is the project details.

The Investigation process is the data collection process. Information has been captured from the remote areas by the government inspectors. Suggestions and solutions will be provided. These are the results of the project operations. Feedback from the projects is stored in the structured information.

The Report process is the results representation. In order to improve the projects, all information that captured from the remote areas is provided. Different perspectives of the information will be developed and also the comments or suggestions will be added in this process. LiMe applied the ontological framework in the report system. It provided the report designing for the users which allowed modifying the report templates based on the individual requirements.

Lastly, the Knowledge Management process is the core process that applies to every process. It involves information classification. Knowledge is the information used to solve the problem. This is fully supported from LiMe methodology which captured the information, extracted the concepts, developed the ontological framework, and translated it into formal languages.

Basically, the different inspectors introduced different meanings for the information. This is the feature the ontology approach was most useful with. The common understanding of the information will be useful to the environment. The accuracy of accessing the right information from the existing framework was very useful. Knowledge itself could be improved from the descriptions. Therefore, this model will be the more successful, the more members are participating to it. For this reason, the system was designed the experiment as a social network (Figure 7), where members can interact by sharing knowledge, experiences, problems, suggestions, comments, etc, not only as text but also images, and video clips.

Figure 7: This inspector application designing

Conclusions

Results

This research contributed various terms or methods to this research. For example, the term "Knowledge Information" was introduced to represent the crucial information that is extracted from research articles. This KI has been captured and described based on the individual perspectives from the researchers. It is very useful information, especially for the researchers that require the articles related to their research areas. Instead of searching the academic journal websites, they can use this information to retrieve the related information and access the articles from the journal websites.

LiMe has classified knowledge into two different kinds: User-defined requirements and Knowledge Information. The User-defined requirements are knowledge that is contributed by the ontology developer, or domain expert. Knowledge Information is knowledge that is captured from research articles. LiMe represents knowledge by developing the combination between knowledge in an Ontological Framework.

An Ontological Framework is the intermediate information that provides the specification of knowledge in the domain. It has been represented with a hierarchy of concepts which is called taxonomy. LiMe integrates the taxonomies in the domain of interest based on the knowledge topics.

In order to interact with this framework, LiMe proposes the Knowledge Blog to aid the knowledge representation. It has been developed for retrieving, describing, and analysing the knowledge from the domain.

Future Works

LiMe has been proposed as an open environment methodology which extends current methodologies. It is still in the development stage. Therefore, it is available to any developer wishing to use this methodology to develop any Ontologies in any research domain.

To perform at its best, LiMe needs a lot of information about the domain. More knowledge information will produce more Metadata to describe the domain. Therefore, LiMe needs a way to integrate Ontological Frameworks, and it could be improved by applying results from the ontology community working on Ontology merging, Ontology mapping, and Ontology alignment methodologies.

LiMe classifies the knowledge based on the individual topics. Therefore, flexible information will be represented in different ways. Similarity methods are needed to resolve the problems of inconsistency in the Ontological Framework.

In practice, LiMe uses description languages such XML(s), RDF(s), and OWL to share and reuse an Ontological Framework. The specification of these languages or versioning will enhance the reliability to describe the context and characteristics of knowledge in Ontological Framework for individual environment domains.

References

Asanee, K. (2004). Automatic Thai Ontology Construction and Maintenance System. Workshop on OntoLex LREC conference, Bisbon, Portugal.

Al-Mubaid, H.N. (2006). Using MEDLINE as Standard Corpus for Measuring Semantic Similarity in the Biomedical Domain. BioInformatics and BioEngineering, 2006. BIBE 2006. Sixth IEEE Symposium on, 315-318.

Andreasen, T.B., H.; Knappe, R. (2003). Similarity from conceptual relations. Fuzzy Information Processing Society, 2003. NAFIPS 2003. 22nd International Conference of the North American, 179-184.

Cooper, P. (2007). Knowledge management. *Anaesthesia & Intensive Care Medicine*, 8(12), 516-517.

Dai, B.A.W. (2005). Architecture for automated annotation and ontology based querying of semantic Web resources. Web Intelligence, 2005. Proceedings. The 2005 IEEE/WIC/ACM International Conference, 413-417.

Eriksson, H. (2007). The semantic-document approach to combining documents and ontologies. *International Journal of Human-Computer Studies*, 65(7), 624-639.

Gahegan, W.P.a.M. (2007). Beyond ontologies: Toward situated representations of scientific knowledge. *International Journal of Human-Computer Studies*, 65(7) 674-688.

Gomez-Perez, A.n., Fernandez-Lopez, M., and Corcho, O. (2004). Ontological engineering : with examples from the areas of knowledge management, e-commerce and the semantic Web. Advanced information and knowledge processing. 2004, London: Springer-Verlag. xii, 403.

Ganjisaffar, Y.A., H.; Neshati, M.; Jamali, M. (2006). A Similarity Measure for OWL-S Annotated Web Services. Web Intelligence, 2006. WI 2006. IEEE/WIC/ACM International Conference on, 621-624.

Hwang, M.K., Hyunjang; Baek, Sunkyoung; Hwang, Kwangsu; Kim, Pankoo (2007). The Techniques for the Ontology-Based Information Retrieval. Advanced Communication Technology, The 9th International Conference, 2, 1365-1369.

Maria Ruiz-Casado, E.A.a.P.C. (2005). Automatic Extraction of Semantic Relationships for WordNet by Means of Pattern Learning from Wikipedia. Natural Language Processing and Information Systems, ed. L.N.i.C. Science. Vol. 3513. Heidelberg: Springer Berlin 67-69.

M del Puerto Paule Ruiz, M.J.F.D. (2008). Francisco Ortín Soler and Juan Ramón Pérez Pérez, Adaptation in current e-learning systems. *Computer Standards & Interfaces*, 30(1-2), 62-70.

Mahalingam, K., and Huhns M. N. (1997). A tool for organizing Web information. *Computer*, 30(6), 80-83.

Nenad Stojanovic, L.S. (2002). Searching for the knowledge in the Semantic Web. American Association for Artificial Intelligence.

Norm Friesen, R.M. (2005). CanCore: best practices for learning object metadata in ubiquitous computing environments. Pervasive Computing and Communications Workshops, 2005. PerCom 2005 Workshops. Third IEEE International Conference, 317-321.

Brewster, C. A., and O'Hara, K. (2007). Knowledge representation with Ontologies: Present challenges - Future possibilities. *International journal of human computer studies*, 5(7), 563-568.

Patil, L. (2005). Ontology-based exchange of product data semantics. Automation Science and Engineering, IEEE Transactions on [see also Robotics and Automation, IEEE Transactions], 2(3), 213-225.

Sheng, H. (2004). Ontology mapping for interaction in agent society. Services Computing. (SCC 2004). Proceedings. 2004 IEEE International Conference on, 619-622.

Hübner, S. R.S., Visser, U., and Vögele, T. J. (2004). Ontology-based search for interactive digital maps. *Intelligent Systems*, IEEE [see also IEEE Intelligent Systems and Their Applications], 19(3), 80-86.

Chung, W., Bonillas, A., Lai, G., Xi, W., and Chen, H. (2008). Organizing domain-specific information on the Web: An experiment on the Spanish business Web directory. *International Journal of Human-Computer Studies*, 66(2), 51-66.

Liu, X. J., Wang, Y.L., and Wang, J. (2006). Towards a Semi-Automatic Ontology Mapping - An Approach Using Instance Based Learning and Logic Relation Mining. Artificial Intelligence, 2006. MICAI '06. Fifth Mexican International Conference, 269-280.

Huang, W. (2003). Contextual knowledge representation, retrieval and interpretation in multimedia e-learning. Information Reuse and Integration, 2003. IRI 2003. IEEE International Conference, 258-265.

Zhang, P.Q.S. (2006). Ontology Mapping Approach Based on Concept Partial Relation. Intelligent Control and Automation, 2006. WCICA 2006. The Sixth World Congress on, 1, 4108-4112.

Zhumin, S.L.M.J.L.L.C. (2006). Fuzzy Similarity from Conceptual Relations. Services Computing, 2006. APSCC '06. IEEE Asia-Pacific Conference on, 3-10.