Ontology-based Semantic Search System for Drought Management

Nattapong Kaewboonma† Kulthida Tuamsuk‡*
Wanida Kanarkard* Marut Buranarach**
†Ph.D. Student, Doctor of Philosophy program in Information Studies, Khon Kaen University
‡Information and Communication Management Program, Khon Kaen University
*Department of Computer Engineering, Khon Kaen University
**National Electronics and Computer Technology Center (NECTEC)
kaewboonma@gmail.com, {kultua,wanida}@kku.ac.th, marut.buranarach@nectec.or.th

Abstract
This paper describes the research project on the development of an ontology-based semantic search system for drought management of the Chi River Basin (CRB), Thailand. The research methods included Ontology Development 101 by adopting Noy & McGuinness (2001) and Ontology-based Application Management framework by adopting Buranarach et al. (2012). Seventeen experts including ten experts in the areas of environmental engineering, water resources engineering, and GIS, and seven government officers who has been involving with water management in the CRB were interviewed. The experts also took part in the processes of developing the knowledge domain, classifying, and structuring the knowledge for water resources management in the CRB.

This paper is mainly focused on the ontology development process for Drought management information which comprised of three types of ontologies: 1) River Basin Ontology 2) Statistic Ontology, and 3) Task Ontology. Approximately 154 concepts of drought management have been defined and classified. The ontology was constructed by using the Hozo Ontology Editor. An example of semantic search system developed by using the drought management ontology is shown in the paper. The result will be useful for the ontology-based semantic search and knowledge visualization system applications, which will provide extensive capabilities for accessibility and decision-making on drought management of the Chi River Basin.

Keywords: Drought management, Drought ontology, Semantic search system, Chi River Basin

1 Introduction
Floods and droughts have been increasingly occurred and caused real problems in Thailand. Heavy rainfall, limited rainwater and natural runoff, ineffective use of water in the agricultural sector and large areas of degraded forest are the main causes of floods or droughts. Droughts also occur annually, and can cause heavy damage in agriculture and industrial sectors. Rainfall, water runoff, and water availability are not uniformly distributed, so drought can occur in many different places within a basin, even if there is a good balance between water demand and supply. In agricultural areas, drought can cause serious damage to local production and to farmers' dwellings. During 1990 to 2013 a severe drought was experienced in the northeastern area; mainly centered in the two larger river basins, Chi and Mun. (The International Disaster Database, 2012; The World Bank, 2011).

Considering the scope of water resources research, data are available in many different sources that use different nomenclature, storage technologies, interfaces and even languages, which make its discovery a hard and time-consuming task (Beran and Piasecki, 2009). It is essential to have effective tools to access these sources and retrieve the data precisely and accurately.

Web search is a key technology of the Web, since it is the primary way to access content on the Web. Current standard Web search is essentially based on a combination of textual keyword search with an importance ranking of the documents depending on the link structure of the Web. For this reason, it has many limitations, and there are a plethora of research activities towards more intelligent forms of search on the
Web, called semantic search on the Web, or also Semantic Web search (Hendler, 2010; Baeza-Yates and Raghavan, 2010).

The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation. It is the idea of having data on the web defined and linked in a way that it can be used for more effective discovery, automation, integration, and reuse across various applications. The semantic search engine deals with many knowledge domains which cover most of current world knowledge (Guha et al., 2003; Horvath et al., 2007). Many of the proposed solutions take advantage of the ontology. Ontology is a formal representation of knowledge as a set of concepts within a domain, and the relationships between those concepts. Ontology is becoming popular in various fields. In this sense, semantic search can provide the means to simplify the search on any data, allowing the users to access rich information with less effort, and lower the access barriers resulting from the complex interfaces of current systems.

In order to collect all the needed water resources information, researchers spend a lot of time with information search. Researchers usually simply browse the data sources, or use data search options available by the government agencies. These conventional systems for drought management information searching, however, do not offer naturally formulated search queries with direct graphical output.

In this paper, we introduce a prototype system of an ontology-based semantic search for drought management. The ontology is used as the input with a structural languages query, then searching across the heterogeneous data sources from the government bodies responsible for providing water resources information in Thailand, and displaying the results in a way that suits best the user tasks.

2 Related Research

There have been several researches on water resources information management which relating to ontology and semantic search. Raquel et al. (2010) proposed a new approach, based on semantic techniques that is able to classify the hits provided by a traditional search engine into categories according to the different meanings of the input keywords. This method considered the knowledge provided by ontologies available on the web in order to dynamically define the possible categories. The results have proved the effectiveness of the technique, especially when the users were searching for information which were not the most popular on the web.

Beran and Piasecki (2009) developed an ontology-aided search engine namely Hydroseek that allowed the users to find and retrieve large numbers of hydrologic parameters and their values from multiple databases using keywords. This mechanism significantly reduced the time spent to discover and download water quality, meteorology and hydrologic domains, that provided a linkage between scientific or everyday language and the language and variable codes used by repositories to store data.

Battista et al. (2007) proposed a modular architecture and basic functionality of Semantic web information Management with automated Reasoning Tool (SMART). The architecture was composed of an application mediator, a web-based user interface, an ontology repository, ontology index, and a hybrid reasoner which optimized the reasoning process by converting a DL expression in to SPARQL, and a SPARQL query engine for querying a database/triple store with a use case scenario. SMART can intuitively query the yeast knowledge base and navigate the results for the biological scientist.

Chen et al. (2007) also developed a prototype system consisting of software modules that belonged to the analysis and visualization and integration components of a eb interface, a data integration, a data warehouse/geodatabase, and an analysis and visualization components. This prototype system allowed the users to the integrated water monitoring data across many federal, state, and local government organizations and provided novel techniques for information discovery.

Rocha et al. (2004) proposed a hybrid search engine which combined the traditional search engine techniques together with the ontology based information retrieval and proposed a spread activation algorithm over semantic web in three different measures: cluster, specificity and combined which useful in developing the system.

The latest project being developed by the Linking Open Data (linked-data.org) is the DBpedia Project. This project consists of a large knowledge base containing structured information gathered from Wikipedia (http://en.wikipedia.org). The idea behinds the project is that people will expose their data and interlink it with the RDF from DBpedia. Such a
large scale repository provides an ideal starting point for implementing a semantic search engine and such a system has already been produced (http://dbpedia.org/search). However, DBpedia is by no means an exhaustive reference for every concept or entity. The project is also in the initial stages and not everyone can be expected to provide data that has been linked to their own. Nonetheless the SPARQL endpoint of DBpedia provides access to a rich set of knowledge that is being used in our system (Jaffri, 2007).

3 System Architecture

Figure 1 shows architecture of semantic search system. It is built for manipulating the information of water data and cooperated with drought cycle management process for integrating water dataset to enable the users to retrieve water information and statistics. The main objective of this system is to support users to monitor the water data of the CRB. Knowledge base in semantic search system is represented in the model instances that mapped water database to Drought ontology. It was built based on document analysis including the Integrated Water Resources Management (IWRM) guidelines at river basin level and suggestion of the domain experts. The system can provide the suitable information based on the search conditions.

4 Ontology Development

In this section, ontology constructing techniques will be explained. As Noy & McGuinness have said ‘there is no correct way or methodology for developing ontologies and ontology development is necessarily an iterative process’. Ontology development should base on the objectives of each application, characteristics of data, roles and context of the information system, and the feasibility of reuse in the future. In this regards, the ontology development process of this research has been designed and divided into two phases: 1) Knowledge acquisition and ontology modeling, and 2) Construction of ontology (Figure 2)

Figure 2 The Ontology Development Process

Phase 1: Knowledge acquisition and ontology modeling

In the first step, the knowledge acquisition approach by Liou (1990) was adopted for acquiring knowledge from the domain experts. The study was comprised of four steps: 1) Document analysis to identify the concept and domain knowledge. 2) Drafting the knowledge classification by using the concept of categorization, considering both common and different attributes by evaluating the data types and the characteristics of the data with a hierarchical clustering of water resources management information. 3) Interviews with ten experts who are researchers in the areas of environmental engineering, water resources engineering, and GIS in the CRB from four universities located on CRB in the northeast of Thailand, and also interviews with seven government officers who are working with water management in the CRB. 4) Summarizing the knowledge domain and reconfirming the results from the experts.

Concerning the difference of geography and climate in each region of Thailand, the area-based problem solving solution should be applied. Thus, it needs the specific information of each area so that the collection of information is enough for further analysis and processing.

Many government agencies take crucial roles in solving the issue of water resources management. The roles involves incident prevention and
preparation, aids during incidents, and recovery. By analyzing all information, it was found that there are three types of information needed to manage flooding, drought and wastewater. These three types are 1) general information about river basin 2) statistics collected by government agencies and 3) disaster management processes.

Given the definition of ontology, in this section, the type of ontology used is decided by the modeling aim. For example, if the ontology is intended to provide the record structure of databases for application fields, it is best to choose the Information ontology. If the ontology will link verbs with the management process or problem solving structure concepts, possibly the task ontology is needed.

In this research, the ontology is divided into three groups, namely, the River Basin Ontology, Statistic Ontology and Task Ontology. 1) The River Basin Domain Ontology represents the key concepts of the domain. 2) The Statistic Ontology shows a drought statistic and attributes of Weather, Water demand, Water supply, Water quality, Water quantity, Land use and Government Provincial Resources Support. It can be observed that there exist various unit of data, namely, functional relations, structural relation, and so on. 3) Task Ontology illustrates the top-level concepts and relationships of a drought cycle management process (Kaewboonma et al., 2013).

The River Basin Ontology

The river basin ontology was developed by following the top-down approach for analysis and study of relevant information sources about water resources management domain. The emphasis was on the ontology components (concepts, attributes, relation and instances). The River Basin Ontology represented the key concepts of the domain (River Basin and Sub-River Basin, Dam and Reservoir, GIS, Land use, etc.).

The Statistic Ontology

The statistic ontology was developed by following a bottom-up approach for outputting a hierarchy of concepts from the feature instances contained in the government bodies repositories which were used as data sources. The statistic ontology described the datasets (e.g. Weather, Water demand, Water supply, Water quality, Water quantity, Land use and Government Provincial Resources Support, etc.) that are needed for drought management.

The Task Ontology

Verbs were defined for task ontology by following a top-down approach. The is-a relation and part-of relation were applied for defining the class hierarchy.

Phase 2: Construction of ontology

Ontology construction in this research was done with a team of ontology engineers and domain experts, who were invited to help verify and provide additional comments on the drafted ontologies. The construction of ontology was done in the following steps.

1) Setting the Scope

In this step, concepts refer to the top-level domain classes, whereas an instance is either a bottom-level domain class or a specific object that could be derived from one concept with various properties. The ontology resulted from top-level approach is a complete model over the domain of interest. On the other side, the resulting ontologies from a bottom-up approach focus on the specification of the data sources. It captures the relationship between the data sources, in terms of the inconsistencies, overlapping, disjointedness between classes in the ontology, equivalence and so on. We decided to use bottom-up approaches from investigating water information databases, such as water level, water quality and weather information and to use top-level approach for analysis and study about the key concepts of the water resources management domain and also emphasizes the drought management.

2) Enumerate Terms and Defining the Class and Class Hierarchy

At this step, the crucial concepts and relationships in the drought management field should be concentrated and specifically defined. Hozo Ontology Editor (Kozaki et al., 2002) is a tool for developing drought management ontology. Class is defined based on basic relations, is-a relation, part-of relation and attribute-of relation. After classifying all the concepts into the main classes, relationships between them should be specified. Figure For example, Planning Process consists of three sub-process (Forecast, Monitoring and Planning). In Monitoring process consists of dataset of Water Information and Weather and Climate Information. Water Level consists of dataset of Water Level and Quality of Dam, Groundwater, Reservoir and Lake and River and Stream Flow (Figure 3).

The ontology is divided into three groups, namely, the River Basin Ontology, Statistic Ontology and Task Ontology.
Currently, Drought ontology has approximately 154 concepts that have been defined.

3) Creating Instances
There are two methods in creating instances. The first one is to use model editor (instance editor) which is an engine provided by Hozo environment, especially for model instantiation. The second one is mapping the existing information to ontology, normally mapping the existing time and location in database into model. This is suitable for organization where database are provided. Here the second method which the water database and Drought management ontology are mapped to create model instances is applied. Several ontology applications apply programming interface, such as Jena and Jaster API, in this step.

5 Semantic Search Development

We adopted the Ontology-based Application Management (OAM) framework (Buranarach et al., 2012) to develop an ontology-based semantic search system to discover the useful drought information for researchers.

The OAM framework is an integrated platform that supports both RDF data publishing from databases and processing of the published data in ontology-based applications, i.e. semantic search, recommender system applications (Figure 4).

The semantic search application template provides a faceted search interface. Using the provided form-based interface, the end-user can select a class in the ontology to search for its instance data and define some search property conditions. A search allows value comparison using both string and number comparators and semantic-based comparators, i.e. IS-A. It also allows the user to customize properties in the displayed search results. The user’s faceted search condition is automatically transformed to a SPARQL query for retrieving the instance data from an RDF database.

In this section, semantic search construction and mapping techniques will be explained. In schema mapping, we defined mapping between ontology classes and database tables. For example of each class, we defined mapping between each property, i.e. either datatype property or object property, with a database column of any table with an optional join condition. The property mapping supports one-to-one, one-to-many and many-to-many relationship types in databases (Figure 5). Based on the mapping configuration, the tool can generate instances of the ontology classes as an RDF/XML file.

In vocabulary mapping, the user can choose a table column and assign each attribute value as label of an ontology class. This will allow synonymous terms to be mapped with a class in ontology that would allow semantic-based processing in applications (Figure 6).
6 Evaluation Results

The prototype user queries were applied for testing the results of our search engine. The experts liked very much the possibility to see an overview of the search response in table format. It enables them to quickly identify major data developments without cumbersome data export and table creation for each single data series. This quick and easy way to preliminary data analysis saves the time which can be used for more in-depth data analysis (e.g. data mining methods).

Retrieval efficiency described by the precision rate was used for evaluating the drought ontology. In an information retrieval context, precision and recall were measured in terms of a set of retrieval documents and a set of relevant documents. Precision is a measure of exactness or fidelity. A perfect precision score of 1.0 indicates that every result retrieved by a search was completed.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Option</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ontology search</td>
<td>has Location &gt; CRB</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>has Location &gt; CRB has Location &gt; Khon Kaen</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>has Location &gt; CRB has Location &gt; Khon Kaen has WaterVolume &gt; &gt;800</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Table 1 Query results of an ontology-search approach (Figure 7-9).

A prototype system successfully shows that applying semantic search techniques this way provides intuitive means for searching and accessing to the complex water information.

7 Conclusion

This paper presents the ontology-based semantic search system of integrated dataset for drought management which has been studied on the case of CRB in Thailand. The research findings can be concluded as follows:

1) The knowledge acquisition of this research was conducted by integrating the explicit knowledge which were information on the drought management currently used by the government offices responsible for the CRB management, and the tacit knowledge from the experts who were involving with the CRB’s research and management. Therefore, the knowledge domain developed under this research is original and can be practically used for the development of integrated information system for the CRB resources management.

2) The ontology development of this research was built for drought management information which comprised of three types of ontology: 1) River Basin ontology, 2) Statistics ontology, and 3) Task ontology. The River Basin ontology was developed by following the top-down approach, based on the available drought information resources and experts in the fields of study. This system enables the information retrieval functions of the system operated automatically, efficiently and accurately. The number of searches and the time of search can also be vastly reduced.

3) From the literature reviews, there have been several systems being used on the studies of integrated water resources management. The research found that in the Thai context, in the needs for risk management of water resources, especially for the CRB resources, there are only three processes concerned: Planning, Response, and Recovery. This finding is completely new for the Thai concerns and has contributed the new system implementation to be used in Thailand, which has not yet been developed.
Figure 7 Example of semantic search results (has Location > Chi River Basin)

(Note: Area - ขอนแก่น = Khon Kaen, DamName - อุบลรัตน์ = Ubolratana)

Figure 8 Example of semantic search results (has Location > Chi River Basin > Khon Kaen)

Figure 9 Example of semantic search results (has Location > Chi River Basin > Khon Kaen > >800)
Acknowledgment

The authors would like to thank the Graduate School, Khon Kaen University for providing financial support on the research. The gratitude is also expressed to the Language and Semantic Technology Laboratory, National Electronics and Computer Technology Center, Thailand (NECTEC) for supporting the tools for ontology development and mapping.

References