AN ONTOLOGY-BASED KNOWLEDGE AS A SERVICE FRAMEWORK: A CASE STUDY OF DEVELOPING A USER-CENTERED PORTAL FOR HOME RECOVERY

Marut Buranarach, Thepchai Supnithi and Passakon Prathombutr (NECTEC, Thailand)

Abstract With increasing information overload on the Web, searching for information requires massive user browsing and searching efforts. Moreover, Web users sometimes need to integrate a lot of information from multiple web sites to support problem solving and decision making. One of the challenges is how to transform existing information on the Web such that:

- Information from different sources can be dynamically integrated to meet with user requirements.
- Information can be autonomously filtered and selected to match with each user’s unique circumstances.

In this paper, a novel knowledge as a service framework is proposed to transform the information on existing Web semi-structured content, i.e. blogs, tags, feeds, metadata information, such that they can be conveniently integrated and searched to match with each user’s requirements. The framework adopts ontologies as a means for integrating information from multiple sources that may be different in terms of data schema and vocabulary. Moreover, ontology is used as a means for modeling user profiles, which allow the information resources to be matched with the user requirements.

One of the differences in information access techniques between the knowledge as a service framework and typical digital library (DL) systems is search personalization. While most DL systems usually rely on keyword-based search and tag-based search, this framework adopted a knowledge-based recommendation technique that relies on concept-based search. In our framework, a concept can be mapped with existing metadata or tags. Thus, a user requirement that is mapped to a concept can be matched with the resources related to the concept.

After the 2011 Thailand floods, Thai home owners increasingly research for information in support for their home recovery. To support the home owners’ problem solving in their home recovery, we adopted the knowledge service framework in developing a user-centered knowledge portal for home recovery -- REBUILD. The portal provides relevant information that are dynamically integrated from multiple sources and present them to match with each home owner’s needs. The portal allows a home owner to identify home location, types of damages and types of home materials or parts, which were affected by flood, in a user profile. The portal can subsequently provide the related recommendations and contacts of home renovation business and construction material retailers based on the home owner’s profile. Based on some preliminary evaluation results, the concept-based search has an overall improved precision over keyword-based search as well as an overall improved recall over tag-based search in searching blog content on the portal.

1. INTRODUCTION

With increasing information overload on the Web, searching for information requires massive user browsing and searching efforts. Moreover, Web users sometimes need to integrate a lot of information from multiple web sites to support problem solving and decision making. One of the challenges is how to transform

Key words: Knowledge Service, Ontology, Personalized Recommendation

1. INTRODUCTION

With increasing information overload on the Web, searching for information requires massive user browsing and searching efforts. Moreover, Web users sometimes need to integrate a lot of information from multiple web sites to support problem solving and decision making. One of the challenges is how to transform existing information on the Web such that:

- Information from different sources can be dynamically integrated to meet with user requirements.
- Information can be autonomously filtered and selected to match with each user’s unique circumstances.

In this paper, a novel knowledge as a service framework is proposed to transform the information on existing Web semi-structured content, i.e. blogs, tags, feeds, metadata information, such that they can be conveniently integrated and searched to match with each user’s requirements. The framework adopts ontologies as a means for integrating information from multiple sources that may be different in terms of data schema and vocabulary. Moreover, ontology is used as a means for modeling user profiles, which allow the information resources to be matched with the user requirements.

One of the differences in information access techniques between the knowledge as a service framework and typical digital library (DL) systems is search personalization. While most DL systems usually rely on keyword-based search and tag-based search, this framework adopted a knowledge-based recommendation technique that relies on concept-based search. In our framework, a concept can be mapped with existing metadata or tags. Thus, a user requirement that is mapped to a concept can be matched with the resources related to the concept.

After the 2011 Thailand floods, Thai home owners increasingly research for information in support for their home recovery. To support the home owners’ problem solving in their home recovery, we adopted the knowledge service framework in developing a user-centered knowledge portal for home recovery -- REBUILD. The portal provides relevant information that are dynamically integrated from multiple sources and present them to match with each home owner’s needs. The portal allows a home owner to identify home location, types of damages and types of home materials or parts, which were affected by flood, in a user profile. The portal can subsequently provide the related recommendations and contacts of home renovation business and construction material retailers based on the home owner’s profile. Based on some preliminary evaluation results, the concept-based search has an overall improved precision over keyword-based search as well as an overall improved recall over tag-based search in searching blog content on the portal.

1. BACKGROUND

2.1 KNOWLEDGE AS A SERVICE SYSTEM

The service paradigm focuses on customization, customer relationships, service focusing, marketing to individual customers and improved information processing [1]. A service is co-creation of value between the customer and the provider. Quality is a measure of value from a customer stakeholder perspective, and productivity is a measure of value from a provider stakeholder perspective [2]. Thus, one of the challenges is to increase value co-creation outcomes of customer and provider interactions. Digital Connections Scaling (DCS) studies how the digital means may increase in the value outcomes [2]. The model proposes that digitization reduces the time and the transaction cost of service co-creation, improves service quality and productivity, and ultimately enhances the utility of service to the customer and the profit of service to the provider.

A service system made up of its entities: customer, service provider, and service experience [3]. The unified services theory [4] emphasizes that, in the service model, the customer provides

1 http://technology.in.th/rebuild/
significant inputs into the production process. The customer-input involvement distinguishes the service model from the non-service model, where customer mostly involves in consuming the output. Thus, a service system is distinguished from other types of systems by the fact that the customer may be actively involved in all nine classes [5]: customer, goals, input, output, process, human enabler (as a resource in the process), physical enabler (providing a resource to the process), informatic enabler (applying knowledge to the process) and environment.

Knowledge as a Service is an innovational service which combines the concept of services innovation [6-8] and knowledge innovation [9]. Its focus is usually on how to effectively organize information resources to meet with individual user requirements especially in problem solving and decision making. In addition, knowledge service also focus on knowledge sharing and integration between systems [10]. Ontologies are typically used for representation and sharing of domain knowledge in knowledge service [9-11]. Table 1 lists some key characteristics of knowledge service system based on the service paradigm.

Table 1. some characteristics of knowledge service system

<table>
<thead>
<tr>
<th>Service System</th>
<th>Knowledge as a Service System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Customization</td>
<td>Search Personalization</td>
</tr>
<tr>
<td>Customer relationship</td>
<td>User profiling</td>
</tr>
<tr>
<td>Service focusing</td>
<td>Problem-solving focusing</td>
</tr>
<tr>
<td>Marketing to individual customers</td>
<td>Recommendation for individual users</td>
</tr>
<tr>
<td>Improved information processing</td>
<td>Improved information integration</td>
</tr>
</tbody>
</table>

2.2 SEARCH PERSONALIZATION Searching resources in digital libraries usually relies on metadata and keyword-based search. Recently social tagging and bookmarking has gained more popularity in allowing searching resources based on user-generated tags [12-13]. However, with the users increasingly overwhelmed by the amount of information resources, search personalization, i.e. personalized recommendation, is an important trend to alleviate the problem [14]. Search personalization typically uses context information to improve searching effectiveness. For example, Amazon.com has adopted a recommendation algorithm that can efficiently recommend product items based on a user’s purchase [15].

Typically, a recommendation engine involves at least one of the following types of recommendation techniques:

1. Content-based (CB) analysis [14, 16] CB techniques focus on comparing text content of the document that the user is reading or selecting with other documents in the database. Using this technique, the recommendation engine will recommend documents that have higher similarity score with the user-selected document based on text analysis. This technique has some limitations that it can not be applied to non-text information resources, e.g. multimedia resources, etc. For these resource types, CB may only be applied to text content in metadata.

2. Collaborative Filtering (CF) [14, 16] CF techniques focus on analysis of behavior and rating of other users on an information resource. Some popular measures include click history and user rating. Using this technique, the recommendation engine will recommend resources based on popularity of the resources. This technique has some limitations for resources that are in more specific subjects. These resources are typically less popular than other resources, thus they are less likely to be recommended, i.e. “sparseness problem”.

3. Knowledge Based (KB) analysis [14] KB techniques focus on analysis of user context information. Some popular used context information include user’s background, search context, preference, etc. Using this technique, the recommendation engine can use a user’s context information defined in a profile to match with resource properties, e.g. resource metadata. Ontology may be used to enable semantic-based matching of user profiles and document properties. The quality of recommendation results also depends on the level of details of profiles or resource properties. Specifically, recommendation results may be too general if profiles or resource properties do not provide sufficiently specific details.

3. AN ONTOLOGY-BASED FRAMEWORK FOR KNOWLEDGE AS A SERVICE SYSTEM

3.1 SYSTEM ARCHITECTURE Our knowledge service system was built using an ontology-based framework. The system consists of four major components as shown in Fig. 1: RSS crawler, user profile management, resource management, and recommendation modules. Each component is described along with the system process in the following sections.

Fig 1. an architecture of ontology-based knowledge as service.

3.2 DATA ACQUISITION The system focuses on collecting semi-structured data such as metadata and tags. Specifically, the system collects metadata of blog article created by some Content Management Systems (CMS) by means of RSS (Really Simple Syndication) feeds. An RSS crawler module will extract information about each blog article of some pre-defined websites such as title, link, description, created date, creators, and tags. The information is stored in the system database and used to guide searching of the articles.

Some additional content, such as reference books and materials in the domain, may be encoded and acquired into the system database. These resources must be assigned metadata to guide searching of the resources. Data from external databases may also be acquired into the system database, e.g by means of XML data sharing.
3.3 RESOURCE MANAGEMENT AND MAPPING WITH ONTOLOGY

The resource management module allows domain experts to organize and catalog the resources in the database based on metadata and ontology. The module consists of three sub-modules: ontology management, metadata management, and tag - ontology mapping. The ontology management sub-module allows the experts to manage ontology concepts. The metadata management module allows the experts to assign metadata terms to the resources. In this framework, the metadata terms will be associated with ontology concepts. The tag-ontology mapping module allows the experts to relate existing tags of the blog content with ontology concepts. The type of link between tags and concepts can be synonymous, e.g. tag1 is synonymous with concept1, or ‘is-a’ type, e.g. tag2 is a specific type of concept2 or ‘part-of’ type, e.g. tag3 is a part of concept3.

Mapping information allows different kinds of resources to be associated with related ontology concepts. Three kinds of associations are established as shown in Fig. 2: concepts-tags, concepts-metadata terms and concept-attribute values. Tags are normally assigned to blog articles. The domain experts will map each tag assigned to the blog articles with a concept in the ontology. Mapping a tag with concepts is a less time-consuming process compared to mapping an article with concepts. In mapping tags, experts do not need to read the articles unlike mapping of an article with concepts. The tag-concept mapping can have many-to-many relationships, i.e., one tag can relate to more than one concept or vice versa.

Similarly, metadata terms are normally assigned to documents or articles. The domain experts will map each metadata term with a concept in the ontology. For the collected database resources, i.e. data records, the domain experts will map some attribute values with some concepts in the ontology. For example, the value ‘air conditioner installation’ appear in the ‘expertise’ database field can be mapped with the ‘air condition’ concept in the ontology. These mapping can also have many-to-many relationships, i.e., one metadata term or attribute value can relate to more than one concept or vice versa.

Fig 2. types of mappings concepts in ontology with existing resource information.

3.4 RECOMMENDATION ENGINE

Our framework focuses only on recommendation engine that applied the KB technique. Modeling of user profile and resource properties are based on ontologies to allow for semantic-based matching of the resources. In addition, modeling profiles and resource properties based on ontology will separate domain knowledge from database. This allows the framework to be more flexible to be applied to different domains.

User requirements are expressed in a user profile as some attribute values that are related to some ontology concepts. The recommendation engine can subsequently search resources whose properties values are matched with those of the user profile. The search process is concept-based, where values in user profiles and resource properties must be mapped to ontology concepts before they are compared and matched. The rules for matching between user conditions and resource conditions can be externally defined by domain experts. One advantage of this recommendation technique is that the user can adjust values in the profile to have the recommendation results updated accordingly.

4. A CASE STUDY OF USER-CENTERED HOME RECOVERY PORTAL DEVELOPMENT

4.1 OVERVIEW

After the 2011 Thailand floods, Thai home owners increasingly research for information in support for their home recovery. Although a large number of supported information is available, the users are generally overwhelmed by information overload. To alleviate the problem, a user-centered knowledge portal for home recovery (REBUILD) was developed using the knowledge service framework. The portal aimed to provide the relevant information that are dynamically integrated from multiple sources and present them to match with each home owner’s needs.

The portal allows a home owner to identify home location, types of damages and types of home materials or parts, which were affected by flood, in a user profile. The portal can subsequently provide related recommendations in problem-solving and contacts of experts, i.e. home renovation business and construction material retailers based on the home owner’s profile. The home recovery ontology was developed in supporting the information integration and recommendation process.

4.2 ONTOLOGY DEVELOPMENT

The first development step is ontology development for knowledge management of the domain. Ontology is used as a means for integration and search of the related information resources. In addition, ontology is used for modeling user requirements in user profile for personalized recommendation. Scope of the ontology was entities related to recovery of homes affected by flooding. The ontology was developed by a team of knowledge engineers and domain experts. The ontology was developed as a light-weighted ontology which focused on building concept hierarchy and less formal in terms of concept placement in IS-A hierarchy.

There are seven main concepts in the ontology: 1) Home recovery knowledge 2) Home part 3) Home recovery process 4) Equipment 5) Knowledge level 6) Media type 7) Home owner. The total number of defined concepts is approximately 200 concepts. Fig. 3 shows a portion of the home recovery ontology.

‘Home owner’ represents user profiles which specifies all ‘Home part’ of a user that were affected by flood and current stage of ‘Home recovery process’ of the user. It also specifies the user’s home type and location. ‘Home owner’ is related with ‘Home recovery knowledge’ which represents information resources to be
recommended to the user. ‘Home recovery knowledge’ can be related to ‘Home part’, ‘Home recovery process’, ‘Knowledge level’ and ‘Media type’.

4.3 DATA COLLECTION AND METADATA MANAGEMENT The supported information was mainly collected and extracted from the Thailand Flood Digital Archive\(^2\) and the Thailand construction industry registry\(^3\). The total number of resources collected was approximately 7,000 resources of blogs and database items. The data used for describing the resources, i.e., tags, metadata terms and attribute values were extracted and presented to some domain experts. The domain experts subsequently associated each value with the related ontology concepts. Approximately 1,500 mapping entries were created by the experts. After this process, all the resources were linked to ontology concepts as shown in Fig. 4.

4.4 PERSONALIZED RECOMMENDATION In providing personalized recommendation for a user, the user profile management module generates an interactive user dialogue interface. The dialogue prompts the user with questions and choices of answers that are some defined concepts in the ontology. Once the user submits his or her answers, the user information and conditions are linked to the ontology concepts as shown in Fig. 4. The recommendation engine subsequently search the resources that are related to the selected concepts in the user profile. The search results are presented to the user as recommended resources as shown in Fig. 5.

4.5 SEARCH EVALUATION A preliminary evaluation was conducted to compare the effectiveness of the system search results, i.e. concept-based search, with conventional techniques, i.e. keyword-based search and tag-based search. The evaluation used the data from the Thailand Flood Digital Archive website, which is powered by a Wordpress CMS and contains nearly a thousand blog articles. Five samples queries (Q1-Q5) were formulated for each search technique. Each query will have one term or tag or concept to represent each user need except for Q3 where a few terms are used for keyword-based search. Only the top ten results of each search were assessed based on an assumption that they are the most likely pages to be viewed. Precision (P) and Recall (R) of each search were assessed and compared as shown in Table 2 based on the following operational definitions:

Precision (P) = Total relevant retrieved / Total retrieved

Recall (R) = Total relevant retrieved / Total relevant

Fig 3. portion of the home recovery ontology.

Fig 4. Links between resource properties and user profiles by means of ontology.

Fig 5. example of personalized recommendation results.

---

\(^2\) http://technology.in.th/thaiflood/

\(^3\) http://www.constructiondb.in.th/
Recall (R) = Total relevant retrieved/ Total relevant discovered by all techniques

Table 2. preliminary results of search evaluation

<table>
<thead>
<tr>
<th>Query</th>
<th>Keyword</th>
<th>Tag</th>
<th>Concept</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>R</td>
<td>P</td>
</tr>
<tr>
<td>Q1: Air cond</td>
<td>50.0</td>
<td>71.4</td>
<td>100.0</td>
</tr>
<tr>
<td>Q2: Water pump</td>
<td>60.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Q3: Fence/Wall/Gate</td>
<td>30.0</td>
<td>66.7</td>
<td>80.0</td>
</tr>
<tr>
<td>Q4: Wallpaper</td>
<td>100.0</td>
<td>63.6</td>
<td>80.0</td>
</tr>
<tr>
<td>Q5: Door</td>
<td>10.0</td>
<td>12.5</td>
<td>100.0</td>
</tr>
<tr>
<td>Average</td>
<td>50.0</td>
<td>62.9</td>
<td>96.0</td>
</tr>
</tbody>
</table>

The results of the tag-based and concept-based search have an overall higher precision than the keyword-based search. The low precision of the keyword search was caused by two major reasons. First, some pages contain the terms but are not relevant to the topics, for example, “... turn off all electric appliances, e.g. air condition... before turn on electric switch board...”, “...collect all garbage and put them in front of your home gate...”, etc. Second, some terms were ambiguous, e.g. the term ‘wall’ is a part of many location names in Thai or the term ‘door’ is a part of the terms water control gates and refrigerator openers, etc. in Thai. Tags and concepts are normally assigned by users and usually represent main topics of the documents, thus resulted in higher precision. In addition, a search based on tags and concepts uses word-matching technique rather than string-matching technique normally used in keyword searching in Thai, thus can eliminate some ambiguities.

In recall evaluation, concept-based search always results in a higher or equal recall comparing to tag-based search. To explain this, in our framework, a concept search is equivalent to searching with one or more tags. The concept-based search can have a higher or lower recall than the keyword-based search. When a document is moderately related with a search term, the keyword search usually results in a higher recall. Put another way, a term that is moderately related to a document is often not assigned as tags, thus will result in a lower recall for tag and concept search. However, when a term can have synonyms, the keyword-based search can result in lower recall because some terms may not be included in the query terms. In such a case, the concept-based search can result in higher recall when a concept is mapped with all the related tags.

5. SUMMARY AND DISCUSSIONS

In this paper, we present a novel knowledge as a service framework as an approach to improving information integration and recommendation to provide support for user’s problem solving. An ontology-based approach is used to support data integration by mapping ontology with existing resource information, such as tags, metadata and database values. In addition, ontology is used in modeling user profile to support personalized recommendation. In this framework, domain experts must be involved in ontology development and defining necessary mapping information. We demonstrated an adoption of the framework in developing a knowledge portal to support home owners in their home recovery after the 2011 Thailand floods. The preliminary evaluation results have demonstrated a comparable or improved overall search effectiveness over conventional search techniques. Based on web log statistics in 2012, approximately more than 1500 unique visitors have used the website in the six–month period or an average of 250 unique visitors per month.

In this framework, the resource management process requires human expert involvements in creating mapping information between existing resources with ontology. Thus, this would not be efficient when a very large number of resources are involved. Future works should combine some automatic data mapping mechanism with ontology to improve scalability of the framework. Further, the home recovery portal can be potentially extended to support other phases of disaster management including knowledge support for user’s problem solving in disaster situations, e.g. logistics, availability and locations of nearby facilities, etc.

6. ACKNOWLEDGEMENT

This work was partially supported by the Service Informatics (SI) and Service Research Innovation (SRI) research programs at National Science and Technology Development Agency (NSTDA), Thailand. The authors would like to gratefully acknowledge the contributions from the Science and Technology Knowledge Services (STKS) and Builk Asia Co., Ltd.

(Manuscript received July 10, 2012, revised October 1, 2012)

References