A Framework for Mobile Web Query Searching using Ontology Domain

Rosliza Hamzah[a], Hamidah Ibrahim[b] and Nur Izura Udzir

Department of Computer Science,
Faculty of Computer Science and Information Technology,
Universiti Putra Malaysia,
43400 UPM, Serdang, Selangor, Malaysia
E-mail : [1]ygkh@yahoo.co.uk

ABSTRACT

Web Access today has been extended from PC/Notebook users to mobile users. There is an increased performance expectation of the web search applications due to the high demand of web usage. Web sites today are only useful to users if the information is meaningful and accessible to users. This paper explains the concept of Mobile Web search capabilities using ontology domain for mobile Internet users. This research also focuses on the specific issue on how query searching works on Mobile Web using ontology can cater for the query's polysemy (one word might have several meanings) and synonymy (several terms, i.e. words or phrases, might designate the same concept) words and also proposing a framework for Query Searching using Ontology for Mobile Web.

Keywords

Query searching, Ontology, Mobile Web

1.0 INTRODUCTION

The Web has become ubiquitous in people's lives in many respects. It follows that the Web in the pocket should be even more widely used, but the reality is that the Web is not used nearly as much on mobile handsets as on fixed terminals (Daniel, 2005).

As the Internet growing up there is an increasing need to provide users with effective search facilities. An increasing number of documents on Web, cause querying for that documents become crucial. Most of the documents are developed based on Natural Language (NL) which is very hard for machine to understand the language. The main problem in accessing Web document with NL based, NL strongly uses many polysemy and synonymy words, thus making the Web becomes unstructured Web. As a result, many documents which are not relevant with the queries are being captured. For an example, if one posed the query "return all chairs", keyword-based search engines return Web sites that contain keyword "chairs", including chairs that refer to furniture and chairs that refer to people.

The present computer system that processes information, knowledge, and data including searching is based on the similarity of structures and mechanisms. As an example, data stored in database system are managed in the form of tables and provide queries and answers using a query language called SQL (Structured Query Language) in which it is necessary to configure pattern matching between keyword written in SQL and the field name in the table. Here, Semantic Web technology is introduced to overcome such technical limitations, provide intelligence to a computer in order to understand semantics, and process knowledge, information, and data and also to cater for unstructured Web pages. The main intent of the Semantic Web technology is to give machines much better access to information resources so they can be information intermediaries in support of humans (Michael, 2003).

The mobile search engines available today is mainly for Location Based System (LBS) and Point of Interests (POIs). For an example, people often want to restrict the query to the location they are in currently. They would like to find a hotel, shop or a restaurant nearby. Users also used their mobile searching for ringtones, movie, and blog. Reachable for web documents is important to people when they are on move or during outstation. Due to the problem stated above, this paper presents a framework for accessing web documents using ontology for Mobile Web. Ontology refers to the specification of a conceptualization. In general, ontology is the backbone for the Semantic Web (Bruijn, 2003). This is because ontology can manage information efficiently.

The objectives of this research are:

- To propose and design a model for query searching using ontology for Mobile Web.
- To cater for the query's *polysemy* (one word might have several meanings) and *synonymy* (several terms, i.e. words or phrases, might designate the same concept) words by using ontology domain.
- To extract the relevant URI's list of Web documents on Mobile Web.

This paper is organized as follows. Section 2.0 briefly discusses about the preliminaries study (search engines, strategies, techniques, and conventional mobile search engine architectures). Section 3.0 briefly reviews related works. Section 4.0 presents the proposed framework for Mobile Web query searching using ontology domain, attempting toward analytical solution how mobile search can pass the user's query to server, how Semantic Queries Model is generated from NL query, and then how the results is displayed at the user's mobile screen as a URIs list. Lastly, section 5.0 concludes the paper and future works.

2.0 PRELIMINARIES

2.1 Search Engines

Search engines for both the conventional Web and the Semantic Web involve the same set of high level tasks: discovering or revisiting online documents, processing users' queries, and ordering search results (Li et al., 2005). A search engine is an information retrieval system designed to help find information stored on a computer system. Search engines help to minimize the time required to find information. Without search strategies or techniques, finding the needed information can be like finding a needle in a haystack.

2.2 Web Search Strategies

Successful searching involves two key steps. First, user must have a clear understanding of how to prepare their search. Secondly, users need to know how to use the various search tools available on the Internet. For an example, search engines (e.g., AltaVista) are very different than subject directories (e.g., Yahoo). Even search engines themselves can vary greatly in size, accuracy, features, and flexibility (http://www.learnWeb skills.com/search/main.html).

2.3 Search Engine Techniques

To use search engines effectively, it is essential to apply techniques that narrow results and push the most relevant pages to the top of the results list. The techniques used for most of search engines are *Keywords*, *Boolean Operator* ("*AND*", "*OR*", and "*NOT*"), *Phrase Searching*, *Title Search*, *URL Search*, *Link Search*, *Crawler*, *Index*, *Page Rank*, and etc.

2.4 Anatomy of a Conventional Mobile Search Engine

According to the research that has been performed at Sonera MediaLab, in its basic form, a search engine consists of a crawler, a parser, an indexer, and a query engine. Refer to Figure 1. The crawler, also known as a spider, fetches Web pages which are files written in various markup languages, and gives the pages to the parser. The right parser module parses the markup, separates URSs to be fetched later, and forwarded to the text indexer. The indexer saves the words and metadata of the page in the index database. The query engine receives search queries from the user, queries the index, and sends the search results back to the user (Sonera MediaLab, 2002).

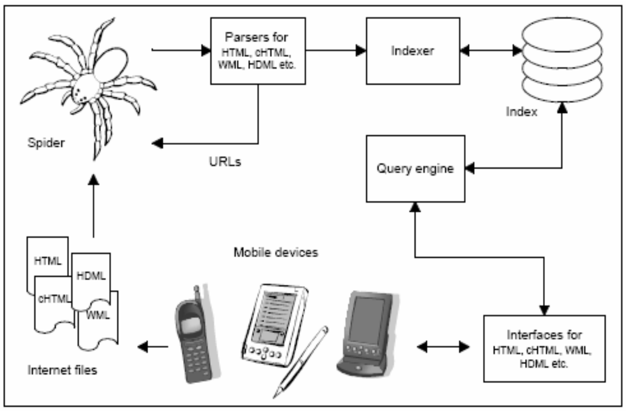

Figure 1: Conventional search engine architecture

2.5 Case Study

Agriculture sector is an equally important industry to the country of Malaysia. Plant Ontology (PO) which is a part of Agriculture Ontology Domain is chosen as a case study for this research. This domain is chosen because Agriculture Ontology Domain is very huge. PO is built by Food Agriculture Organization (FAO) Consortium. This ontology domain contains a controlled vocabulary of plant's morphological and anatomical structures representing organ, tissue and cell types, and their relationships. Examples are stamen, gynoecium, petal, parenchyma, guard cell, etc. The main purpose of these vocabularies is to facilitate cross database querying and to foster consistent use of these vocabularies in the annotation of tissue and/or growth stage specific expression of genes, proteins, and phenotypes.

The PO represents a first step toward a unified vocabulary for flowering plants. Plant phenotypic descriptor (e.g. gynoecium, leaf) is often common English words that have been applied with varying degrees of precision; the same term can be applied to quite different structure (e.g., floret in Compositae and in Poaceae) or conversely different terms can be applied to quite similar structures (e.g., some legumes, follicles) (http://www.plantontology.org/).

3.0 RELATED WORKS

3.1 Web Search Engines

The first Web search engine was Wandex, a now-defunct index collected by the World Wide Web Wanderer. It also used Web crawler technique to fetch Web pages. It was developed by Matthew Gray at MIT in 1993. Another very early search engine, AliWeb, also appeared in 1993, and still run today. One of the first "full text" crawler-based search engines was WebCrawler, which came out in 1994. Unlike its predecessors, it let users search for any word in any Web page, which became the standard for all major search engines since. Soon after, many search engines were introduced such as Excite, Infoseek, Inktomi, Northern Light, and AltaVista. Yahoo! was among the most popular ways for people to find Web pages of interest, but its search function operated on its Web directory, rather than full-text copies of Web pages. Information seekers could also browse the directory instead of doing a keyword-based search (http://en.wikipedia.org/wiki/Search_engine).

3.2 Web Documents Searching using Conventional Search Engine

AltaVista offers both a Simple Search and an Advanced Search page. Simple Search requires the use of implied Boolean logic (plus and minus), while Advanced Search requires full Boolean logic (and, or, and not). The Advanced Search page also requires that the most important keyword appears in the "Sort by" search box in order to return the most relevant results at the top of the list. Both the Simple and Advanced pages support phrase searching, field searching (title, domain, host, URL, and link searches), and truncation (*) (http://www.learnWebskills.com/search/altavista.html).

Google uses mathematical formulas to rank a web page based on the number of "important" pages that link to it. When a search is conducted, Google determines the websites that meet the search criteria and then lists the most popular sites among high-quality resources at the top of the list. Google uses combination of techniques such as keyword, link, title, domain, and host searching. Using the Image Search database, News Search database, or Discussion Group Search database, visitors can search for pictures/graphics, news articles, and newsgroup postings [http://www.learnWebskills.com/search/google. html).

3.3 Web Documents Searching using Ontology Domain

The next generation of intelligent search engines allows elaborated more queries (for example, in natural language) and understands what people ask. In addition, these are implemented by using ontology and its related technologies. Ontologies will form the brains of the search engines because it tries to understand the user queries.

Ontology driven to Semantic Retrieval of Web Documents is a system to generate a semantic document model for each search session which can be semantically searched and browsed by user. The use of domain specific ontology is to reflect the user's "shared agreement" on the meaning of information. This system allows users to submit query in the form of natural language. This query is syntactically and semantically analyzed using natural language analysis (NLA) technique and domain ontology to form a query model. Commercial search engine is used to execute the query model and result from the search engine is analyzed to extract candidate concepts that are potential describe the document content. Heart Ontology which is part of Medical Subject Heading (MeSH) domain ontology was used for purpose of the research (Arifah et al., 2007).

Ontobroker is one of the search engines that comprises of language and tool to semantically markup content on Web pages and semantically query the WWW by taking advantage of semantic inference. At its core Ontobroker makes use of ontologies, which carry out the semantic markup of Web documents. Ontologies also provide for querying interface that formulate semantic rules for the domain (Roopa et al., 2005).

3.4 Mobile Web Searching

Mobile search used to consist solely of SMS (Short Message Service), which provides simple information on weather, flights, stock quotes, etc. based on text message requests. Similarly, MMS (Multimedia Message Service) provides ringtones, wallpapers, animated icons, and small videos upon request. Now, mobile search results are collected from numerous databases and indices, which are generated by spiders that crawls the Mobile Web. These collections consist of WAP sites with content similar to SMS and MMS services, or are filtered from WWW-centric real-time feeds, like Google and Yahoo!, drawing upon their huge database of local search Web pages.

3.4.1 Non Document Based

Live Search for Windows Mobile is the best way to find key local information on the move. Live Search via the mobile browser searches the broadest range of information and is available on any mobile phone with Internet access. Live Search provides streamlined access to all of the Web's information such as for latest weather forecast, a stock quote, news headlines, or a Website. It provides powerful local features such as local business search, driving directions, and detailed maps and traffic that work on almost any phone with Internet access (http://livesearchmobile.com/any_mobile_phone.htm).

Figure 2: Google image search result

Google Image Search uses text based image retrieval. A crawler (also referred to as a spider or robot) follows all links on the Web and finds keywords. These keywords are indexed in a database and when a query is performed using the search engine's Web browser, links to the Web pages containing the images are displayed as a list. The results are sorted on relevance and how popular the site is. To use Google Image Search, the user submits a text-based query against their database. The results are in form of a thumbnail of an image and a link to the site where the original image is situated. Refer to Figure 2 (Anne, 2007).

3.4.2 Document Based

Figure 3: Google interfaces for a WAP enable mobile phone and PALM PDA

The famous web search engine Google also has wireless search solution for WAP, PDA, I-mode, and J-Sky devices on the www.google.com.server. Figure 3 shows search results for the query 'sonera' on a NOKIA 6210 WAP phone and on a Palm M505 PDA (Sonera MediaLab, 2002).

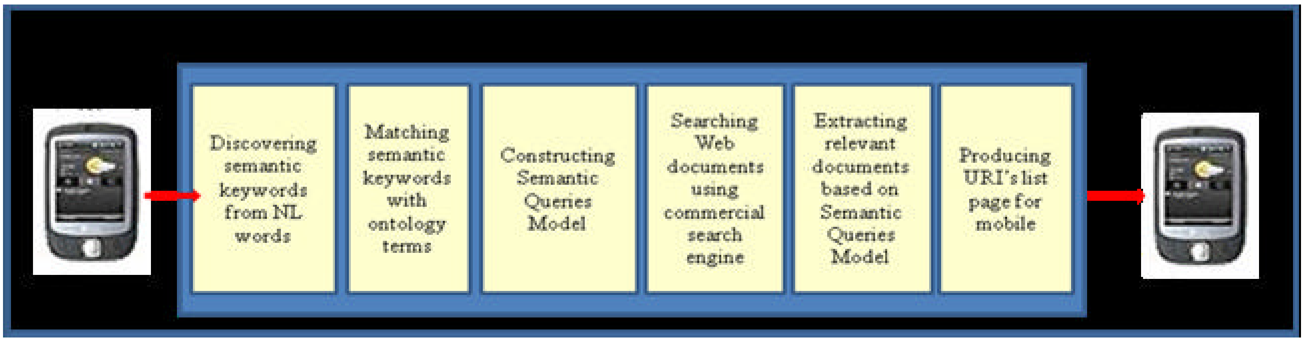

Figure5: Query searching using ontology for Mobile Web model

Google Web Search, on the other hand allows searching all the content in the Google index for desktop Web browsers. Because this content is not written specifically for mobile phones and devices, and thus might not display properly, Google automatically translates these pages by analyzing the original HTML code and converting it to a mobile-ready format. To ensure that the highest quality and most useable Web pages are displayed on mobile devices, Google may resize, adjust, or convert images, text formatting, and certain aspects of Web page functionality.

4.0 THE PROPOSED FRAMEWORK

Web information access is supported by software system that is distributed across a number of network elements. A program running on the user's mobile terminal interacts with the network infrastructure Web server to retrieve the requested data. The program on the mobile terminal can be a full-fledge application with a proprietary user interface. An example of such an application is a mobile search engine. In this case, the application needs to run on top of the mobile terminal's operating system (OS). The OS provides the developer with an API to access services such as user interface, communication, etc.

Figure 4: Query searching using ontology for Mobile Web networks architecture

The mobile device provides query searching interface that is communicated through a wireless network to send user's NL queries to a server. All processing for syntactic and semantic analysis, and accessing to Web documents occur in the server. Finally, all the results for the relevant Web documents are displayed at the user's mobile screen. The main architecture of a query searching using ontology for Mobile Web model is as shown in Figure 4.

4.1 The Mobile Web Model Components

Figure 5 shows the components of a query searching using ontology for mobile Web model. There are 8 components involved which are explained below:

4.1.1 Mobile Query Interface

The mobile query interface is an interface for user to key in the NL based query which is developed in the mobile device. The interface captures the query and submits the query from mobile network to a server to perform the NL analysis.

Example 1: A user submits the NL query, "What mutation affect in Inflorescence" from the mobile device, and then the query is submitted to the server.

4.1.2 Discovering Semantic Keywords from NL

The second component discovers semantic keywords from the user's query. The NL analysis tool is used to parse the NL query to determine the semantic keywords. The semantic keywords created are used in the next component.

Example 2: (Continued from Example 1), after the server received the user query, which is "What mutation affect in Inflorescence", the NL analysis tool parsed the words into semantic keywords.

4.1.3 Matching Semantic Keywords with Ontology Terms

The third component matches semantic keywords with ontology terms. This process is to generate all of the direct and alternate synonym/polysemy words which are relevant to the semantic keywords.

Example 3: (Continued from Example 2), the semantic keywords, "mutation", "affect", and "Inflorescence" are matched with PO Domain Database to search for synonym words. The words such as "panicle", "cyme", "raceme", "cob", and "spike" are synonym for the "Inflorescence".

4.1.4 Constructing Semantic Queries Model

The fourth component generates all of the direct and alternate synonym/polysemy words to create possible structured queries. The structured queries created are used to construct Semantic Queries Model.

Example 4: (Continued from Example 3), the constructing Semantic Queries Model component constructed a list of words that contains "mutation", "affect", "Inflorescence", "panicle", "cob", "cyme", "raceme", and "spike".

4.1.5 Accessing Web Documents using Commercial Search Engines

The fifth component is where a Semantic Queries Model is used to access all relevant Web documents through WWW using a commercial search engine. A popular search engine which is Google is chosen to get all relevant Web documents to the user's queries.

Example 5: (Continued from Example 4), the Semantic Queries Model which contains a list of "mutation", "affect", "Inflorescence", "panicle", "cob", "cyme", "raceme", and "spike" becomes a keyword to Google search engine. All documents related to those keywords are captured and displayed to server.

4.1.6 Extracting Relevant Documents based on Semantic Queries Model

The sixth component extracts the relevant documents from the URI's being captured by Google search engine. The URI list which is captured by Google is filtered based on the Semantic Queries Model to get relevant URI Web documents and remove all irrelevant URIs which do not match with the Semantic Queries Model.

Example 6: (Continued from Example 5), the URI's Web document captured by Google, e.g., URI#1: "*A line, **Inflorescence**. The flowering axis or other **panicle**. Specialised flowering structure of a plant such as an umbel, **raceme**, **spike** etc. Immature seed**...** www.seed tamilnadu.com/glossary1.htm - 208k*" are filtered based on the keywords "mutation", "affect", "Inflorescense", "panicle", "cob", "cyme", "raceme", and "spike" of Semantic Queries Model.

4.1.7 Producing URI's List Page for Mobile

The seventh component develops page which contains all of the extracted relevant URI's documents for mobile environment.

Example 7: (Continued from Example 6), the constructed page for mobile is as shown in Figure 6 below:

```
<wml>
<card id="card1"
title="Synonymy/Polysemy for">
&Inflorescense
<p>
<a href= URI#1> </a>
<a href= URI#2> </a>
<a href= URI#5> </a>
</p>
</card>
</wml>
```

Figure 6: Mobile page coding

4.1.8 Displaying the Relevant Results on Mobile Screen

The last component displays the entire relevant Web documents on the user's mobile screen. The results for the user's query are displayed on the user's mobile screen as a list of URI link.

Example 8: (Continued from Example 7), the relevant Web documents are displayed as URI's list as shown in Figure 7:

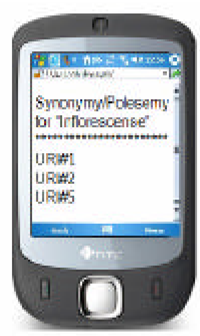

Figure 7: Relevant documents on mobile screen

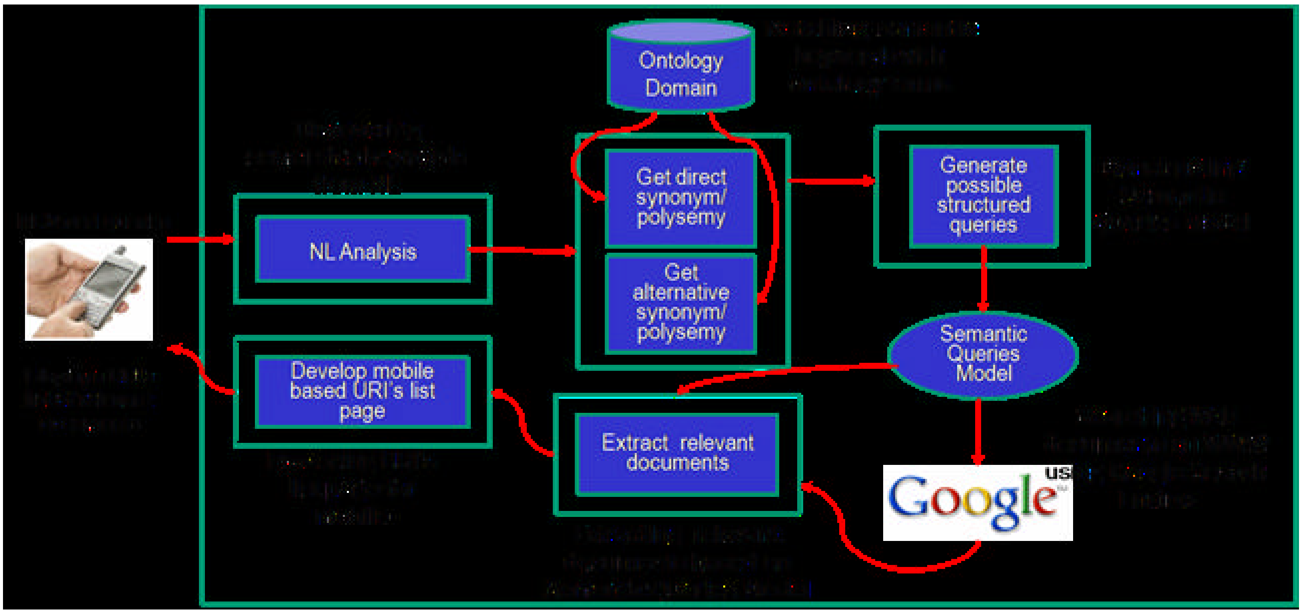

Figure 8: Querying Searching using Ontology Detail Process

The detail processes of a query searching using ontology for Mobile Web model is as shown in Figure 8.

5.0 CONCLUSION

Mobile search plays a critical role in agriculture environment, bringing tangible benefits to people by helping them to find desirable information. As we all know, synonymy and polysemy had effect on precision and recall, so that using the ontology domain can benefit people, mainly in agriculture field to access as much as information that meet with their expectation everywhere and anywhere.

Future development work will integrate to other Agriculture Ontology domain such as Gene Ontology, Trait Ontology, Environment Ontology, and etc. Secondly, all background process at server level can be done at mobile environment and lastly converting the Web pages from conventional Web to Mobile Web.

REFERENCES

Arifah Che Alhadi, Shahrul Azman Noah, & Lailatul Qadri Zakaria. (2007). Pendekatan Ontologi dalam Capaian dan Perwakilan Semantik Dokumen Web, *Jurnal Teknologi*, Universiti Teknologi Malaysia, 46, D, 103-120.

Anne Staurland, Aarbakke. (2007). *M2S and CAIR:Image based Information Retrieval in Mobile Environments*, Master's thesis. Faculty of Science, Department of Computer Science, University of Tromsø.

Bruijn, J.D. (2003). Using Ontologies: Enabling Knowledge Sharing and Reuse on the Semantic Web. *Technical Report DERI-2003-10-29*, Digital Enterprise Research Institute (DERI).

Daniel Appelquist. (2005). Position Paper: Enabling the Mobile Web through Semantically-Driven User Experiences. *10th MobEA III Workshop, Chiba, Japan*.

Li Ding, Tim Finin, Anupam Joshi, Yun Peng, Rong Pan, & Pavan Reddivari. (2005). Search on the Semantic Web. *Computer Magazine*, 38, 10, 62-69.

Michael Uschold. (2003). Where are the Semantics in the Semantic Web? *AI Magazine*, 24.3.25-36.

Roopa Jakkilinki, Nalin Sharda, & Imran Ahmad. (2005). Ontology-Based Intelligent Tourism Information Systems: An Overview of Development Methodology and Applications. *TES2005: Tourism Enterprise Strategies: Thriving-and Surviving-in an Online Era, Centre for Hospitality and Tourism Research (CHTR)*, Victoria University, Melbourne, Australia.

Sonera MediaLab. (2002). Mobile Search Engines, White Paper, November 15 [www.medialab.sonera.fi/workspace/MobileSearchEnginesWhitePape.pdf]

http://en.wikipedia.org/wiki/Search_engine [Accessed on February 29, 2008]

http://livesearchmobile.com/any_mobile_phone.htm [Accessed on January 15, 2008]

http://www.learnWebskills.com/search/main.html [Accessed on March 11, 2008]

http://www.learnWebskills.com/search/altavista.html [Accessed on February 29, 2008]

http://www.learnWebskills.com/search/google.html [Accessed on February 29, 2008]

http://www.thelearningsite.net/cyberlibrarian/searching/lesson3.html [Accessed on March 2, 2008]

http://www.plantontology.org [Accessed on February 29, 2008]