Searching the Web Querying Services over Ontological Process

Kavya. R[1], Seema.S[2]

PG Student, Department of Computer Science & Engineering, M.S.Ramaiah Institute of Technology,

Bangalore, India[1]

Associate Professor, Department of Computer Science & Engineering, M.S.Ramaiah Institute of Technology,

Bangalore, India[2]

Abstract: Modern hand held devices such as smart phones and PDA's have become increasingly powerful in recent years. In particular, most cell phones regularly include cameras, processors comparable to PCs from only a few years ago, and internet access. As mobile devices become more like PCs they will replace object we tend to carry around such as cameras, planners, mp3 players etc. In short, we will be using them to accomplish our daily tasks. One application that falls into this category is the Internet service provider application developed for Android phones. The prime objective of "Searching web querying service" is to create a fully fledged Android application which helps the user to get exact information based on the query provided. In this application the user gets the synonyms based on the query provided and those synonyms can be used to fetch exact information from the Web. Thesaurus can be used as a form of controlled vocabulary to aid in the indexing of appropriate metadata for information bearing entities.

Keywords: Ontological and terminological resource, Android, Thesaurus, Synonyms.

I. INTRODUCTION

Web is not a set of semistrctured documents interconnected via hyperlinks. In order to integrate data, a preliminary step consists of harmonizing external data with local ones i.e., external data must be expressed with the same vocabulary as the one used to index the local data. Searching the semantic web querying service over fuzzy ontologies is a model for an Ontological and Terminological Resource that is dedicated to the task of n-array relations annotation in Web data tables. This task relies on the identification of the symbolic concepts and the quantities, defined in ontology. Ontology deals with the synonyms which could denote the same concept in a multilingual context. The structure of the ontology is dedicated to the task of n-array relations as a particular domain of interest. The model of ontology and its use in an existing method for querying of Web server is been presented.

Thesaurus can be used as a form of controlled vocabulary to aid in the indexing of appropriate metadata for information bearing entities. A thesaurus helps with expressing the manifestations of a concept in a prescribed way, to aid in improving precision and recall. The semantic conceptual expressions of information bearing entities are easier to locate due to uniformity of language. Additionally, a thesaurus is used for maintaining a hierarchical listing of terms usually single words or bound phrases that aid the indexer in narrowing the terms and limiting semantic ambiguity.

The Google Web Search API lets us to put Google Search in the web pages with JavaScript. A simple, dynamic search box can be embedded and displays the search results on the web pages.

This system is developed as an android application. Linux operating system kernel is enhanced in mobile system that delivers low level interface with the hardware, memory management and process control.

A. Problem Definition

- The recognition and the representation of enormous data appearing in the list.
- Computation and explicit representation of the semantic terms using ontological and terminological resource.

B. Objective

The main objective of the semantic annotation method is to identify which relations of the OTR are represented in a data list, this concept is the simple target concepts and annotation of the extracted web data is shown as an output to the user.

C. Scope of the study

To provide efficient and exact access to data provided in the form of query by the user.

As the query is provided by any user it finds the related data and provides the user with the synonyms which can be used by the user to get semantically obtained data from the web and then it is provide to the user a proper output.

Each user has different thinking process and search for different data. This system holds good for the user creating different events. User can provide the query in the form of text which is used to retrieve the data from web documents as the synonyms and this synonym is used to fetch the related links.

II. LITERATURE SURVEY

An Ontology-driven workflow [1] feeds and queries a data warehouse opened on the web. Data are extracted from data tables in Web documents. Web documents are very heterogeneous in nature and the key issue in this workflow [1] is the ability to access the reliability of the retrieved data. Main step is to annotate the data and query web data tables driven by domain ontology. Methods are used to assess web data table reliability by means of evidence theory. Workflow is extended to integrate the reliability assessment step[1].[2] Companies, governmental agencies and scientists produce a large amount of quantitative (research) data, consisting of measurements like e.g. the surface temperatures of an ocean to the viscosity of a sample of mayonnaise.[2] Such measurements are stored in tables in e.g. spreadsheet files and research reports.

To integrate and reuse such data, it is necessary to have a semantic description of the data. A new disambiguation strategies based on an ontology is introduced, which allows to improve performance on "sloppy" datasets.[3] It is crucial to associate linguistic information with ontologies and more expressive models, beyond the label systems implemented in RDF, OWL and SKOS, are needed to capture the relation between natural language constructs and ontological structures. [3]In the light of tasks such as ontology-based information extraction (i.e., ontology population) from text, ontology learning from text, knowledge-based question answering and ontology verbalization, and currently available models are not sufficient as they only allow us to associate literals as labels to ontology elements. LexInfo has been implemented as OWL ontology and is freely available together with an API. A clear motivation is provided that why more elaborate models for associating linguistic information with ontologies are needed. [4]There are a large number of ontologies currently available on the Semantic Web. However, in order to exploit them within natural language processing applications, more linguistic information that can be represented in current Semantic Web standards is required.

There are a large number of lexical resources available representing a wealth of linguistic information, but this data exists in various formats and is difficult to link to ontologies and other resources. [4] A model called lemon (Lexicon Model for Ontologies) that supports the sharing of terminological and lexicon resources on the Semantic Web as well as their linking to the existing semantic representations provided by ontologies. Lemon can succinctly represent existing lexical resources and in combination with standard NLP tools it can easily be generated, new lexica for domain ontologies according to the lemon model. It is demonstrated by combining generated and existing lexica It can be collaboratively developed, rich lexical descriptions of ontology entities. [5] A real-world semantic information retrieval tool for auto- motive diagnosis. Troubleshooting documents have always been popular within car work- shops / manufacturers as a simple and direct way to capitalize knowledge on the one hand and to access repair information on the other hand. However, with more and more complex vehicle architectures, troubleshooting bases have grown so much nowadays that finding relevant pieces of information can become harder than looking for a needle in a haystack. Based on a limited knowledge model of automotive diagnosis, software aims at relieving car mechanics from the burden of storing and semantically searching through a large set of breakdown cases. Maximizing the data [6] and progressing the system using ER with a limited amount of work using "hints," which give information on records that are likely to refer to the same real-world entity. A hint can be represented in various formats (e.g., a grouping of records based on their likelihood of matching),A family of techniques is introduced for constructing hints efficiently and techniques for using the hints to maximize the number of matching records identified using a limited amount of work. Using real data sets, system illustrate the potential gains of our pay-as-you-go approach compared to running ER without using hints.

III. METHODOLOGY

A. System Design

The systems architect establishes the basic structure of the system, defining the essential core design features and elements that provide the framework. The systems architect provides the architects view of the users' vision. Given Figure1 shows that the user provides query and query will be converted into search based on ontology & terminology based queries. Thesaurus web service is service used to get the synonyms based on ontology process and then the related semantic words are provided to the user where the user selects for synonym required and then the search API is used. Search API is a Google search API which helps in extracting and displaying the web links to the user.
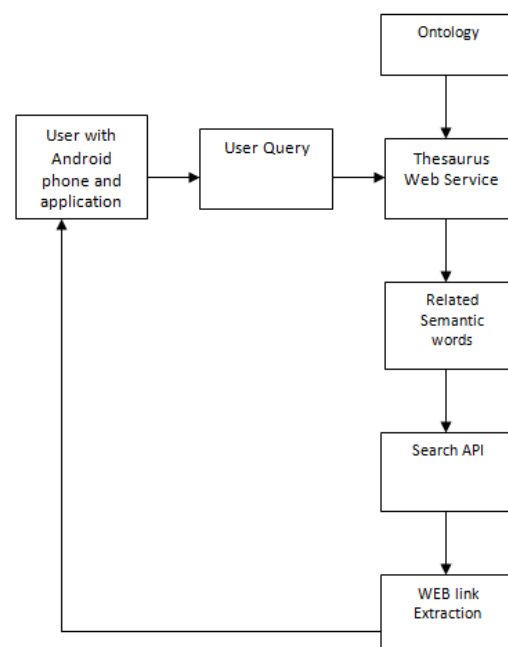

Figure 1: System Architecture

Thesaurus can be used as a form of controlled vocabulary to aid in the indexing of appropriate metadata for information bearing entities. A thesaurus helps with expressing the manifestations of a concept in a prescribed way, to aid in improving precision and recall. The semantic conceptual expressions of information bearing entities are easier to locate due to uniformity of language. Additionally, a thesaurus is used for maintaining a hierarchical listing of terms usually single words or bound phrases that aid the indexer in narrowing the terms and limiting semantic ambiguity.

In this application, the data or input given by user will be searched for the synonyms using an ontological process that is thesaurus service and the obtained synonyms are provided to the user. After which the synonym will be used by other process to provide the list of links to the user. Here query, list of synonyms, selection of synonym and the list of links are the use cases and the user is an actor. Users send the query and receive the synonyms then again send the synonym and obtain the links.

IV.   IMPLEMENTATION

Android is made up of several essential and dependent parts such as a hardware reference scheme that defines the capabilities required for a mobile device to maintain the software stack. A Linux operating system kernel that delivers low-level interface with the hardware, memory management, and process control, are enhanced for mobile devices. Open-source libraries for application improvement, includes SQLite, Web Kit, OpenGL, and a media executive. A run time used to perform and host Android applications, including the Dalvik virtual machine and the core libraries that run Android-specific functionality.

A. Algorithm

There are four modules in the system Mainactivity, Synonymlist, Linklists ans splash. Each module works for the system in different ways. The main activity is used to obtain the user query and check for query if it is correct. Then the process continues with the synonym selection. In this process Synonymslist is used to fetch the synonym and provide to the user. The technique used is ontological process where thesaurus web service is used for synonyms selection. Then the process continues for the link selection where links are fetched from the Google search engine based on the annotation.

Step 1: Start
Step 2: User provides the query
Step3: Checks the related synonyms using the ontological process if (the keyword matches) then synonym is displayed
Step 4: All related synonyms are displayed
Step 5: Select the required synonym.
Step 6: Extraction takes place using Search API if (link in the document is related to the synonym) then it display the link
Step 7: Links related to that synonym are displayed
Step 8: Select any link to open the page
Step 9: End

The run time is designed to be trivial and efficient for use on mobile devices. An application framework that agnostically uncovering system services to the application layer, plus the window manager and location manager, content providers, telephony, and sensors. A user interface structure used to host and launch applications. Preinstalled applications transported as part of the stack.

As this process run on the system the bytecode is converted to dex file using Delvic virtual machine and that in turn is converted to apk file. This is an executable file which can be installed in to an android mobile and user can run it in mobile.

In this application user provides the query in the form of text and this text is used to search the synonym which is related to the query based on the keyword the synonym is searched. Thesaurus web service is used to search the synonym and the relevant synonyms are displayed to the user. The user can select the synonym required so that again the keyword is used to query the data from the web database and provide the relevant links to the users.

Information retrieval thesauri are formally organized so that existing relationships between concepts are made clear. For example, "fruits" might be linked to the broader concept of "fruits", and the narrower ones of "oranges", "lemons", etc. When the terms are displayed online, the links between them make it very easy to surf around the thesaurus, selecting useful terms for a search. When a single term could have more than one meaning, like tables (furniture) or tables (data), these are listed separately so that the user can choose which concept to search for and avoid retrieving irrelevant results. The idea is to guide the searchers to use the same term for the same concept, so that search results will be as complete as possible. Thesaurus is multilingual, equivalent terms in other languages are shown too. International standards, concepts are generally arranged hierarchically within facets or grouped by themes or topics. Unlike a general thesaurus used for literary purposes, information retrieval thesauri typically focus on one discipline, subject or field of study.

The JSON format is syntactically identical to the code for creating JavaScript objects. Because of this similarity, instead of using a parser (like XML does), a JavaScript program can use standard JavaScript functions to convert JSON data into native JavaScript objects. A common use of JSON is to read data from a web server, and display the data in a web page.

With the rise of AJAX-Powered sites, its' becoming more and more important for sites to be able to load data quickly and asynchronously, or in the background without delaying page rendering. Switching up the contents of a certain element within the layouts without requiring a page refresh adds a wow factor to the applications, not to mention the added convenience for the users. Because of the popularity and ease of social

media, many sites rely on the content provided by sites such as twitter, flicker and others. These sites provide RSS feeds, which are easy to import and use on the server side, but if tried to load them with AJAX, its run into a wall.

Parsing is the lost art of identifying all the components of a text, and once one of the fundamental exercises that tested and informed pupils in English. To parse a phrase such as 'man bites dog' involves noting that the singular noun 'man' is the subject of the sentence, the verb 'bites' is the third person singular of the present tense of the verb to bite, and the singular noun 'dog' is the object of the sentence.

Information retrieval thesauri are formally organized so that existing relationships between concepts are made clear. For example, "fruits" might be linked to the broader concept of "fruits", and the narrower ones of "oranges", "lemons", etc. When the terms are displayed online, the links between them make it very easy to surf around the thesaurus, selecting useful terms for a search. When a single term could have more than one meaning, like tables (furniture) or tables (data), these are listed separately so that the user can choose which concept to search for and avoid retrieving irrelevant results. The idea is to guide the searchers to use the same term for the same concept, so that search results will be as complete as possible. Thesaurus is multilingual, equivalent terms in other languages are shown too. International standards, concepts are generally arranged hierarchically within facets or grouped by themes or topics. Unlike a general thesaurus used for literary purposes, information retrieval thesauri typically focus on one discipline, subject or field of study.

V.   TEST CASES

A. User Mobile

- Battery is checked so that it does not create problem during the execution of the android app. If there is no charge it can be connected and charged.
- In case of charging it has to show it is charging.
- In case of not charging and battery is low it has to show the user that the battery is low.
- In mobile key pad is important it has to be checked if all the numbers and alphabets are working.
- Network is very important in this project were the mobile data has to be kept enabled which helps the user to easily search for the query.
- If the application is not in use mobile data can be kept as disabled.

B. Search

- Access to the search page is when the user selects for the application it has to execute. OTR Search is an application which has to be selected to enter into the search page.
- User then provides the query in the search space so that the related data is obtained. The Synonyms based on the query is fetched and displayed to the user.

- If the mobile data is not enabled then the synonyms are not fetched and the output for the user will be no synonyms.
- If the spelling of the query is wrong then the synonyms are not fetched.
- If the query is any special character or any number then the synonyms are not fetched.

C. Output

- When the query is obtained by the user then the synonyms has to be checked whether it is related to the query or not.
- After the synonyms are selected links obtained has to be checked if it is related to that particular synonym.
- The number of links obtained has to be checked if it is the same as required.
- List of link is provided from which any one of the links can be selected which then opens that link page.
- If there is no proper browser then the link page is not opened.

VI.   CONCLUSION

This is the system used for searching the web querying service over ontological and terminological resource. It is built using the recommendation of the W3c, on a generic ontological process. It allows the web data extracted from web documents, to be displayed. The model for an Ontological and Terminological Resource dedicated to the task of n-array relations in Web data servers. This task relies on the identification of the concepts and the quantities, defined in the ontological and terminological resource. This system allows a separation between the terminological and conceptual components and allows dealing with synonyms which could denote the same concept in a multilingual context. The ontological and terminological resource is composed of a generic part to represent the structure of the ontology dedicated to the task of n-array relations annotation in data tables for any application and of a specific part to represent a particular domain of interest.

REFERENCES

[1] S'ebastien Destercke ,Patrice Buche and Brigitte charnomordic,"Data reliability assessment in a data warehouse opened on the Web",2011.
[2] Mark van Assem,Hajo Rijgersberg,Mari Wigham and Jan Top,"Converting and Annotating Quantitative Data Tables",Vol.Part I,Nov. 2010.
[3] John McCrae,dennis Spohr and Philipp Cimiano, "Linking Lexical resources and ontologies on the semantic web with lemon", Springer-Verlag Berlin, Heidelberg ,2011.
[4] Axel Reymonet,Jerome Thomas and Nathalie Aussenac – Gilles,"Ontology Based Information Retrieval: An application to automotive diagnosis",Conference 2009.
[5] Steven Euijong Whang,David Marmaros and Hector Garcia-Molina,"Pay-As-You-Go Entity Resolution",IEEE Transaction on knowledge and data engineering,Vol.25,May 2013.
[6] http://obitko.com/tutorials/ontologies-semantic-web/
[7] http://en.wikipedia.org/wiki/Fuzzy_logic