Intelligent Health Information System to Empower Patients with Chronic Diseases

Sifat Islam
PhD Candidate
Computer & Electrical Engineering and Computer Science
Florida Atlantic University
Boca Raton, FL

Glenn Freytag
President
Shilpico, Inc.
Deerfield Beach, FL

Ravi Shankar
Professor and Director
Center for Systems Integration
Florida Atlantic University
Boca Raton, FL

Abstract

Semantic Web technologies can be utilized to keep patients informed on the latest research on chronic diseases such as diabetes, by gathering online information published by both government and corporate sources. To assemble our Intelligent Health Information System, we created an ontology of terms related to diabetes. Then we leveraged an open-source API to read the ontology into a graphical application, which we also built in an open-source development environment. The application searched for articles on diabetes using a web service and a web crawler. The ultimate goal for our system is to synthesize the information that it gathers, make it customer-specific, and present it to the patient in an easy to understand format.

Keywords: Intelligent Health Information System, Semantic Web, Biomedical Informatics, Web Services, Web Crawler

1. Introduction

An Internet search using any of today's popular search engines may return thousands or even millions of results for a given set of keywords. Typically, only a handful of the results are relevant to the user's needs, and the user has to review many pages of results to find useful documents. If such documents are not found, the user may have to repeat the search with different keywords, in the hope that the desired documents will contain at least one of them. This process soon becomes tedious, time-consuming and frustrating. In a business environment, it wastes money and hinders productivity. There is thus a need for an application which automates the process of obtaining relevant results from Internet searches. This proposal describes an application which is under development for that purpose.

Internet search engines often return excessive numbers of results for several reasons:

- They consider any document relevant if it contains even one of the specified keywords.
- They make no distinction as to the context or type of document in which a keyword is used.
- They have no knowledge of the user and his/her needs.

Search engines may also miss relevant documents because they are unaware of synonyms or other terms related to the specified keywords.

We are utilizing Semantic Web technologies to address these shortcomings of generic Web searches. Our specific goal is to update patients with chronic diseases on the latest research related to their conditions. A patient will create his or her profile over 30 or 40 dimensions, such as height, weight, ethnicity, gender, blood chemistry, medication, allergies, disease condition, socioeconomic status, and state residency. Then, our tool will automatically dispatch a web crawler and make web service calls that will search the Web and medical databases to find and retrieve the latest research. The tool will synthesize this information, make it customer-specific, and present it to the patient in an easy to understand format. Our initial goal was to search only the www.fda.gov, and Medline database. The tool is being extended to other medical databases and websites.

To access such resources and align data from them, we will have to develop a semantic web infrastructure. This will require us to develop an ontology and to set up rules for translation from medical databases to our Semantic web database. The results will be displayed to the patient as RSS (Really Simple Syndication) feeds.

2. Background

The Semantic Web, aptly labeled Web 3.0, has had important applications for Web 2.0 social networking and collaboration aspects. The Semantic Web offers a powerful, practical approach to gain mastery over the multitude of information and information services. Tim Berners-Lee, the visionary behind the World Wide Web, has said that "… if properly designed, the Semantic Web can assist in the evolution of human knowledge as a whole." [1] At a more practical level, the Semantic Web is a strategic technology that truly provides a solution with a significant advantage and lucrative opportunities. SEC's XBRL [2], Berners-Lee's Linked Data [3], and GeoNames [4] are large scale success stories. However, more focused smaller scale applications have the potential to personalize the web-experience of the individual. They can also empower the individual and protect his/her rights and personal information.

Several different components were required to assemble our system. First, we utilized Common Terms from the American Diabetes Association (ADA) and the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) [5] to create our ontology via Protégé Ontology Editor [6]. The ADA formed a list of diabetes-related terms and their definitions, which was adapted from the NIDDK. The NIDDK performs and assists fundamental and clinical research on many of the most life-threatening diseases affecting public health. We used our ontology in reference to sample profiles that represent various diabetes patients. To read the ontology into our Semantic Web application, we leveraged the open source, Java-based OWL API from the University Of Manchester [7]. The OWL API is proving popular with many developers around the world. It is a very clean API that closely follows the OWL specification, and the parser is optimized to be faster and use less memory. To request relevant medical information, we utilized the European Bioinformatics Institute's (EBI) CiteXplore Web Service [8], a Simple Object Access Protocol (SOAP) based service, to fetch data populated from the Medline database. The CiteXplore literature database offers integrated databases of literature information from a range of resources and contains references to biological databases, text mining findings and links to locations of the abstract or full text version of the citation. To search federal medical websites, we utilized Lucene Web Crawler code from chapter 2 of Algorithms of the Intelligent Web by Marmanis and Babenko [9].

3. Methodology

Our search methodology is depicted in Figure 1. An ontology is a graph of the relationships between terms that are relevant to a particular topic. A reasoner is a program which interprets these relationships to find terms related to a specified term. The user profile contains personal information that the patient provides which is relevant to his/her condition. The patient selects search terms from the ontology and the user profile. The reasoner adds whatever terms it may find in the ontology that are related to the terms the patient selected. The Web Crawler and Web Service Engines then use the terms to search exhaustively through online databases that are related to the patient's condition. The engines generate search results, and the reasoner sorts them by relevance to the patient's needs, taking the user profile into account. Thus, our methodology can obtain valuable results that might otherwise be missed, discard results which are not relevant to the patient's needs, and rank the results in order of interest to the patient.

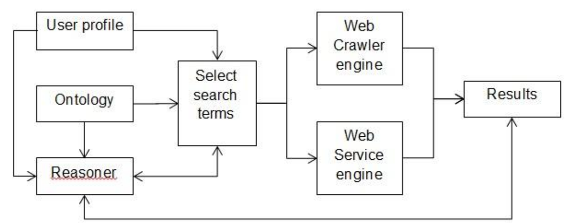

Figure 1. Elements of search methodology.

4. Semantic Web Application

Our application organizes both the ontology and the user profile as two-level lists, as shown in Figure 2. The ontology consists of Categories of terms relevant to the patient's condition. The terms themselves are called Keywords. The user profile is divided into Parameters, each of which may have one or more Values. The patient may select any desired combination of Parameters, Values, Categories and/or Keywords as the terms to be used in a web search.

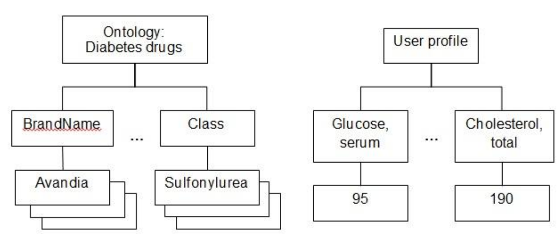

Figure 2. Sample ontology and user profile.

Figure 3 shows how the user interface of our prototype application is organized. The interface was built in the open-source Eclipse development environment using the Java language and the Swing library of user interface components. The interface has two tabbed panels, Profile and Results. On the Profile panel, the patient can edit the information in their profile or in the ontology for their condition (Edit mode), and then use the information to initiate a web search (Search mode). The Categories, Keywords, Parameters and Values are each displayed in one of four list boxes. The patient may click on a Category to view its list of Keywords or click on a Parameter to see its Values. In the Edit mode, the patient can add or delete terms in any of the four lists. In the Search mode, the patient selects terms from the four lists to add to a line of search terms. The application, by use of the reasoner, may suggest additional search terms based on the information in either the user profile or the ontology. The patient can even edit the line of search terms manually if desired. The patient then clicks a Submit button to begin the search.

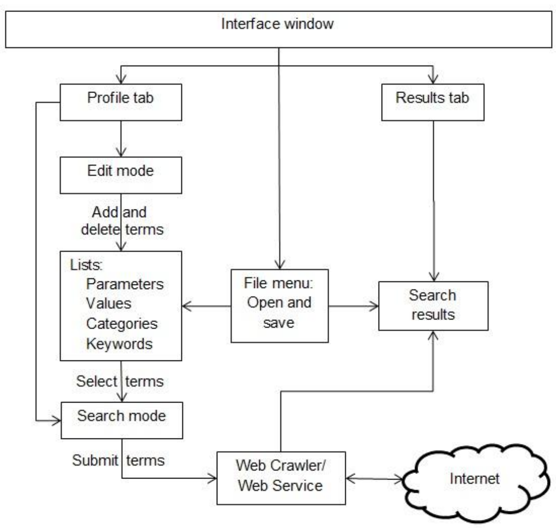

Figure 3. Organization of user interface.

Figure 4 shows that our application returns search results on the Results tab, which is divided into lists of Web Crawler and Web Service results. In both lists, each result includes a URL and a relevance Score. For example, the relevance score for Lucene Web Crawler is a value between 0 and 1. Highest relevance score is 1 and lowest is 0 [9]. On either tab, the File menu allows the patient to save the information being displayed and retrieve it later.

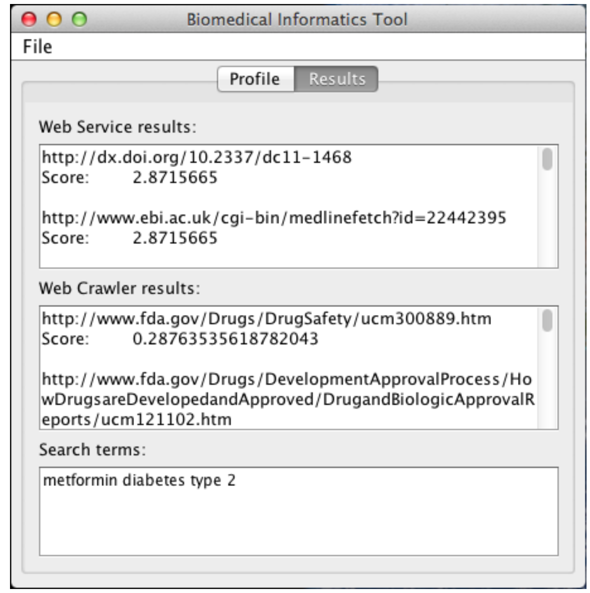

Figure 4. Our Semantic Web Application

5. Discussion

To improve the tool with regard to its results, performance, and code size, one can quantify the meta information and use the metrics. There are algorithm challenges with ontology and rule set creation. One of the main advantages of utilizing ontologies is that they can be managed by a reasoner. The reasoner can determine if a class is a subclass of another class. This functionality enables a reasoner to discover the inferred ontology class hierarchy. Another functionality of reasoners is consistency checking. Taking into account the conditions of a class, the reasoner can verify if the class can have any instances. A class is considered to be inconsistent if it cannot have any instances. Reasoning in Web Ontology Language (OWL) is founded on the open world assumption (OWA). It is often called open world reasoning (OWR). The open world assumption is the understanding that one cannot assume that something does not exist until it is clearly specified that it does not exist. Another way of saying the same thing is that just because something hasn't been specified to be true, it cannot be thought to be false. The understanding is that the knowledge has not yet been included in the knowledge base [6].

Utilizing data mining techniques, such as cluster analysis and association rule mining, one might be able to discover patterns to improve search results. Machine learning can be used to automatically identify complex patterns and make intelligent decisions based on feedback from the patient. We assume that the databases are reliable (for now), since they are generated by government agencies. Based on how accurate and recent the data is, one can consider assigning weight to each site and database. In addition to the desktop version, a mobile phone version can be developed. Since the profile is created by the patient, there is potential for incorrect information which can cause our tool to provide inappropriate results. The patient may wish to get the profile authenticated by his/her doctor. A tool for communication with the doctor and validation from the doctor can be helpful. To see our group's semantic web related activity, visit Florida Atlantic University's semantic web research site [10].

6. Conclusion

Google is an example of an excellent search engine that helps us navigate through the vast world of information that is available on the internet. As we gather more and more personal data, and want to stay up to date with the latest information available to us, we will need tools that are aware of our personal data and are able to gather relevant data from different sources. We have made progress in developing an intelligent health information system to empower patients with chronic diseases utilizing semantic web technologies. We integrated an ontology, web services, and web crawler to create an application that will inform patients with the latest research on their conditions from diverse sources of data. To further improve the intelligence of our application, we will be extensively utilizing the semantic reasoner. Our intelligent health information system is just one example among many of how semantic web techniques can help with the increasing need to obtain information from the growing amounts of data available to us.

References

[1] T. Berners-Lee, J. Hendler and O. Lassila, "The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities," Scientific American, May 2001.

[2] U.S. Securities and Exchange Commission. *Protal for information about eXtensible Business Reporting Language*, 2012
http://xbrl.sec.gov/

[3] Tim Berners-Lee. *Linked Data Design Issues,* 2006
http://www.w3.org/DesignIssues/LinkedData.html

[4] GeoNames, 2012
http://www.geonames.org/

[5] American Diabetes Association. *Diabetes Basics – Common Terms,* 2012
http://www.diabetes.org/diabetes-basics/common-terms/

[6] M. Horridge, "A Practical Guide To Building OWL Ontologies Using Protege 4 and CO-ODE Tools Edition 1.3," The University Of Manchester, March 2011.

[7] M. Horridge, and S. Bechhofer, "The OWL API: A Java API for Working with OWL 2 Ontologies" OWL Experiences and Directions (OWLED 2009), Washington D.C., USA.

[8] H. Mcwilliam, F. Valentin, M. Goujon, W. Li, M. Narayanasamy, J. Martin, T. Miyar, and R. Lopez, Web services at the "European Bioinformatics Institute-2009," Nucleic Acids Res. 2009 July 1; 37(Web Server issue): W6–W10. Published online 2009 July 1.

[9] H. Marmanis, and D. Babenko, "Algorithms of the Intelligent Web," Manning publications Co, 2009.

[10] FAU Semantic Web Course, 2012
http://semanticweb.fau.edu/