

## Development of a Semantic-based Search System for Immunization Knowledge

Li-Hui Lee<sup>a</sup>, Hsing-Yi Chu<sup>b</sup>, Der-Ming Liou<sup>c</sup>

<sup>a</sup> Institute of Public Health, National Yang-Ming University, Taiwan (R.O.C.)

<sup>b</sup> Institute of Health Informatics and Decision Making, National Yang-Ming University, Taiwan (R.O.C.)

<sup>c</sup> Institution of Biomedical Informatics, National Yang-Ming University, Taiwan (R.O.C.)

### Abstract

*This study developed and implemented a children's immunization management system with English and Traditional Chinese immunization ontology for semantic-based search of immunization knowledge. Parents and guardians are able to search vaccination-related information effectively. Jena Java Application Programming Interface (API) was used to search for synonyms and associated classes in this domain and then use them for searching by Google Search API. The searching results do not only contain suggested web links but also include a basic introduction to vaccine and related preventable diseases. Compared with the Google keyword-based search, over half of the 31 trial users prefer using semantic-based search of this system. Although the search runtime on this system is not as fast as well-known search engines such as Google or Yahoo, it can accurately focus on searching for child vaccination information to provide search results that better conform to the needs of users. Furthermore, the system is also one of the few health knowledge platforms that support Traditional Chinese semantic-based search.*

### Keywords:

Semantic-based Search, Vaccination, Ontology.

### Introduction

Vaccination significantly prevents the development of specific diseases [1, 2]. High vaccination rates are seen for comparatively older mothers and for those with greater levels of knowledge about child immunizations [3]. In addition, with the development of the World Wide Web (WWW), health resource websites have increasingly become an important source of knowledge about child vaccinations for parents and guardians. There are free medical and health resource sites with large amounts of evidence-based or clinical medical literature such as the Cochrane Library, but this literature can be difficult for lay people to comprehend. These sites often contain medical terminology and are usually targeted towards clinicians (doctors, nurses, radiologists, etc.) or academic researchers. This is in contrast to the child immunization information available on general websites, which is more readable and at a more appropriate level of difficulty. These websites can provide easily accessible information for parents and guardians who do not have medical or health backgrounds.

However, some issues were mentioned in searching information through general websites. These include the infor-

mation being filtered on the web, and the possibility that keyword-based searches can generate a large number of websites that do not contain relevant information, or even inadequate or incorrect information. The information sought by parents and guardians may not always be available from keyword-based searches [4]. In order to ensure more effective search strategies, the semantic web has been seen increasingly as an important foundation for current web search strategies [4, 5].

### Children's Immunization Information Search

Current well known children's immunization informational websites and search engines in English and Traditional Chinese largely adopt the keyword-based search strategy: examples include the Taiwan Centers for Disease Control (CDC), BabyHome, the National Taiwan University Hospital (NTUH), Everyday Health, Google, Yahoo, and Microsoft Bing. Apart from BabyHome, all of these websites provide synonyms for users to refine their searches. Websites that provide semantic search functions include the health information sites Web MD and HealthMash. While WebMD provides semantic-based searches in English only, HealthMash provides English and Simplified Chinese semantic searches [6]. HealthMash utilizes National Library of Medicine (NLM) Medical Subject Headings Thesaurus and the Unified Medical Language System (UMLS) as foundations for its semantic structure [7]. While it can provide searches in Simplified Chinese, it is still lacking in support for health and medical information search conducted in Traditional Chinese [6, 8]. Medstory, purchased by Microsoft in 2007, is perhaps the website that most closely matches the needs of web users in Taiwan; however, to date no formal website has been developed for it; one can obtain information regarding its earlier semantic functions only in the literature.

### Keyword-based and Semantic-based Search Design

The top search engines in use like Google, Yahoo, and Microsoft Bing provide keyword-based search functionality in order to meet users' needs, in which the suggested keyword searches are based on popular keyword combinations. However, these suggested keywords may distract users from their initial keywords, and simultaneously increase the popularity of existing suggested keywords; in the meantime, initially desired keywords that more accurately reflect the search topic may be ignored. If search phrases can undergo semantic analysis, or be presented in a domain ontology, this will avoid unnecessary high frequency phrases that might inaccurately reflect search intentions and might influence user behavior.

The foundation of semantic web construction lies with ontology design. According to expressivity and complexity of structure, an ontology could be a Taxonomy (terms with simple hierarchic structure), a Thesaurus (terms, synonyms, broader/narrower term and association), a Conceptual Model (concepts structured, relations, properties, attributes and instances) or a Logical Theory (extending with axioms and rules) [9]. The key development stages for an ontology include defining the domain definition, considering whether an existing related ontology is applicable, proposing terms from the related ontology, providing concept and hierarchy definitions, and establishing instances [10, 11]. Regarding existing related ontology concerning immunization, the UMLS Metathesaurus, not only is in relevant domains but also can be applied in the semantic web, thereby providing a more accurate and flexible search phrase to obtain precise biomedical knowledge. In addition, through BioPortal website, one can identify existing ontological structures for reference [12]. However, what is notable at present is that the majorities of related domain ontological structures are available only in English and cannot be used for search in Traditional Chinese characters directly.

### Research Aim

The purpose of the study is to develop a semantic-based search system for parents and guardians searching for immunization-related information in English and Traditional Chinese and then to evaluate its search ability. For developing the immunization ontology, domain experts are consulted. The ontology is designed in English and Traditional Chinese. Once users have keyed in their search phrases, the system uses the ontological relationships to identify synonyms and related words and then adopts the Google Search API to perform semantic search. This allows users to quickly obtain desired immunization information. Additionally, for presenting the complete immunization information in search results, the vaccinations and related diseases descriptions collected by the study are simultaneously displayed to users. Finally, to examine the usability of the immunization ontology in Traditional Chinese as a semantic-based search for Taiwanese users, a comparison with the search results of other search tools is also conducted.

### Methods

The following paragraphs detail the methodology and system structure used in the present research, as well as the evaluation processes used to verify the usability of the designed system.

#### System Architecture

In developing the semantic-based search system for children's immunization knowledge, in addition to referring to the literature, we also interviewed four mothers with young children and two specialists in the area. After system analysis and design, we proposed the functions and architecture as demonstrated in Figure 1. In addition to user account management, immunization schedule, immunization reminder, and schedule update, the system also provides related disease and vaccination information. The search function is the main core of this study shown in Figure 1. In an attempt to provide comprehensive information for the semantic-based searches, basic information on related diseases and vaccination are also shown in the search results. The complementary basic information derives from pediatric literature, 56 documents from web resources, and 13 documents from Taiwan CDC.

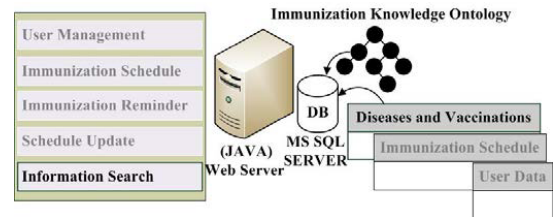


Figure 1 - System Functions and Architecture

The system was developed using Java, JavaScript, and Java Server Pages (JSP). The website is run by Apache Tomcat, and all individual user information, children's immunization records, and related-diseases and vaccinations information were stored in a MS SQL server database. The Protégé was used to implement the immunization ontology because it can export file into Web Ontology Language (OWL). After that, Jena API and Protégé OWL API were utilized to access the ontology content.

An immunization information search consists of four interface options. The first and second involve the user directly clicking on the vaccination name in the immunization schedule, and choosing from among preset frequently encountered vaccine and disease names that appear on the system. The third option involves the user typing in a search phrase that then undergoes a semantic-based search, and the fourth option utilizes a keyword-based search from the typed search phrases.

The processing workflow for options 3 and 4 are shown in Figure 2. Once a user has entered a search phrase, the system will convert all capitalized letters into lowercase in the English alphabet, and conduct either a keyword search or search using the entire phrase. If the search is conducted through the semantic-based search, the system will first check for Traditional Chinese and English synonyms in the immunization ontology. If synonyms identified, then variants of the synonyms are defined as keywords in the search. Apart from searching for synonyms in the search phrase, the system will also conduct ontological analysis (see Figure 3). For example, if the term belongs to the class named "Vaccine", then the disease prevented by this vaccine can be obtained by the relationship of "Prevents". Conversely, if the term belongs to class named "Disease", then the vaccine that can prevent this disease are linked by the relationship "Prevented-By". The obtained vaccine or disease names are designated as final search terms.

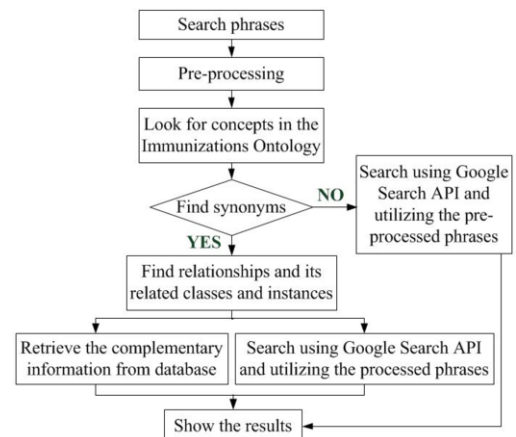


Figure 2 - Flowchart of search phrases pre-processing and execution

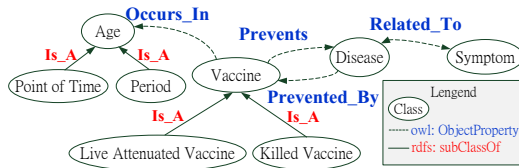


Figure 3 - Immunization ontological structure

Moreover, for search interface options 1 and 2, once the user has selected the vaccine or disease name, the system will use semantic-based search in a similar way, to compare against synonyms and classes from the immunization ontology and to conduct subsequent information searches. The benefit of these two search options is that it saves the time of having to type out the search phrase, and the user can target specific questions for further searching.

### System Evaluation

The system evaluation includes three key parts. The first part is the evaluation of the immunization ontology designed for this research. The team consulted specialists in clinical pharmacy and immunization to comment on and suggest improvements for the comprehensiveness of the ontology. The second part involves verifying the system performance. In addition, semantic-based search runtime, accuracy, comprehensiveness of results, and synonym searches are also analyzed. Lastly, the four mothers involved in the initial focus group interview, eight medical practitioners and nineteen computer science students were invited to use the system, providing feedback on their experiences with the system and suggesting areas for improvement.

### Results

The results broadly cover two areas. The first consists of findings concerning the immunization ontological design, and the second outlines the functional design of a semantic-based search for children's immunization management system. Assessments and explanations are provided for results from both areas.

#### Children's Immunization Ontology Design

Protégé was used for the final design of the ontological structure, as shown in Figure 3. Considering the majority of searches on children's immunization are on topics related to vaccine information, the disease prevented by a vaccine, symptoms of diseases, and optimal point-of-time or period for vaccination, after consulting the UMLS Metathesaurus, the four classes of vaccine, disease, symptom, and age were established. The vaccine class was further divided into live attenuated vaccine and killed vaccine, and age was separated into two classes of point-of-time and period, depending on the vaccination requirement. In addition, to improve the reliability of semantic-based searches, symptoms characterizing different diseases were established as additional relationships. Each class is interrelated via its ObjectProperty settings, including the terms "Occurs\_In", "Prevents", "Prevented\_By", and "Related\_To". Synonymous terms are presented by the DataProperty. Once the ontological structure is completed, additional instances are added to respective classes, and new instance synonymous are added as well. Finally, the relationships between instances are assigned by defined ObjectProperty (relationship).

### Support for Semantic-based Search in the System

In order to store resources on vaccines and relevant diseases, two tables were established. The vaccine table contains the ID, Traditional Chinese and English vaccine names, vaccine properties, an introduction to the vaccine, the suitable ages for recipients, dosage, mode of inoculation, injection sites, adverse reactions and contraindications, seroconversion rate, and duration of immunity. The disease table contains fields for ID, disease name, infectious agent, mode of transmission, period of communicability, symptoms of infection, the case definition, and methods of disease prevention, treatment, and related public health regulations.

The main system function "Information Search" has two sections on web page. The upper section displays the search function where a user can type in their search phrase from both semantic-based search functions of our system or a Google keyword-based search (Figure 4a). The lower section lists vaccines and related diseases identified from our internal database, where users can select these terms directly without performing another keyword search: this is the aforementioned option 2 search interface. Search option 1 is located on the "Immunization Schedule" web page, where users can directly select vaccine names to obtain further information. The snapshot of search result is shown in Figure 4b.

#### (a) Information Query

Please Enter Search Phrases

Semantic-based Search  Query

Google  Google-Query

Please Click on the Name for Search

Vaccine	
Hepatitis B (HepB)	Bacille Calmette-Guerin vaccine (BCG)
Oral polio vaccine (OPV)	Diphtheria and tetanus toxoid children's dose (DT)
Inactivated polio vaccine	Varicella vaccine (Varicella)
Measles, mumps, rubella and varicella vaccine (MMRV)	Japanese encephalitis vaccine (JE)
Tetanus toxoid (TT)	
Disease	
Hepatitis B	Tuberculosis (TB)
Diphtheria	Tetanus
Whooping cough	Polio
Chickenpox	Measles
Mumps	Rubella

#### (b) Semantic-based Search Results

Semantic Structure	
<b>+Vaccine</b> - Live Attenuated Vaccine 麻疹腮腺炎德國麻疹混合疫苗 麻疹疫苗 卡介苗 水痘疫苗 黃熱病疫苗 德國麻疹疫苗 口服小兒麻痺疫苗 <b>+Killed Vaccine</b> Synonyms : Bacille Calmette Guerin Vaccine - BCG Prevents : 結核菌 (Tuberculosis)	
The Number of Search Results: 16	
No.	Name
1	有團體的和個別的接種
2	預防接種有「建議接種」和「自願接種」
3	下次的接種時間會因疫苗種類而異
4	有關卡介苗副作用的新發現
5	全型多功能保護疫苗 小兒六合一疫苗來了
6	寶寶一出生就要注射疫苗，讓爸媽們準備好了嗎？
7	您該了解的b型肝炎性感冒病毒傳染
8	小兒疫苗接種的迷思
9	疫苗接種10大迷思
10	疫苗接種出新 爸媽該大?
First Previous 1/2 Next Last	
卡介苗 - Basic Information	
Name	卡介苗、BCG
Vaccine	牛型分枝桿菌（活性疫苗）
Definition	是一種牛的分枝桿菌所製成的活性疫苗，經滅毒後注入人體，可產生對結核菌的抵抗力，一般對初期侵染的預防效果

Figure 4 - System Interface and Display<sup>1</sup>

### System Evaluation

The system evaluation is divided into three sections. The ontological design and improvements developed from consulta-

<sup>1</sup> The original system interfaces were designed in Traditional Chinese. For the publication reason, web pages were fully or partially modified in English.

tions with specialists are shown in Figure 3. Furthermore, comparisons of search results among various search engines and the current system are listed in Table 1. We tested the runtime of semantic searches using the term “BCG” with the current system, Google, and Yahoo, and found Google used the least time, while the current system took the most. The

indicator of relevance of search results is assessed by whether relevant documents are listed at the first in the search results. It shows that HealthMash and current system produced links to relevant articles in the 1st search results, while Google, Yahoo, and Microsoft Bing had the relevant articles appearing between 6th and 8th in the search results.

Table 1 - Comparisons among search tools using the term “BCG” as an example

Comparisons		The Study	HealthMash	Google	Yahoo	Microsoft Bing
Runtime of search (microsecond)		188-220	ungettable	80	170	ungettable
The relatedness of results Which result is related to search phrases in the result list.		The 1st	The 1st	The 7th	The 8th	The 6th
The complete-ness of results (1) Functions	Sameness	<ul style="list-style-type: none"> <li>To show the ontological structure</li> <li>To present the definition of search phrases</li> </ul>		<ul style="list-style-type: none"> <li>To show related compositional terms.</li> <li>To filter the search location</li> <li>To filter the search language</li> <li>To translate the content of web</li> </ul>		
	Difference	The Study: <ul style="list-style-type: none"> <li>The ontological structure is shown in Traditional Chinese.</li> <li>The target resource for search is the same as Google's search engine.</li> </ul> HealthMash: <ul style="list-style-type: none"> <li>The ontological structure is shown in English.</li> <li>The target resource for search includes limited Taiwanese websites or Chinese articles are more likely in Simplified Chinese, not in Traditional Chinese.</li> </ul>		Google: <ul style="list-style-type: none"> <li>To filter the published date or duration of target resource</li> <li>To present the definition of search phrases from dictionaries</li> <li>To show the result according to the read difficulty level</li> </ul> Yahoo & Microsoft Bing: <ul style="list-style-type: none"> <li>To configure the number of search results per page</li> </ul>		
(2) To show synonyms		Yes	No	No	No	No

For the comprehensiveness of the search, Google, Yahoo, and Microsoft Bing all displayed terms that were related to the search phrase, but they were frequently terms entered by other users in previous searches or colloquial terms. HealthMash displayed BCG-related terms and definitions, with the search results also classified based on semantic structure. For example, with the term “BCG” entered as the search phrase, the results reflected the semantic structure of: Causes (including disease), Treatments (including therapy and prevention), and Risks and Effects (including bladder cancer and high risk). The sources of the results can also be classified into Health News, Videos, Images, Drugs and Substances, Clinical Trials, Web, Articles, Books, Blogs, and Twitter. The current system displays the definition of BCG, suitable ages for vaccination, disease prevented and related conditions in its semantic structure. Users can select any of these classes in the semantic structure to further their search. In addition, the system also provides synonymous terms so that users can familiarize themselves with context, and to increase their knowledge on the current search topic, a brief introduction to BCG is also displayed on the results page.

Additionally, to explore the differences in search outcomes between semantic-based and keyword-based searches for existing health information websites, the current study took “BCG” and its Traditional Chinese synonym “卡介苗” as search terms and typed them into a number of well-known health information websites in Taiwan: the results of this are described in Table 2, and suggest that other websites likely did not conduct their searches with synonymous terms, nor did they combine searches with vaccines or related diseases, and as a consequence a larger number of unrelated search results were produced.

After the functional comparisons, 31 individuals were invited to use the system. The participants all agreed that there is a

practical need for children's immunization management systems. Compared with the keyword-based search, 67.7% of responders preferred the semantic-based search option provided by the current system. Feedback for improvement covered the topics of the system interface design, ontological structure, and the system providing more related information in its results. Figure 4 is the modified interface design based on this feedback. With regard to ontological design, authors believe that changes should reflect knowledge advances in the field in order to provide the most accurate and up-to-date information for users. In order to improve the comprehensiveness of the search, the inclusion of vaccine pictures, current vaccine prevalence in Taiwan, and related health insurance information should be considered.

Table 2 - Comparisons of the related search results number among Taiwanese health information websites

Websites	The Study	Taiwan CDC	Baby Home	Health Education Center of NTUH
BCG	127,000	1,320	0	22
卡介苗	127,000	1,730	94	15

## Discussion and Conclusion

The present study designed a vaccine and related-disease information database whose implementation differs from the keyword-based search. This database presents accurate and relevant immunization information to parents and guardians by combining a semantic-based search function with an immunization ontology. While at present the search runtime lags behind other well-known search engines, the search results better target vaccination information and can reduce the amount of

time users spend looking for relevant information. Ontological design and semantic-based search runtime are discussed below, followed by an outline of the limitations of our research and future work.

Concerning runtime of semantic-based search, the present system's search runtime is more than that of other well-known search engines. This could be due to the fact that loading the embedded Google API requires extra time. In addition, the current search takes the time to focus more on relevant information sought by users, thereby reducing the overall time spent researching for relevant information. Search result relevance in general is defined by precision and recall. It is possible the numerous semantic restrictors will limit the amount of information identified; however, the search settings will avoid situations where no results are identified [8]. In order to avoid excessively strict search conditions, the system will modify search terms into those with less restriction when initial conditions produce no results without sacrificing the meaning of the original terms.

The related health information website are WebMD and HealthMash [6, 13]. Compared with the ontological designs and display of these two websites, the current system has a more concise ontological structure. WebMD and HealthMash have more numerous and finer classes within their databases. This is because these websites are general health information providers and therefore need to encompass all possible health information into their design, whereas the ontological design of present research is concerned only with immunization knowledge and does not require its information classes to be as detailed. HealthMash provides both English and Simplified Chinese semantic-based search, but such search results contain little Taiwanese information that maybe not easy to intuitively comprehend for a Traditional Chinese user. While there are a number of online translation tools that can translate between English and Traditional Chinese, the integrity and accuracy of such translations are not guaranteed. In comparison, our search system displays its semantic structure in Traditional Chinese and allows searches in both Traditional Chinese and English, meeting the needs of the Taiwanese general public.

The current research has three main limitations. First, concerning semantic-based search design, while the research design considered the relationship between vaccines and related diseases, these relationship sets may restrict information that may otherwise have been identified, as mentioned previously. In addition, the ontological design of the present research is based on Traditional Chinese and English, and therefore information written in other languages is excluded from the search. Lastly, concerning pre-processing of search phrases, the current system places emphasis on its semantic-based search design and did not target structural analysis of lengthy search phrases, the assumption being that users will mostly likely type in common terms. If search phrases are lengthy or overly complicated, the system may not identify a synonym or related term in the ontological structure. This may reduce the semantic-based search capacity.

Future semantic-based search functions should utilize more polished pre-processing tools, and incorporate user feedback on individual research results as a parameter when determining the listing order of search results, in order to improve search quality. But in the long term, a key task is the maintenance and continued renewal of the ontological structure. Aside from the need for collaboration between field specialists and information engineers, processing runtime, cost, and comprehensiveness of resources are all factors that should be considered.

## Acknowledgments

The authors thank the National Science Council of Taiwan for supporting this research under the grant of NSC 101-2221-E-010-007-.

## References

- [1] Centers for Disease Control and Prevention Morbidity and Mortality Weekly Report. Impact of vaccines universally recommended for children--united states, 1900-1998. *J Am Med Assoc* 1999; 281(16): 1482-1483.
- [2] Heinonen S, Silvennoinen H, Lehtinen P, Vainionpaa R, Ziegler T, and Heikkinen T. Effectiveness of inactivated influenza vaccine in children aged 9 months to 3 years: An observational cohort study. *Lancet Infect Dis* 2011; 11(1): 23-29.
- [3] Borrás E, Dominguez A, Fuentes M, Batalla J, Cardenosa N, and Plasencia A. Parental knowledge of paediatric vaccination. *BMC Public Health* 2009; 9(154).
- [4] Antoniou G and van Harmelen F. *A semantic web primer*. 2nd ed. Cambridge: MIT Press., 2008.
- [5] Rajan JM and Lakshmi MD. Ontology-based semantic search engine for healthcare services. *IJCSE* 2012; 4: 589-594.
- [6] WebMD LLC (2012). Webmd vaccine tracker. Available: <https://vaccinetracker.webmd.com/Reminders.aspx>
- [7] WebLib (2011). Next generation search and knowledge discovery tools for the web. Available: <http://weblib.com/products/healthmash>
- [8] Morville P and Callender J. *Search patterns: Design for discovery*. Sebastopol: O'Reilly Media, 2010.
- [9] Obrst L. Ontological architectures. In: Poli R, Healy M, and Kameas A, eds. *Theory and applications of ontology: Computer applications*. Dordrecht: Springer, 2010.
- [10] Brusa G, Caliusco M, and Chiotti O, A process for building a domain ontology: An experience in developing a government budgetary ontology, presented at the Second Australasian Ontology Workshop (AOW 2006), Hobart, Australia, 2006: pp. 7-15.
- [11] Jimeno-Yepes A, Jimenez-Ruiz E, Berlanga-Llavori R, and Rebholz-Schuhmann D. Reuse of terminological resources for efficient ontological engineering in life sciences. *BMC Bioinformatics* 2009; 10 Suppl 10: p. S4.
- [12] National Centers for Biomedical Computing (2011). Bioportal. Available: <http://bioportal.bioontology.org/>
- [13] WebLib (2011). Healthmash: The best semantic health search engine and health knowledge base. Available: <http://healthmash.com/>

## Address for correspondence

Der-Ming Liou, Tel: +886-2-28267187, Email: [dmliou@ym.edu.tw](mailto:dmliou@ym.edu.tw)  
Address: No.155, Sec.2, Linong Street, Beitou District, Taipei City, 112 Taiwan (R.O.C.)