

Association for Information Systems

AIS Electronic Library (AISeL)

ACIS 2013 Proceedings

Australasian (ACIS)

2013

Development of User Warrant Ontology for Improving Online Health Information Provision

Arya Prakasa

Monash University, appra3@student.monash.edu

Daswin De Silva

Monash University, daswin.desilva@monash.edu

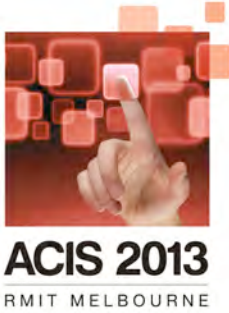
Follow this and additional works at: <https://aisel.aisnet.org/acis2013>

Recommended Citation

Prakasa, Arya and De Silva, Daswin, "Development of User Warrant Ontology for Improving Online Health Information Provision" (2013). *ACIS 2013 Proceedings*. 120.

<https://aisel.aisnet.org/acis2013/120>

This material is brought to you by the Australasian (ACIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ACIS 2013 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.



Information Systems: Transforming the Future

**24th Australasian Conference on Information
Systems, 4-6 December 2013, Melbourne**

Proudly sponsored by



Development of User Warrant Ontology For Improving Online Health Information Provision

Arya Prakasa
Daswin De Silva
Centre for Organisational and Social Informatics
Faculty of IT
Monash University, Australia
Email: daswin.desilva@monash.edu

Abstract

Health information portals (HIP) are gateways to reliable and personalised online health information. In practice, however, searching for information in HIP is still far from effective due to the intricate nature of health information provision. Previous studies have shown the emerging trend of using domain ontology to address the retrieval issue in online healthcare information. Yet, the suitability of domain ontology alone for HIPs is still questionable due to the varied levels of user behaviour and preferences in information search. Motivated by this problem, we propose an ontology development method grounded on the collaboration between user warrant principles, knowledge engineering, and design science framework. The paper reports the development method and the implementation of such an user-warrant ontology that accommodates user-sensitivity into HIP. The evaluation process is conducted by domain experts responsible for portal management by validating the external semantic of the ontology according to a set of pre-defined evaluation criteria. Results from the application of this methodology to an actual HIP are also reported as this research demonstrates the potential of user warrant ontology to resolve information retrieval problem in HIP.

Keywords

User-warrant Ontology, Ontology Development Methodology, Health Information Portals, User-centred Information Provision.

INTRODUCTION

Health information portal (HIP) is an internet based system which aims to assist user to find reliable and useful online healthcare information (Fisher et al. 2009). HIP act as an intermediary between the chaotic world of online health information and its user by evaluating the quality of information, and classify the information according to its user preferences (Luo and Najdawi 2004; Suominen et al. 2009). To be able to achieve its objectives, HIP should provide a set of mechanisms to enable user to find and retrieve relevant information easily. Such mechanism, however, is not an easy task to provide due to the complex nature of online health information, as well as varied level of users' information seeking skills (Fisher et al. 2007).

Semantic web is claimed to be beneficial in supporting HIP due to its capability to support aggregation, integration, and presentation of information, with ontology as the main pillar (McGuinness et al. 2012). The need for further research and development for ontology to support information retrieval system in HIP is well recognized (Hyvönen et al. 2007; Fisher et al. 2007; McGuinness et al. 2012). Suominen et al. (2009) argue ontology for online health information provision need to be flexible enough to cover a large number of possible subjects within the portal's domain of knowledge. In practice, most research and development on ontology in healthcare domain are expansion or adaptation from the existing scientific ontologies/controlled vocabularies (d'Aquin and Noy 2011). However, the suitability of scientific ontology for HIP is still questionable as they are designed for general user with diverse knowledge and information seeking skills (Burstein et al. 2005; Fisher et al. 2007).

Development of ontology for HIP requires a holistic approach which is able to ensure the quality, comprehensiveness, and relevance of the ontology to the users of the portal. Motivated by the lack of research on ontology methodology for HIP, in this paper we propose an ontology development method grounded on the collaboration between user warrants, knowledge engineering and design science framework. The research aims to redefine, implement, and evaluate the construction of domain ontology in the context of Breast Cancer Knowledge Online (BCKOnline) systems, an Australian health information portal specializing in woman breast

cancer. The portal's user-sensitive approach was identified as a suitable case study to develop ontology with user warrant principles.

The next section reviews health information portals, the key issues in information search and retrieval followed by the case study on BCKOnline. Section three discusses ontology development approaches with specific focus on the user warrant approach. Section four present the research design followed by section five focusing on the development of a user-warrant ontology for the BCKOnline portal. Finally, section six concludes the paper.

HEALTH INFORMATION PORTALS

HIPs have become imperative in addressing the popularity of the Internet as a source of online health and medical information. Hopkins and Fogg (2002) state the various levels of health information, wide selection of options, and convenient access to the information as the three valuable points of internet healthcare information. Despite its potential benefits, consumer reliance on internet healthcare information has raised concerns among health practitioners regarding the quality of the information, accuracy, completeness, and currency in comparison to the current medical research (Hopkins and Fogg 2002; van den Haak and van Hooijdonk 2010; Bundorf et al. 2006). Other important issues are the enormous quantity of healthcare information available on the Internet, and the different levels of information seeking skills among internet users (Zeng et al. 2003).

Suominen et al. (2009) point out the importance of ontology as one of the main components of HIP in supporting semantic portal. The author further specifies ontology roles as to describe all relevant concepts in application domain of the portal, as well as describing document subject matters, content genres, and target audiences. Bhojaraju and Buck (2007) emphasize on the two benefits of ontologies for information portal. The benefits are: 1) to facilitate appropriate information delivery, and 2) to provide efficient search functionality for users. Ontologies offer multidimensional searching which empower users to narrow and/or expand the query. Such feature is possible due to the nature of relationships between concepts in ontology. The effective search functionality supported by ontology will enables user to retrieve relevant information with least effort.

Case Study of Breast Cancer Knowledge Online (BCKOnline)

BCKOnline is an Australian user-centred personalised HIP for breast cancer health information. The portal consists of quality-assured health information, evaluated by a panel of domain experts, for a diverse range of users (Fisher et al. 2007; Xie and Burstein 2011). The portal is specifically designed to provide efficient and effective search functionality which can be personalized according to user preferences (Burstein et al. 2005; Evan et al. 2009). BCKOnline is based on user-sensitive research design, incorporating methodologies from various disciplines within interpretivist framework (McKemish et al. 2009). The user-sensitivity of the research design is reflected in the BCKOnline's six iterative research processes and based on user needs and values.

Nguyen et al. (2011) research finding on BCKOnline taxonomy usage problems shows the lack of content for specific information need and indexing problems as the most significant content issues (25.96% and 36.49%), while the use of short and simple strategies (26.36%), low persistence (36%), and limiting search with narrow topic/options (27.37%) are the three significant findings on user problems. The research highlights problems relating to the mismatch between portal content and user needs. The study also highlights the use of three different vocabularies for single portal as inefficient due to the simple structure and the lack of functionality of the controlled vocabularies in the portal. The authors state the need for better glossary interface which can encourage consistency and capability to shows the relationships using semantic web technologies. This research reveals the needs to accommodate user-generated terms and common query faults for indexing, which demonstrate the potential of developing ontology based on user-generated terms to overcome the mismatch between user and system terminologies.

ONTOLOGY DEVELOPMENT

There are a wide number of studies proposing a variety of approaches in developing ontology (Chorco et al. 2003). Fernández-López's (1999) state the need to unify the approaches to present the most representative methodologies. The author further mentions the possibility to apply IEEE standard 1074-1995 (software development process) to ontology development. The authors argue ontologies are part of a software product, thus ontology should be developed according to the software standard in general. Yun et al. (2011) state the reliance of ontology engineering to IEEE 1074-2006 standards follows the common practice of evaluation in knowledge

engineering as the superset of ontology engineering. The authors argue ontology engineering as a relatively immature discipline and in need of further exploration and refinement.

Fernández-López's (1999) study proposes METHONTOLOGY which combines SKELETAL methodology by Uschold (1998) with TOVE methodology proposed by Gruniger and Fox (1995). Fernández-López's (1999) claims METHONTOLOGY as the most representative method for ontology construction which able to comply with IEEE standard 1074-1995. Recent study conducted by Yun et al. (2011) proposed a knowledge engineering approach to build domain ontology. The proposed frameworks update and expand the previous study conducted by Fernández-López (1999) based on IEEE standard 1074-2006 and design ontology criteria proposed by Gruber (1993). The methodology consists of six phases which are: 1) purpose and requirement identification, 2) ontology acquisition, 3) ontology implementation, 4) evaluation/check, 5) confirmed formal ontology, and 6) documentation.

Ontology Development and Warrant Principles

The process of selecting and expressing terminology in a controlled vocabulary is based on the concept of warrant, which is also the fundamental principle that guides terminological choices (Wallace 2007). The concept of warrant is closely related with semantic theories which render the methodological implication for determining meaning and relationships in semantic tools such as thesauri and semantic network (Hjørland 2007). The concept of warrants primarily derived from Thesaurus construction (Library of Congress 2007; Lancaster 1986; Hjørland 2007; Mai 2011). However, thesaurus and ontology are actually related as they both are classified as pre-coordinate indexing tools (Tomić 2008).

For the purpose of this research, there are two types of warrants that are covered in this paper. The first one is literary warrant which relies heavily on the use of terminology and concepts captured in the documents (Miwa and Kando, 2007). The well-established controlled vocabulary in medical information such Medical Subject Heading (MeSH) is constructed specifically to assist in scientific research. In this type of warrant, the ontology structure is based on main stream of scientific medical classification, and the selection of terms directed from scientific terminology found in relevant scientific literatures (Hahn et al. 1999). The use of this type of ontology is proven to be fruitful in medical scientific community as they tend to use the same terms and have a good understanding about the domain. However, implementing scientific literary-warrant terms within HIP can trigger a mismatch between system terms and user terms leading to difficulties for users to find relevant information.

While most of ontologies are built based on scientific literary-warrant principles, a few researches are proposing user warrant to accommodate user terms. Wallace (2007) states user warrants relies primarily on the language of people who are expected to use the system, which include both people who construct the systems, and people who use the controlled vocabulary. Giunchiglia and Zaihrayeu (2007) argue user warrant ontology approach as the intermediary between the two extreme types of ontology: 1) the one with high semantic expression that requires complex query language which is not really appropriate for common user, and 2) the other one which is very simple query language that almost reduces the ontology function into dictionary of synonyms used in Boolean operator. A few research studies have been conducted to analyse the possibilities to construct ontology based on user terms, all those researches share similar patterns on the use of user terms as the preferred index terms, proposed for general users, and based on creating ontology from the scratch (Miwa and Kando 2007; Weng et al. 2007; Kotis and Vouros 2006; Brewster 2002).

User warrant principle is relevant to this research because user-sensitive HIP is based on user-centred method which requires active participation of users. User warrant principles in this research uses the following three elements: 1) portal terms, in this case BCKOnline which specifically developed by the domain expert of the portal, 2) user terms which generated from the portal search engine logs, and 3) evaluation from the domain expert responsible to manage the semantic relationships of the portal. The combination between these three elements promotes the suitability of ontology terms and relationships to support the user-sensitive approach of the portal.

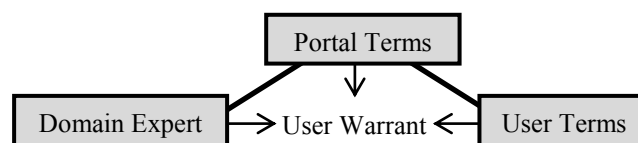


Figure 1: User warrant elements

RESEARCH DESIGN

This paper addresses the development of user warrant ontology to improve online health information provision. Development of user warrant ontology for HIP requires a comprehensive ontology development approach which comprises an ontology engineering method, ontology validation, and evaluation. The research uses design science research (DSR) as a suitable method for ontology construction (Weber 2002; Indulska and Recker 2010). The standard involves the three major processes, which are: pre-development processes, development processes and post-development processes. In this study, we modify Delir Haghighi et al.'s (2012) method on development of DO4MG and Hongyan's (2009) method on ocean domain ontology. Figure 1 below shows the proposed ontology framework for this research.

Pre-Development Stage

In this phase, we identify the scope and objectives of the ontology. The study follows Fernández-López et al. (1997) Methontology development steps which consists of four steps: i) define intended uses, scenarios of use, and end-users, ii) level of formality of (informal, semi-informal, semi-formal or rigorously formal), and iii) scope, which consist of a set of terms to be represented, its characteristics and granularity.

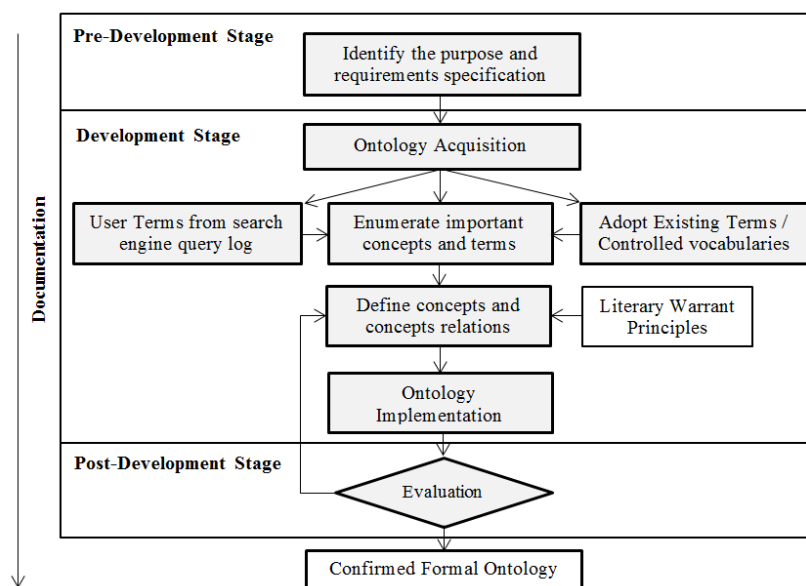


Figure 2: Framework for HIP ontology development

Development Stage

In knowledge acquisition, the research consists of five steps:

- (i) Enumerating Important Concepts and Terms: The process of enumerating important concepts and terms will be based on the topic search page on BCKOnline (<http://www.bckonline.monash.edu.au/search/patopics.do>). The terms also collected from several related websites to develop basic understanding about breast cancer.
- (ii) Adopt Existing Terms from the Portal: The second part is to adopt existing terms within the portal which consist of three controlled vocabularies: BCKOnline vocabulary, terms adopted from medical subject heading, and terms adopted from BreastCare glossary. The terms gathered from each vocabulary then sorted based on its original sources and then cleaned from duplication.
- (iii) Harvest User Terms from Search Engine Query Log: The research will follow on Jansen et al. (2000) method on web user characteristic analysis to capture highly used search terms on the portal. For the purpose of this study, the usage data analysis will focus on the analysis of user query to gain insight to the characteristic of the Web user. The data analysis and manipulation conducted in MySQL server. The step consists of identifying trends among user based on the frequency of term appearance. The next step is to

clean the data from duplication and harvest the unique term according to its occurring score. From this analysis, it is expected to gain more accurate and meaningful data for ontology terms.

- (iv) Define Concepts and the Relationships between Terms: The next step is to define concepts and the relationships between terms and organized them into hierarchy. This process is based on the result on the first process of enumeration of important concepts and terms found in BCKOnline Portal, and input from subject expert. This research adopts top-down design strategy to organized concept hierarchy.
- (v) Ontology Implementation: The aim of this step is to explicitly represent the conceptualization model in a formal language, so that it can be used by computer to process. The BCKOnline ontology is implemented in Concept Map (Cmap) which supports OWL. Terms from BCKOnline glossary will be implemented as classes to define and describe the concepts, subclasses, properties, and associated relationships of the domain interest, while the other terms such from BreastCare Victoria and MeSH and the extraction of data usage analysis will act as a complementary of BCKOnline glossary.

Post-Development Stage

The post-development stage consist of the evaluation process which need to be conducted to check whether the proposed ontology is able to meet the specification requirements. This study is limited to the development of ontology without further implementing it to the systems. Due to the research limitation, the evaluation is based on criteria-based evaluation. Criteria-based evaluation is suitable for this study because the evaluation is grounded on external semantics of the ontology which can only be done by a domain expert (Brewster et al. 2004). The criteria are limited to the characteristics of the ontology (Yu et al. 2009). The study follows six criteria-based evaluation propose by Delir Haghighi et al. (2012). The criteria are described below:

Table 1: Criteria-based evaluation propose by Delir Haghighi et al. (2012)

Clarity	Consists of three aspects: a) the ontology terms should be defined formally without subjectivity; b) the ontology needs to be documented with natural language, and c) the terms must convey 'the intended meaning' with regards to the requirements of social situations and computation rather than their context.
Consistency/ coherence	The ontology concepts and elements should have a logical consistency and avoid contradictions or ambiguity.
Conciseness	The proposed ontology should not take unnecessary concepts or redundancies.
Expendability/ extendibility	The criterion refers to the ability of ontology to extend further or to be applied to a specific application domain.
Correctness	The ontology must represent the correct modelling of the real-world concepts.
Completeness	The criterion applies to the completeness of the individual definitions of the ontology. This can be achieved through competency questions which include the queries and requirements that the ontology must be able to answer.
Coverage	The criterion refers to the completeness and coverage of terms and concepts to represent and information domain.

DEVELOPMENT OF USER-WARRANT ONTOLOGY FOR BCKONLINE

Pre-Development Stage

The pre-development stages consist of three elements: 1) define intended uses, scenarios of use, and end-users, 2) Level of formality of (informal, semi-informal, semi-formal or rigorously formal), and 3) Scope, which consist of a set of terms to be represented, its characteristics and granularity.

The intended use of the proposed ontology is to improve BCKOnline portal's retrieval systems capabilities by integrating the portal's current controlled vocabularies into single ontology based on user-warrant principles. The proposed ontology is expected to provide efficient and effective search functionality, which can be personalized according to user preferences. The ontology's scenarios of use are: 1) to assist user in finding relevant health information available in the portal by linking user terms and systems terms. 2) enhance user search experience by providing query expansion capability. 3) assist domain expert in expressing and updating subject terms consistently, and 4) assist the portal's administrator to develop the semantic relationships of the portal. The end user of the ontology is the users of the portal which consists of domain experts, systems administrator, and customer of the portal.

The ontology is developed in semi-formal language to accommodate the variance of terms in the existing controlled vocabularies which consists of user preferred terms and formal health and medical terms. The scope of terms is limited to all existing terms available in the portal, which are: MeSH to represent formal medical and health terms, BreastCare Victoria Glossary to accommodate Victorian health and medical terms, BCKO Controlled Vocabulary to accommodate terms which constructed based on user needs and analysis, and user query data to facilitate user preferred terms.

Development Stage

Enumerate important concepts and terms: After identifying scope and objective of the ontology, the next step is to conduct knowledge acquisition. The first step is to enumerate important concepts and terms. The important terms were collected from information available in topic search page (<http://www.bckonline.monash.edu.au/search/patopics.do>). The topic search page consists of two tables: 1) browse by content, and 2) list of most searched content. Data on the first table is used as input to identify the main contents of the portal. The terms were extracted as a basic glossary.

Concepts and terms are also collected from other related sources, mainly from relevant websites which provides useful information regarding breast cancer such Breast Cancer Network Australia (<http://www.bcna.org.au/>), BreastCancer.org (<http://www.breastcancer.org/>). Those websites are useful in providing basic understanding about breast cancer and other related issues. The next step is to combine the extracted terms to build a primary glossary.

Adopt existing terms: The BCKOnline systems consist of three controlled vocabularies: 1) BCKOnline glossary which is specifically developed by the portal's domain experts, 2) BreastCare Victoria which was adopted from BreastCare Victoria glossary, and 3) terms adopted from Medical Subject Heading (MeSH). The terms were gathered from each vocabulary, sorted based on its original sources and duplicates removed. From the review of the three vocabularies, it is found that the majority of terms population are from BCKOnline (335 terms or 44%), following by The Breast Care Victoria (225 terms or 29%) and MeSH (208 terms or 27%).

The next step is to classify terms from each vocabularies into a systematic scheme to enable us to conduct further analysis. The process is important due to the absence of semantic relationships between each term. From the review, it is found that all terms can be classified into four major classes: 1) breast cancer symptom and diagnosis, 2) treatment and side effects, 3) day-to-day matters, and 4) non-medical terms. Each of the major classes then further specify into more specific subclasses. The basis of this classification was the BreastCancer.org (<http://www.breastcancer.org/>) sitemap because the website structure is simple and easy to understand for common people without sufficient knowledge in breast cancer.

The four major classes are further subdivided into 22 subclasses. The aim of this classification is to gather relevant information regarding the composition of each vocabulary, and later on, enable us to clean up the duplicates. The grouping of terms conducted through analysis of the relationships shows on the portal's metadata resource description. Based on this, we can justify the relationships between subject terms and glossary terms.

Harvest user terms from search engine query log: The next part of usage data analysis focus on the analysis of user query to gain insight to the characteristic of the Web user based on Jansen, Spink and Saracevic's (2000) theory. We analysed transaction logs containing 1,942 queries posted by 548 users. The data sample taken from period: 6th of March, 2008 to 27th of August, 2009. The data analysis and manipulation was conducted in MySQL server.

Table 2: BCKOnline terms frequency

Term	Frequency	Term	Frequency	Term	Frequency
Chemotherapy	64	Diagnosis	36	Recurrent breast cancer	25
Complementary Therapy	60	Lymphedema	35	Radiotherapy	19
Surgery	59	Risk factor	30	Breast Cancer	17
Breast Reconstruction	50	Early breast cancer	28	Tamoxofin	17
Advanced Breast Cancer	38	Hormonal therapy	26	Cancer	13

Occurrences: We constructed a rank-frequency table for all 1,942 queries. From there, we sort the data with frequency variant from 64 to 5. Based on that, we could identify 30 terms with the most high frequency.

Chemotherapy, complementary therapy, surgery and breast reconstruction are the four terms with the highest frequency ranging from 64 to 50. The second most used terms are advanced breast cancer, diagnosis, lymphedema, and risk with the frequency ranging from 38 to 30 (Table 2).

The next step is to clean the data from duplication and harvest the unique term according to its occurring score. From this analysis, it is expected to gain more accurate and meaningful data for ontology terms.

Table 3: BCKOnline unique terms

Measure	Original	Cleaned	Percentage
Total Terms	1,942	450	23,17%
Unique Terms	450	138	30,6%
Terms occurring once	436	138	31,65%
Terms occurring more than 50	3	0	–

The data from rank-frequency table below then cleaned from any duplication to gain the actual number of unique term. From 1,942 queries, only 450 are actually unique. The result then further compared to the portal's existing vocabularies gathered from the previous step. The result shows only 138 terms from 450 terms are unique. Thus, the unique terms that can be utilize from user queries are 138 from the total original 1,942 queries (7%).

The data finding shows most terms with the highest queries are terms that are already covered in the existing portal's vocabulary. While most unique terms appear less than 5 times. The terms are suitable to be included to the system despite of its low frequency. This decision is taken to accommodate the potential user needs in the future. The filtering process is conducted twice, first, is to eliminate the duplication within the log data itself, and second, to eliminate terms that already covered in the existing controlled vocabularies. Based on this step, we were able to gather 138 unique terms as the source of terms for the ontology.

Define concepts and the relationships between terms: The main goal of this step is to create a set of concepts categorization, and organize these concepts into hierarchical structure. There are four terms gathered from the development stages which are: BCKOnline glossary, Breastcare Victoria, MeSH, and extraction of user query log. Terms from BCKOnline glossary in particular, resemble user warrant approach due to the fact that the glossary was constructed specifically to support information retrieval of the portal content. The terms also closely maintained by the portal's domain expert. Based on above consideration, BCKOnline glossary is implemented as main concept (classes) of the proposed ontology to define and describe the subclasses, properties, and associated relationships of the domain interest, while the other terms such from BreastCare Victoria and MeSH and the extraction of user query log will act as the complementary.

Ontology Implementation: The BCKOnline ontology is implemented in Cmap which support OWL (Web Ontology Language), an ontology language widely used to define and describe classes, subclasses, properties and its association within a domain interest. The ontology construction process is conducted through top-down design strategy. By communicating with subject expert of the portal, we developed a seven level of concept hierarchy for breast cancer ontology. Here, we describe the top two level concept hierarchies.

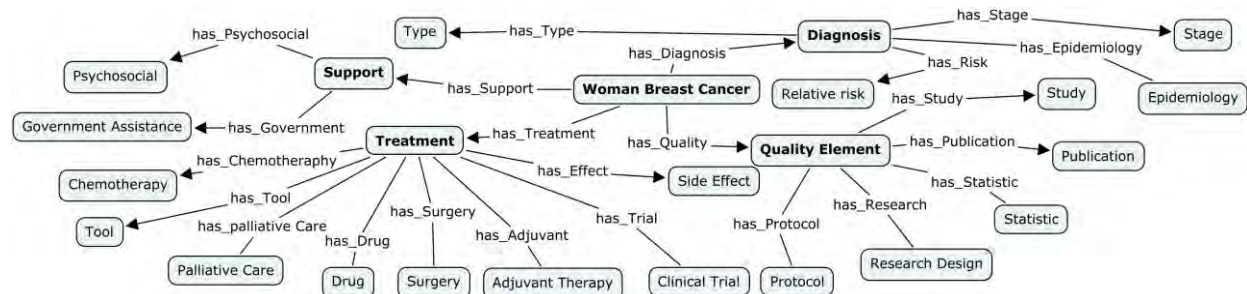


Figure 3 : Top level of the breast cancer ontology

The top two level of the BCKOnline ontology is illustrated in Figure 3 as a combination of nodes and arrows which represent class/concept and its property/relation. The top-class in the ontology is woman breast cancer which specifies the scope of terms contained within the ontology. The concept is further divided into four aspects consisting of the following properties: First, the *Treatment* subclass refers to all procedures or applications that are intended to relieve illness or injury resulting from developing breast cancer. This class consists of eight subclasses. Second, the *Diagnosis* subclass refers to a statement about what disease someone has, based on the type of examination procedure. This class includes the type, stage, epidemiology, and relative

risk. Third, the *Support* subclass refers to any kind of non-medical support from community and government. The support subclass consists of two sub-subclasses, which consists of government assistance, and psychosocial. The last subclass is *Quality Element*. The subclass refers to all criteria to evaluate the quality of information within BCKOnline. This subclass is further divided into five sub-subclasses.

Post-Development Stages

Ontology Evaluation: This paper follows six criteria-based evaluation propose by Delir Haghighi (2012). The criteria are described below:

- (i) Clarity. There are three requirements to be addressed: a) the ontology terms should be defined formally without subjectivity; b) the ontology needs to be documented with natural language, and c) the terms must convey 'the intended meaning' with regards to the requirements of social situations and computation rather than their context.
 - (a) For the first context of formal definition without subjectivity, the ontology terms are generated from three existing controlled vocabularies which almost all the terms is already pre-defined prior ontology construction process. Terms from BCKOnline glossary are specifically defined by domain expert of the portal, while terms from BreastCare Victoria and MeSH contains formal definition from its core organization (BreastCare Vicotoria and Library of Congress). Terms generated from user query log analysis are defined based on its context and relationships in ontology.
 - (b) The ontology is documented using natural language; this is aligning with the proposed ontology development methodology which contains a documentation element during the whole development processes.
 - (c) The proposed ontology is based on literary warrant approach, where one of the components is evaluation by domain expert. The BCKOnline domain expert was invited to provide some input and clarification regarding the scope, terms selection, and its relationships. The evaluation results were being implemented to clarify and refine the ontology.

Example: During interview process with domain expert, the proposed four main classes actually consist of support, diagnosis, treatment, and research. The 'research' class was proposed due to the high number of terms related to research activities, such as: publication of results, statistics etc. the subject expert clarify the intended meaning of such terms as quality elements encoding scheme. We then modify and update the term accordingly.

- (ii) Consistency/coherence. The ontology concepts and elements should have a logical consistency and avoid contradictions or ambiguity. The ontology has undergone several consistency checks with domain expert, and some terms have been updated to accommodate user's input.

Example: Previously, the term 'psychosocial' and 'government support' was separated as different subclasses. However, both terms and its subclasses actually narrowly related. We then merge both terms under 'support'.

- (iii) Conciseness. The conciseness criterion means that ontology should not take unnecessary concepts or redundancies. This concept has been carefully considered through defined concepts and relationships during the development processes. However, crosscheck with the domain expert shows some terms tend to be general. For example term "protocol", the domain expert was unsure about the use of the terms in HIP, and the term able to bring multiple perception such 'protocol' for 'medical treatment', 'research', or 'trial'. We then decided that the term actually related to 'clinical 'trial'.
- (iv) Expendability/extendibility. The criterion refers to the ability of ontology to extend further or to be applied to a specific application domain. The proposed ontology for breast cancer was built as a pilot project for further improvement and further expansion. There are many concepts within breast cancer ontology that can be further develop for any domain ontology, such as palliative care, quality elements, diagnosis, psychosocial, and government support.
- (v) Correctness. The ontology must represent the correct modelling of the real-world concepts. The ontology was evaluated by the domain expert to ensure the suitability of the ontology in terms selection and relationships.

Example:

The term ‘palliative care’ previously was grouped under ‘psychosocial’; this concept was incorrect due to the scope of ‘palliative care’ is closer to medical care.

- (vi) Completeness. The criterion applies to the completeness of the individual definitions of the ontology. This can be achieved through competency questions which include the queries and requirements that the ontology must be able to answer. Below is an example of competency questions that were answered by the proposed ontology.

Table 4 : Competency Questions Test

Competency Questions	Concept	Relations
What kind of support Victorian government could deliver for people who live in rural area?	Patient accommodation and travel scheme	hasProgram
How help support group could help me?	Social support Advocacy Self-help group	hasSupport

Ontology Documentation: The whole process of ontology development needs to be documented clearly according to the intended user. In this case, subject experts and systems administrator of the portal. The documentation consists of a formal OWL representation of the ontology, and ontology development manual for further improvement and refinement.

DISCUSSION AND CONCLUSION

HIPs provide personalised and quality assessed health information to target audiences. Information retrieval in HIP is still ineffective due to the use of scientific/medical ontology structure which leads to a mismatch between user and systems terms which further decrease the likelihood of finding relevant information. This research learns the possibilities of how ontology can be built by integrating the existing controlled vocabularies, and user terms to construct ontology based on user warrants principle. User warrant principle is suitable for user-sensitive HIP due to its ability to elaborate users’ preferences for ontology construction which is not highly covered in the existing literature. The research design consists of three major stages, which are: pre-development processes, development processes and post-development processes. The methodology comprising three elements: 1) of literary warrant, 2) knowledge engineering framework which is based on IEEE 1074-2005/2006, and 3) design science. The evaluation process is conducted by domain expert of the portal, following a pre-defined criteria-based evaluation approach. Such implementation demonstrates the potential of using ontology for resolving information retrieval problem in HIP through accommodating user terms and establishing consistency of terms selection. This study is a pilot project for BCKOnline system improvement and is anticipated to be extended in the future. Future research includes the implementation of the user warrant ontology within the portal and integrating contextual factors in its evaluation with actual end users.

REFERENCES

- Bhojaraju, G., and Buck, S. 2007. “Ontologies in Portal Design,” In: *Encyclopedia of Portal Technologies and Applications*, A. Tatnall (Eds.). PA: Hershey.
- Brewster, C., Ciravegna, F., and Wilks, Y. 2002. “User-Centred Ontology Learning for Knowledge Management,” In: *Proceedings of the 6th International Conference on Applications of Natural Language to Information Systems-Revised Papers* (NLDB '02), Springer-Verlag, London, UK: pp 203-207.
- Bundorf, M.K., Wagner, T.H., Singer, S.J., and Baker, L.C. 2006. “Who searches the internet for health information?” *Health Service Res* (41:3 Pt 1), pp 819–36.
- Burstein, F., Fisher, J., McKemmish, S., Manaszewicz, R., and Malhotra, P. 2005. “User Centred Quality Health Information Provision: Benefits and Challenges,” In: *Proceedings of the Proceedings of the 38th Annual Hawaii International Conference*.
- Chandrasekaran, B., Josephson J.R., and Benjamins V.R. 1999. “What are ontologies, and why do we need them,” *IEEE Intelligence Systems* (14:1), pp 20-26.
- Corcho, O., Fernández-López, M., and Gómez-Pérez, A. 2003. “Methodologies, tools and languages for building ontologies: where is their meeting point?” *Data Knowledge Engineering* (46:1), pp 41-64.
- d’Aquin, M., and Noy, N.F. 2011. “Where to publish and find ontologies? A survey of ontology libraries,” *Web Semantics: Science, Services and Agents on the World Wide Web* (11), pp 96-111.

- Delir Haghighi, P., Burstein, F., Zaslavsky, A., and Arbon, P. 2013. "Development and evaluation of ontology for intelligent decision support in medical emergency management for mass gatherings," *Decision Support Systems* (54:2), pp 1192-1204.
- Evans, J., Manaszewicz, R., and Xie, J. 2009. "The role of domain expertise in smart, user-sensitive, health information portals," In: *Proceedings of the 42nd Annual Hawai'i International Conference on System Sciences*, IEEE Computer Society, Los Alamitos, CA, USA, pp 1-10.
- Fernandez-Lopez, M., Gomez-Perez, A. and Juristo, N. 1997. "METHONTOLOGY: from Ontological Art towards Ontological Engineering," In: *Proceedings of the AAAI97*, pp 33-40.
- Fernandez-Lopez, M. 1999. "Overview of methodologies for building ontologies," In: *Proceedings of the IJCAI-99 workshop on Ontologies and Problem-Solving Methods: Lessons learned and future trends*. CEUR, Aachen, Germany.
- Fisher, J., Burstein, F., Lynch, K., Lazarenko, K., and McKemmish, S. 2007. "Health information websites: Is the health consumer being well-served?" In: *AMCIS 2007 Proceedings*, Keystone, Colorado.
- Fisher, J., Burstein, F., Manaszewicz, R., and Lazarenko, K. 2009. "Health information portals: How can we improve the user's search experience?" In: *ECIS 2009 Proceedings*, Paper 167.
- Giunchiglia, F. and Zaihrayeu, I. 2007. "Lightweight ontologies," In: *Technical Report DIT-07-071*. University of Trento, Italy.
- Gruber, T. R. 1993. "A translation approach to portable ontology specifications," *Knowledge acquisition* (5:2), pp 199-220.
- Gruninger, M., and Fox, M. S. 1995. "Methodology for the Design and Evaluation of Ontologies," Paper presented at the *Workshop on Basic Ontological Issues in Knowledge Sharing*, IJCAI-95, Montreal, QC, Canada.
- Hahn, U., Romacker, M., and Schulz S. 1999. "How knowledge drives understanding: matching medical ontologies with the needs of medical language processing," *Artificial Intelligence* (1), pp 25-51.
- Hjørland, B. 2007. "Semantics and knowledge organization," *Annual Review of Information Science and Technology* (41:1), pp 367-405.
- Hongyan, Y. U. N. 2009. "Research on building ocean domain ontology," In: *Computer Science and Engineering WCSE'09 Second International Workshop* (1), pp 146-150.
- Hopkins, H. and Fogg, S. 2002. "Assessing health information on the internet," *The Australian Health Consumer* (2), pp 21-23.
- Hyvönen, E., Viljanen, K., and Suominen, O. 2007. „HealthFinland—Finnish health information on the semantic web," In *The Semantic Web* (pp. 778-791). Springer Berlin Heidelberg.
- Indulska, M., and Recker, J. 2008. "Design science in IS research: a literature analysis," Paper presented at the *proceedings 4th biennial ANU workshop on information systems foundations*. Canberra, Australia.
- Jansen, B. J., Spink, A. H., and Saracevic, T. 2000. "Real life, real users, and real needs: a study and analysis of user queries on the web. *Information Processing and Management*," (36:2), pp 207-227.
- Kotis, K. and Vouros, A. 2006. "Human-centred ontology engineering: The HCOME methodology," *Knowledge Information Systems* (10: 1) pp 109-131.
- Lancaster, F. W. 1986. *Vocabulary Control for Information Retrieval*, 2nd ed. Arlington, VA: Information Resources Press.
- Library of Congress. 2007. "Pre- vs. post-coordination and related issues". Retrieved 23 July, 2013, from www.loc.gov/catdir/cpsd/pre_vs_post.pdf.
- Luo, W. and Najdawi, M. 2004. "Trust-building measures: a review of consumer health portals". *ACM* (47:1), pp 108-113.
- Mai, J.E. 2011. "Folksonomies and the new order: authority in the digital disorder," *Knowledge organization* (38:2), pp 114-122.
- McGuinness, D. Shaikh, A., Lebo, T., Ding, L., Courtney, P., McCusker, J., Moser, Morgan, G.D., Tatalovich, Z., Willis, G., Contractor, N., and Hesse, B. 2012. "Towards Semantically-Enabled Next Generation

- Community Health Information Portals: The PopSciGrid Pilot,” In: *Proceedings of Hawaii International Conference on System Sciences*.
- McKemmish, S., Manaszewicz, R., Burstein, F., and Fisher, J. 2009. “Consumer empowerment through metadata-based information quality reporting: the breast cancer knowledge online portal,” *Journal of the American society for information science and technology* (60:9), pp 1792-1807.
- Miwa, M. and Kando, N. 2007. “A naïve ontology for concepts of time and space for searching and learning,” *Information Research* (12:2), paper 296.
- Nguyen, B.V., Burstein, F., Fisher, J., and Wilson, C. 2011. “Taxonomy of usage problems for improving user-centric online health information provision,” In: *AMCIS 2011 Proceedings - All Submissions*, Paper 65.
- Suominen, O., Hyvönen, E., Viljanen, K., and Hukka, E. 2009. „HealthFinland—A national semantic publishing network and portal for health information,” *Web Semantics: Science, Services and Agents on the World Wide Web* (7:4), pp 287-297.
- Tomić.T. 2008. “Thesauri or ontologies? or both? a comparison between two kinds of subject heading systems with regard to their enhancement of effective information retrieval,” *Masteruppsats, 30 högskolepoäng, vt Institutionen för ABM Uppsatserinombiblioteks & informationsvetenskap*, nr 431.
- Uschold, M. 1998. “The Enterprise Ontology,” *The Knowledge Engineering Review* (13:1). doi: 10.1017/S0269888998001088.
- van den Haak, M. and van Hooijdonk, C. 2010. “Evaluating consumer health information websites: The importance of collecting observational, user-driven data,” *Professional Communication Conference (IPCC)*, IEEE International.
- Wallace, D.P. 2007. *Knowledge Management: Historical and Cross-Disciplinary Themes*. Westport, CT: Libraries Unlimited.
- Weber, R. 2002. “Ontological issues in accounting information systems,” In: V. Arnold and S. Sutton (Eds.), *Researching Accounting as an Information Systems Discipline* (pp. 13-33) Sarasota, Florida: American Accounting Association.
- Weng, C., Gennari, J.H. and Fridsma, D.B. 2007. “User-centred semantic harmonization: A case study,” *Journal of Biomedical Informatics* (40:3), pp 353-364. on System Sciences (HICSS'05) - Track 6 - Volume 06 (HICSS '05), IEEE Computer Society, Washington, DC, USA, 138.3-.
- Xie, J., and Burstein, F. 2011. “Using Machine Learning to Support Resource Quality Assessment: An Adaptive Attribute-Based Approach for Health Information Portals,” In: J. Xu, G. Yu, S. Zhou and R. Unland (Eds.), *Database Systems for Adanced Applications* (Vol. 6637, pp. 526-537): Springer Berlin Heidelberg.
- Yu, J., Thom, J.A., and Tam, A. 2009. “Requirements-oriented methodology for evaluating ontologies,” *Journal Information Systems* (34:8), pp 766-791.
- Yun, H., Xu, J., Xiong, J., and Wei, M. 2011. “A Knowledge Engineering Approach to Develop Domain Ontology,” *International Journal of Distance Education Technologies* (9:1), pp 57-71.
- Zeng, Q.T., Kogan, S., Plovnick, R.M., Crowell, J., Lacroix, E.M., and Greenes, R.A. 2004. “Positive attitudes and failed queries: an exploration of the conundrums of consumer health information retrieval,” *International Journal of Medical Information* (73:1), pp 45-55.

ACKNOWLEDGEMENTS

This research is partly supported by the University of Ballarat 'Self-sustaining Regions Research and Innovation Initiative', an Australian Government Collaborative Research Network (CRN) and the Telematics Trust, project: Breast Cancer Knowledge Online (BCKOnline) Portal.

COPYRIGHT

Arya Prakasa and Daswin De Silva © 2013. The authors assign to ACIS and educational and non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive licence to ACIS to publish this document in full in the Conference Papers and Proceedings. Those documents may be published on the World Wide Web, CD-ROM, in printed form, and on mirror sites on the World Wide Web. Any other usage is prohibited without the express permission of the authors.