A Patient Profile Ontology in the Heterogeneous Domain of Complex and Chronic Health Conditions

Tara Sampalli
Doctoral Candidate
Faculty of Computer Science
Dalhousie University, Canada
tara.sampalli@cdha.nshealth.ca

Michael Shepherd, Ph.D.
Faculty of Computer Science
Dalhousie University, Canada
shepherd@cs.dal.ca

Jack Duffy, Ph.D.
Faculty of Computer Science
Dalhousie University, Canada
jack.duffy@dal.ca

Abstract

There is growing interest in recent years in applying ontologies to represent disease concepts because they have the ability to depict the domain knowledge with a superior level of expressiveness and precision. Ontologies have been predominantly used to represent well-categorized disease concepts. However, there are challenges in representing the domain knowledge for heterogeneous and poorly categorized systems. In this study, a methodology to create an ontology to represent the domain knowledge for complex and chronic health conditions is explored. The domain of complex chronic conditions can be viewed not only as heterogeneous but also as dynamic with new knowledge continually evolving. The methodology includes the development of a controlled vocabulary to create the first layer of semantic interoperability. The controlled vocabulary is then converted into a patient profile ontology to add deeper semantics, conceptually and relationally in the heterogeneous domain knowledge.

1. Introduction

There is growing evidence demonstrating the increasing burden of complex or medically unexplained health conditions on the health care system. The burden on the health system is due to lack of standardized treatment options, lack of standardized tests to confirm diagnosis leading to repetitive or unnecessary tests for patients and limited understanding in the medical community [1, 2]. Conditions such as Multiple Chemical Sensitivity, Chronic Fatigue Syndrome, Fibromyalgia, Chronic Pain Syndrome, and Irritable Bowel Syndrome are often grouped in the category of medically unexplained or complex conditions in the literature. Primary care is seeing an increasing number of patients who cannot be diagnosed, assessed or referred to specialists for treatments in a timely and cost-effective manner [3-6]. Multidisciplinary teams of clinicians involved in the management of these conditions need to work in a collaborative care management system with providers getting involved at various points in care [7-9]. However, there are many factors that can create a barrier in the collaborative management of patients [10, 11]. Causal factors could include limited documentation on these conditions, unclear etiology or standardized treatment strategies, and poor categorization of the domain knowledge. Much of the knowledge is evolving in an ad hoc manner among small expert groups. Despite these challenges, there is a clear need for health professionals to share information in a timely and accurate manner in order to enhance patient care experience and reduce medical errors [12-14]. There is a further need to represent the domain knowledge in a sharable format so interdisciplinary knowledge can grow among multiple users [15-17].

In this study, a methodology is presented to develop an ontology in a heterogeneous knowledge domain. An ontology was developed in the patient profile domain of a complex chronic health condition. A controlled vocabulary was developed as a first step to generating a community’s perspectives on the domain knowledge. A patient profile ontology was then developed from the concepts in the controlled vocabulary to provide a deeper level of semantics to the concepts in the domain.

The rest of the paper is organized as follows: Section 2 presents an overview of the literature review is presented. Section 3 discusses the methodology of the study which includes the scope of the ontology, a brief description of the domain, development of the controlled vocabulary and development of the profile ontology. Section 4 presents a summary of the results. Sections 5 and 6 are devoted to the discussion and conclusions of this work.
2. Literature review

An ontology approach to capturing domain knowledge is explored as a possible modus operandi to organizing knowledge in a heterogeneous knowledge base [18]. Ontologies have gained importance in recent years as a knowledge management platform in many areas including healthcare. Ontologies are preferred to conventional classifications due to the higher level of expressiveness that is possible in describing concepts and their relationships [19, 20]. Hine [21] discusses the use of ontologies in integrating and standardizing experimental data from sources in the genetic, cognitive and neuroanatomical domains, with the Gene Ontology being the most prominent among them. The Open Biomedical Ontologies website contains an extensive collection of open biomedical ontologies and related information [22]. The challenges outlined by Hine include building a common and formalized knowledge framework that can be used across different disciplines and communities of practice. Dominigue et al. [19] identify the key requirements for an ontology approach to knowledge management as a community’s perspectives being stable on an issue with “well defined roles”, “specified criteria” and “codified procedures”.

A study by Lin et al. [23] examines the potential to create a common understanding among a group of mental health professionals through an ontology approach. This study outlines the importance and the challenges of building ontologies in heterogeneous knowledge systems. The primary interaction in this study is between a domain expert and an ontology engineer. The study highlights the challenges that exist in the knowledge capture for a domain that has obscure definitions, unstructured data, inconsistent use of vocabulary and assessment scales, and emerging knowledge with time. Despite these challenges, the authors identify the need to enable the domain knowledge in a machine retrievable format to facilitate efficient information retrieval. The study identifies the importance of developing structure and consistency in such situations as a preliminary step to developing the domain ontology.

In heterogeneous knowledge domains, it may be necessary to build a layer of standardization and consistency prior to building a descriptive knowledge of the domain. In a study by Qin and Paling [24], the development of knowledge models in environmental science and ecological science in the collaborative work of a community of scientists and information managers is presented. This study outlines a method to transfer a stable and well categorized [19] controlled vocabulary into an ontology that can introduce a higher level of enrichment and expressiveness to the domain knowledge. Digital objects which are concepts in the Gateway to Educational Materials vocabulary are multi-dimensional in their characteristics. The authors suggest that a conventional cataloguing code will be inadequate to describe crucial and essential details as many of these elements do not even exist in the vocabulary. An ontology was thus created from the controlled vocabulary to provide a higher level of expressiveness to the semantics in the controlled vocabulary. When an information repository has heterogeneous digital objects, interoperability becomes the first obstacle to providing access to these resources.

The primary objective of this research was to develop a model for the development of a patient profile ontology in the heterogeneous domain of complex and chronic health conditions. The proposed model is described in the next section. Section 4 describes the application of the model to develop an ontology in the domain of a complex health condition. Section 5 presents results from this study followed by discussions and conclusion in section 6.

3. Model development for ontology in a heterogeneous domain

The generic framework for the development of an ontology can be divided into three important phases: a specification phase, a conceptualization phase, and an implementation phase [26-28]. Figure 1 is a depiction of the generic steps and the variations that are involved in the development of an ontology in a heterogeneous domain such as a controlled vocabulary and involvement of domain experts.

Figure 1. Model for the development of ontology in heterogeneous domain
The specification phase is to acquire knowledge of the domain. This can be achieved by identifying the primary knowledge sources in the domain, which in this case are the patient charts and the domain experts, i.e., the clinicians.

The goal of the conceptualization phase is to organize and structure the knowledge such as outlining the key classes and relations among these classes in the ontology. The implementation phase involves verifying the ontology through consistency check and user feedback.

The model proposed in this study uses the standardized framework for the development of an ontology as outlined above but with special considerations for organization of knowledge in a heterogeneous domain. The specific steps developed to address this in our design have been explained in the subsequent subsections.

3.1 Specification

3.1.1 Goal and scope of the ontology. Establishing the goal and scope was the first step in developing the ontology in the heterogeneous domain of complex health conditions. The goal of an ontology determines the overall objective for developing an ontology. The goal of the profile ontology developed in this study is to create a comprehensive hierarchical controlled vocabulary and a representation of the multidisciplinary and multidimensional relationships that exist among the concepts in the vocabulary. The scope helps identify the key elements that need to be included in the ontology. The explicit knowledge in the domain for these conditions is limited and the implicit knowledge is very vast. Therefore the scope of the ontology for this research was maintained in the first layer of organization, that is, in the domain of patient profiles. A patient profile or problem list generation is typically the starting point of a care management scheme for patients. The elements in this ontology were maintained on the knowledge that existed in this domain.

3.1.2 Domain description. Complex chronic conditions typically have symptoms that are attributed to a multitude of contributing factors such as physical and psychosocial [29]. Multidisciplinary care teams have come to the forefront as an effective management strategy for these conditions. The variety and recurring nature of symptoms in patients with complex chronic illness has led to the belief that management may benefit from the insights of different bodies of knowledge such as medical, nursing, occupational therapy, psychology, and nutrition [9]. Patient profile categorization is therefore more detailed than the typical problem lists that are generated for well categorized medical conditions [29]. Given the involvement of multiple health care disciplines in the care management scheme, the domain contains multidisciplinary concepts. Patient profile knowledge for most health conditions exists primarily in the intake assessments conducted by clinicians.

3.1.3 Organization of domain knowledge. The domain knowledge is unstructured and poorly categorized in the medical records [31]. There is also a significant amount of knowledge that exists as tacit knowledge among experts. Knowledge in this domain is heterogeneous and dynamic as outlined in the previous section.

It is thus essential to first generate operational knowledge in the patient profile domain. This process will include developing a layer of consistency to the concepts frequently used in the domain [19] [24]. The controlled vocabulary developed in this study will help achieve this layer of consistency. Furthermore, it is essential to develop a level of standardization to the concepts in the controlled vocabulary so the knowledge can be shared among multiple users crossing communities of practice.

3.2 Conceptualization

The conceptualization of the ontology commences with the concepts that exist in the patient profile domain for complex health conditions. As identified in the previous section, it is essential to create a standardized controlled vocabulary that contains the key concepts that exist in this domain that following which the relations between these concepts were established. As the knowledge in this domain is still evolving among domain experts, it is essential to involve the domain experts in the conceptualization phase. A key phase of the conceptualization phase is the development of the standardized controlled vocabulary for complex health conditions. The development of the controlled vocabulary includes reviewing patient charts, compiling a list of routinely used terms in the patient profile domain with the help of domain experts, standardization of the controlled vocabulary and review of the vocabulary by the domain experts. The creation of the controlled vocabulary will help in the development of top level classes of the ontology. The relations and attributes are then developed with the help of knowledge in the patient charts and domain experts. Described in the next few sections are the key phases involved in the development of the controlled vocabulary. Figure 2 shows the schematic for the development of the controlled vocabulary. The subsequent sub-sections
provide explanation for the components of the schematic.

3.2.1 Review of patient charts. A retrospective audit of patient charts is conducted to identify recurring concepts and terminologies used by the multidisciplinary team of clinicians caring for patients. The identified chart audit terms are grouped under the relevant areas of health focus and by the frequency of occurrence. A review is done by a multidisciplinary team of clinicians to validate the retrieved terms for accuracy and relevancy.

3.2.2 Standardization of terms using systematized nomenclature of medicine – clinical terms, SNOMED CT®. SNOMED CT® is used as a reference terminology to standardize the terms retrieved in the chart audit process. The standardization of the chart audit concepts includes the following steps.

i. A clinical term is identified for standardization from the chart audit process when the term is a recurring term used to describe a patient profile.

ii. A manual search for an identical match to the source term is made using a SNOMED CT® browser.

iii. A search is made for alternative terms / synonyms when an exact match is not found with the same clinical meaning (i.e. concept match) to the source term.

iv. No match terms are identified.

The multidisciplinary team will review the standardized terms for accuracy and completeness.

3.2.3 Re-coding of patient profiles using the controlled vocabulary. A controlled vocabulary with the relevant grouping of standardized concepts in the patient profile domain is created. The multidisciplinary team of clinicians re-codes 3 prototypical patient charts using the new vocabulary. The clinicians use a web-based form to generate the profiles for patients using the controlled vocabulary. The domain experts test the accuracy, completeness and relevancy of the standardized concepts included in the controlled vocabulary.

3.3 Implementation of the ontology

Protégé 3.4.2 is used to implement the patient profile ontology. The profile ontology will be exported into the Web Ontology Language (OWL). The profile ontology presents the formalized description of concepts in the domain of complex health conditions. It includes basic concepts and properties that characterize the profile of patients from multiple areas of health. The information for the ontology is developed from the knowledge contained in the controlled vocabulary.

A consistency check of the classes in the ontology is conducted. Consistency checking helps detect classes that cannot have instances. The ontology evaluation also includes the review by domain experts for accuracy and completeness of the knowledge represented in the ontology.

4. Methodology

A complex and chronic health condition was selected to test the viability of the proposed methodology for a profile ontology in a heterogeneous knowledge system. Multiple Chemical Sensitivity (MCS) is a chronic condition that affects multiple body systems. There is limited knowledge in the literature and limited acceptance of this condition in the medical system. There is a lack of readily available clinical practice guidelines and standardized treatment strategies. Due to these challenges, patients spend a lot of time trying to obtain a diagnosis with repetitive medical tests and various modalities of treatment thus increasing the burden of costs on the healthcare system. Furthermore, there is yet to be a common consensus on suitable diagnostic criteria, symptom profile or treatment strategies among experts in the field. However, the importance of a multidisciplinary team approach is gaining importance as an effective way to manage patients with these conditions.

The domain knowledge for this condition is considered to be heterogeneous as it is unstructured, multidisciplinary, existing as tacit knowledge and
fragmented. This health condition is thus considered to be suitable to test the proposed model to organize knowledge using ontology approach in a heterogeneous knowledge domain. This study has used a convenience sample approach in testing the model.

A convenience sample of 9 multidisciplinary clinicians and 100 patients participated in the study to develop the controlled vocabulary and the profile ontology in the patient profile domain for MCS.

The creation of the controlled vocabulary included auditing patient charts, interviewing clinicians, and a review of terminology by clinicians for relevancy and accuracy in the domain of patient profile knowledge for MCS. The source terms were then standardized using SNOMED CT® [34] as the reference terminology. The domain experts were also involved in reviewing the standardized vocabulary to ensure accuracy and completeness of the concepts.

5. Results

The results section has been organized under the categories of specification, conceptualization and implementation phases to show the development of profile ontology in the heterogeneous domain for MCS.

5.1 Outcome of the specification phase

5.1.1 Scope of the ontology. A total of five hundred and twelve concepts were retrieved relevant to the patient profile domain for MCS from the retrospective chart audit process. The concepts were grouped under five major areas of health focus: medical, physical, psychosocial, rehabilitation and nutrition. The number of concepts in the medical, physical, psychosocial, nutrition and rehabilitation areas of health care focus was 356, 136, 122, 80 and 118 respectively. Four hundred and twenty two of these concepts (82%) were standardized using SNOMED CT. However, some key concepts related to Multiple Chemical Sensitivity were absent in the reference terminology.

Post co-ordination was applied to improve the availability of necessary concepts for the health condition. For instance, the term Multiple Chemical Sensitivity does not exist in SNOMED CT. Seventy five percent of the missing concepts were included this way (n = 67). The remaining terminologies of relevance (n = 23) were maintained in the controlled vocabulary as determined by the domain experts. The clinicians tested the usefulness of the vocabulary by applying it to recode the profiles for three patients with a diagnosis of MCS. Eighty percent of the clinicians agreed on the overall usefulness of organizing the domain knowledge through a controlled vocabulary.

5.1.2 Description and organization of the knowledge in the domain. An important consideration of this phase was to develop consistency in the description of the knowledge in the domain. This was achieved through the development of the controlled vocabulary. The standardization offered another layer of clarity to assist in the organization of the knowledge in the domain. For instance the concept “Light sensitivity” can be interpreted as sensory intolerance or pain in the eye. The intended clinical meaning relates to sensory intolerance which can now be tagged using a consistent terminology and a unique concept ID.

Table 1. Example of standardization of concepts

<table>
<thead>
<tr>
<th>Terminologies in clinical notes</th>
<th>SNOMED CT® concept</th>
<th>Concept ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fatigue, low energy, very tired, extremely tired, heavy feeling</td>
<td>Fatigue</td>
<td>84229001</td>
</tr>
<tr>
<td>Light sensitivity, hypersensitivity to light, intolerance to light</td>
<td>Light intolerance</td>
<td>62481005</td>
</tr>
<tr>
<td>Fibromyalgia, FM, Myalgia</td>
<td>Fibromyositis</td>
<td>24693007</td>
</tr>
</tbody>
</table>

Table 1 shows examples of the inconsistencies in the use of terminologies found in the patient charts, the corresponding standardized terms and their unique concept IDs.

The controlled vocabulary was organized in a hierarchical structure with concepts being presented the way they were depicted or captured in the patient charts by the various health care disciplines. Figure 3 shows the controlled vocabulary categorization in the domain of psychosocial health focus.
5.2 Outcome of the conceptualization phase

Protégé 3.4.2 [36] was used to develop the ontology. The ontology presents a detailed taxonomic overview of the complex health condition domain. The ontology contained 220 classes describing the profile concepts for the condition of MCS. At the basic level there are five relevant super-classes under the primary areas of health focus identified for the condition of MCS: Medical, Physical, Psychosocial, Rehabilitation and Nutrition.

Figure 4 presents the Protégé tool displaying the relevant super classes from the Complex Conditions ontology.

The profile ontology contains standardized expression of concepts created in the controlled vocabulary to explicate the intended clinical meaning as seen by experts in the domain. Standardized concepts are specified with their SNOMED CT ID number (Concept Unique Identifier) and with a list of synonyms. Class Fatigue has a SNOMED CT concept ID of 84229001 with parent concept being Energy and Stamina and synonyms Weariness and Tiredness (Figure 5).

Examples of more intricate concepts that benefit from standardization and consistency relevant to this health condition include heightened visual perception, heightened auditory perception, emotional hypersensitivity, impairment of balance, emotional regulation or emotional state finding and hypervigilant behaviour.

Individuals or instances are used in the profile ontology to present list of concrete concepts of relevance for each class. For example, the ontology contains 100 individuals with MCS in the class Organization. A patient has data properties such as education, marital status and object properties such as HasProfile and HasOrganization that link it to other classes in the ontology.

The ontology also contains instances of 100 profiles for the class Patient Profile. The instances in profiles show the multifaceted nature of symptoms as substantiated under each area of health focus that exist in the domain of a complex patient (Figure 7).

The properties in the ontology introduce relations among concepts. A patient HasOrganization and the organization is inversely linked to the class Patient by HasPatient. The class Profile is linked to the class Management Scheme by property hasCollaborativeManagement. The class Psychosocial Profile is linked to the management scheme by property Management Required which has individual dietitian_refferal or physician_refferal.

The profile ontology includes definitions of over 70 properties, data and object properties.

Jambalaya, a Protégé plug-in, was used to visualize the knowledge created in the ontology [39]. Clinicians can use the visual platform to browse and query the
knowledge in the domain, such as the nature of multidisciplinary interactions in the categorization of a symptom profile. For instance, the clinicians can view in a graphical interface the way a particular symptom or area of concern may play out as the focus of care for multiple disciplines. Bullying identified in the super class Education and school finding in SNOMED CT® can be a consideration for a psychologist and a rehabilitation coordinator.

---

**Figure 6. Education and school finding using Jambalaya**

![Figure 6. Education and school finding using Jambalaya](image)

**Pain symptom** as presented in the patient charts has been viewed in the patient charts by a psychotherapist, physician or physiotherapist from various angles of importance such as pattern of pain, anatomical site or in relation to the pain threshold. Figure 7 shows the view of a patient profile shown using OwlSight browser [37].

---

**Figure 7. Display of a patient profile in the ontology**

---

### 5.3 Implementation phase – next steps

Protégé 3.4.2 was used to do the consistency check for the ontology. There were no inconsistencies in the profile ontology.

The next steps of the ontology evaluation will include the domain experts checking the ontology for accuracy and completeness of knowledge representation. The clinicians will also explore the multidisciplinary interactions and multidimensional concepts available in this comprehensive ontology. A user friendly interface will be developed for the clinicians to browse the ontology as the existing platforms such as Jambalaya [38] and OwlSight [37] may prove to be complicated for navigation and browsing.

At least two clinicians in each of the six health disciplines that were involved in the creation of the controlled vocabulary will be recruited to participate in this segment of the study (n=12). Clinicians will use the user friendly browser to navigate through the ontology and will have the option to query the multidisciplinary interactions and the patient profile knowledge that exists in the populated instances of the ontology. Clinicians will then use a survey questionnaire to provide feedback on the usefulness of the ontology.

The next steps of the validation of the ontology will also include applying the model on one other chronic and complex health condition, Chronic Pain [8] which satisfies the essential characteristics for a
heterogeneous knowledge domain considered in this domain.

6. Discussion

Providing timely and effective care to individuals who present with medically unexplained symptoms is an important problem in the field of medicine. There is significant time spent by medical practitioners due to the limited knowledge of the conditions and ineffective communication systems [3]. This in turn can leave the patients feeling lost in the health system.

Developing knowledge systems that can be shared is an important advancement in the collaborative care environment for these conditions. The ontology approach brings the advantage of flexibility and semantic interoperability to the shared knowledge on these conditions [18]. The ontology developed in this study adds another layer of expressiveness and clarity to the concepts in the controlled vocabulary in terms of the level of abstraction and relationships that are defined among the concepts in the domain.

The methodology described in this paper is a key contribution in the development of ontologies in heterogeneous systems. The knowledge had to be extracted from inconsistent, poorly categorized and unstructured patient information similar to the challenges outlined by Lin et al. [23]. Developing a controlled vocabulary and standardization of the terms in the controlled vocabulary added a layer of stability to the knowledge in the domain. Furthermore, the standardization of the concepts in the controlled vocabulary is seen as a stable method to facilitate the contribution of new knowledge by multiple users of the ontology [24].

As discussed in the literature [39, 40], in order to facilitate the use of technological advances in medicine it is important to engage the end users and experts in the development phase of the work. User involvement thus formed an important aspect of this methodology to ensure accuracy and completeness of the controlled vocabulary. This stability of the knowledge in the domain thus allowed for the development of the profile ontology.

The development of the profile ontology in this study was divided into three phases: specification, conceptualization and implementation. These are the steps outlined in the METHONTOLOGY framework [26-28]. The phases constitute an interactive framework by which the knowledge in the domain is formulated and validated. In this paper the specification and conceptualization phases of the ontology have been presented.

A key limitation of this work is the use of a convenience sample of patients and the domain experts. This raises the question of the variation in knowledge that can exist among other expert groups which may contradict the opinion and the knowledge of the group used in this study. However, an ontology can reach a wider audience and has been deliberately selected to explicate the knowledge of lesser known and complex health conditions. Ontologies provide a semantically interoperable format for collaborative sharing of knowledge across communities of practice. The ontology has the potential to get richer as more users contribute new knowledge and as more patient instances are populated in the ontology.

6. Future work

A model for development of an ontology to organize the knowledge in the heterogeneous domain of a complex and chronic health condition has been explored. The next step of this study is the implementation phase of the profile ontology. Validation of the methodology using another complex chronic health condition, chronic pain will be conducted. Future work around the evaluation of the usefulness of the ontology in enhancing the understanding and communication in the collaborative care environment for complex conditions is planned.

7. References


