K. S. Raghavan (Indian Statistical Institute, Bangalore)
C. Sajana (Indian Statistical Institute, Bangalore)

NeurOn
Modeling ontology for neurosurgery

Abstract
Patient records constitute an important source of valuable information for health-care personnel. Unfortunately, however, in most health-care institutions these are not indexed and organized to support decision-making. A patient record is complex and has data on so many parameters that conventional methods are not capable of fully exploiting the information and knowledge in these. This paper reports the initial results of an ongoing experiment to build an ontology based on concepts extracted from the patient records in the domain of neurosurgery of a large hospital. Results suggest that the ontology could prove to be an acceptable way of providing a decision support system for health care personnel.

1: Introduction
Health care and delivery is a knowledge intensive activity. Appropriate and relevant knowledge made available in the right form could prove to be a crucial factor in decision-making. In the present day context, recorded knowledge has become a fluid mix of scholarly papers, new experiences, contextual information, and expert insights collectively providing a framework for decision-making. The domain knowledge in health care gets continuously enriched by the experiences of practitioners. Much of this knowledge and new experience get recorded in patient records and may not even appear in published sources or become a part of standard texts until years later. The quality of health care could be enhanced if health care personnel are provided speedy access to patient records with defined similarities to the problem on hand. Similarities could be in terms of one or more of a very wide range of parameters and could include, for example, diseases, symptoms, drugs administered, therapeutic techniques, diagnostic techniques, age, gender, etc. Three principal factors limit the retrieval capability of conventional information systems:

- a patient record contains data on so many parameters that it is difficult to fully and adequately represent the knowledge (in a patient record) using traditional knowledge organization tools such as controlled vocabularies;
- classification schemes and thesauri recognize and represent only a limited number of relations and it is not desirable to reduce and represent the entire range of relations between concepts in a complex domain such as health care and delivery to these limited number of relations;
- current search technologies employed in controlled-vocabulary and free-text based retrieval systems require construction of complex Boolean queries and work well only when the user knows how the information in the target documents is worded and so typically users spend a lot of time browsing irrelevant information, or face the frustration of finding nothing at all (Dahlgren 2007).

Ontologies are being seen as effective tools for comprehensively representing domain knowledge as they not only specify concepts but also the relationships between them at the required level of specificity. Ontology is a term borrowed from philosophy, where an ontology is a systematic account of existence (Gruber 1993). An ontology can also
include if-then rules to support inferencing. This paper reports the initial results of an ongoing project to design and implement an ontology as a decision-support system for the Neurosurgery unit of a hospital in Chennai, India.

**2: Building the ontology**

There are two principal approaches to ontology construction; the first approach is largely manual while the second relies on machine learning and automated language-processing techniques. Ontology tools such as Kaon, Protégé, and Chimaera use the first approach (Navigli 2003 cited in Na & Neoh 2008, 32-46). This project employed Protégé. The domain concepts were derived from patient records. Each patient had been diagnosed for some neurological disorder and was given specific treatment. The patient records also contained information about the symptoms displayed, disease name (final diagnosis) and other associated illnesses. The general information included name of the patient, a unique patient ID number, doctor’s name, disease index number, gender of the patient, age, age range, consciousness level, visual acuity details, etc. It was realized that each one of these factors could be an important parameter in information retrieval.

**2.1: Domain concepts**

Neurosurgery is focused on treating diseases of the central and peripheral nervous systems and spinal column requiring and amenable to surgical intervention. Neurosurgical problems include primarily disorders of brain, spinal cord, vertebral column and peripheral nerve. An examination of the records suggested that there were four broad types of concepts. The nature of domain concepts and the processes are shown in Figure 1.

![Domain concepts and the processes](image)

*Fig. 1: Domain concepts and the processes*

These concepts could be grouped under the 3 top-level categories of S.R. Ranganathan viz. Personality, Matter Property and Energy, as shown in Table 1. The principal classes in each of the three broad categories are (the sub-classes are not exhaustive):
Table 1: Concept categories

<table>
<thead>
<tr>
<th>Personality</th>
<th>Property</th>
<th>Energy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients</td>
<td>- Age</td>
<td>- Diagnosis (using) Methods / Tests</td>
</tr>
<tr>
<td></td>
<td>- Sex</td>
<td>- Treatment (using) Radiation Therapy</td>
</tr>
<tr>
<td>Body Organs</td>
<td>- Brain</td>
<td>- (using) drugs Chemotherapy</td>
</tr>
<tr>
<td></td>
<td>- Spinal cord</td>
<td>- Surgery</td>
</tr>
<tr>
<td>Health Care Personnel</td>
<td>- Physicians</td>
<td>- Method / Technique X</td>
</tr>
<tr>
<td>Doctors</td>
<td>- Surgeons</td>
<td>- Physiotherapy</td>
</tr>
<tr>
<td></td>
<td>- Anesthetists</td>
<td>-</td>
</tr>
<tr>
<td>Para medics</td>
<td>- Nurses</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>- Lab Assistants</td>
<td>-</td>
</tr>
<tr>
<td>Institutions</td>
<td>- Hospitals</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>- OT</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>- Wards</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>- ICU</td>
<td>-</td>
</tr>
<tr>
<td>Hospital Managers / Administrators</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.2: Query library

While a clear picture of the nature of queries is possible only when the system is implemented, it was important to have an idea of the potential queries that the decision-support system was expected to respond to. A small query library was built that served as the basis for defining the classes and properties in the ontology being built.

3: The ontology

The ontology is implemented in Protégé, a frame-based open source ontology editor developed by the Stanford Center for Biomedical Informatics Research. Protégé ontologies can be exported into a variety of formats including RDF(S), OWL and XML schema. The decision-support system proposed here is conceived to have at least two major components:

- the ontology as part of the search interface to support inferencing; and
- a database of patient records linked to related records in other databases, e.g. of images (scans, X-rays, etc).

There is no one perfect way of building an ontology. Fundamentally, any ontology is a semantic network of concepts grouped into classes and subclasses and linked by means of a well defined set of relations. Ontologies also have a set of rules that support inferencing. In general decisions regarding classes, subclasses, and relations are made on the basis of what best meets the task on hand. Instances are distinguished from classes. For example, while patient is a class, an individual patient is an instance of the class. The decisions regarding how detailed the classes and/or relations should be is largely based on the purpose and ease of maintenance. In the first phase of development of the present ontology, only concepts related to the domain of medicine – neurological disorders including associated symptoms and characteristics and their treatment – and those related to patients populate the ontology. Other sub-domain concepts will be added at a later stage. This also limits the number and nature of relations that the ontology has to handle in the first phase of development of the decision support system.
3.1: Classes and subclasses

The taxonomic structure (see Table 1) gives an idea of the classes and subclasses; however, the structure merely reflects hierarchical relations or “is a” relations. The ontology is also required to specify more complex relations between concepts. It is useful to illustrate this with a few examples. The central theme of all patient records ran thus:

![Diagram](image)

**Fig. 2: Classes and subclasses**

The faceted approach was adopted in view of its powerful features in breaking complex concepts into elementary concepts and allowing for more powerful searches at the time of retrieval. Decisions regarding the relations were largely based on purpose and convenience of handling. To give an example, the concept female patients could be seen as a subclass of patients; it could also be handled by seeing it as: patient <has gender> and by fixing acceptable values for gender. The relevant concepts found in the hospital’s patient records were gathered and grouped into broad classes: Body Structure [P], Patient [P], Health-Care Personnel [P], Clinical Findings [M] and Procedure [E]. The relations between the concepts in these classes could be inferred from the classes that were related.

- Patient <has age> e.g. 45 years
- Patient <has gender> e.g. female
- Patient <has blood group> e.g. B positive
- Patient <has disease> Disease name
- Body Structure <has disease> Disease name
- Body Structure <has part> Sub-organ
- Sub-organ <is part of> Body Structure
- Patient <has symptoms> e.g. nausea, headache, vision impairment
- Disease <has symptoms> e.g. nausea, headache, vision impairment
- Disease <has recommended action> Treatment/Therapy

3.2: Terminology

When the terms were structured into a hierarchy, it was realized that the terms used by the health care personnel did not display the required degree of consistency. SNOMED (Systematized Nomenclature of Medicine-Clinical Terms) was used to build the hierarchy of domain concepts as well as to standardize the terminology. This is a comprehensive clinical terminology, originally created by the College of American Pathologists (CAP) and, as of April 2007, owned, maintained, and distributed by the International Health Terminology Standards Development Organization (IHTSDO), a not-for-profit agency in Denmark. However, the terms assigned by the healthcare personnel and the corresponding MeSH terms were also built into the ontology as relations to allow search using these terms also:
3.3: Properties and instances
Properties are referred to as *slots* in Protégé. There are two different types of slots:

- **Object property**
- **Data-type property**

An *object property* relates instances of two classes, while a *data-type property* specifies an attribute for a class instance and associates the class instance with a value for that attribute (e.g., integer or character string). Examples of data type property for a person are name, age, and date of birth. Thus the properties (relations) are used to relate instances, rather than to specify abstract relations between classes (Khoo et al. 2007).

Another type of property that we come across is *annotation property*. Annotation properties can be used to add information (metadata: data about data) to classes, individuals, and object/data type properties (Horridge et al. 2004). Figure 3 indicates partially the class hierarchy and also the properties in the present ontology.

4: Representation
The complexity of concept relations involved could be gauged from Figure 4 (based on a single patient record). Although the idea is to have a separate database of patient records and have the ontology as a part of the search interface, in the first phase all data about patients have been made a part of the ontology schema. Every concept has been linked to instances through properties—either object or data-type. While it is eventually the responsibility of the doctor/health care personnel to decide on a course of action, the primary objective was to develop a decision-support system that would assist the doctors (and other health-care personnel) in retrieving and accessing relevant patient records that might alert them to consider additional relevant factors thus contributing to enhance the quality of health care.

![Fig. 3: Class hierarchy and properties](image-url)
Many of the issues in designing an ontology relate to the design of the relation types and the trade-off between designing more specialized relations or more specialized concepts. Choices also have to be made between simpler and easy to understand representations, or more complex and detailed representations. The top level categories of the ontology are shown in figure 4. A patient is related to symptoms and diseases which are subclasses of clinical findings.

5: Issues and conclusions
This paper presents experience in building an ontology that could serve as a decision support system for the domain of neurosurgery. The present system has been developed incorporating concepts (and relations) extracted from a few hundred patient records. The system was tested with both simple and complex queries. The system responded to quite effectively for such queries as “identify & retrieve records of patients: age => 40 with astrocytoma; with brain diseases and with symptoms of nausea and pain in the neck”.

The system is being expanded and the plan of action is as below:
- To include all the patient records of the Neurosurgery Division of the hospital.
- To extend the ontology to include concepts related to healthcare institutions and personnel.
- To link relevant manuals, reference sources, text books and papers with a view to widen the knowledge base available to the users of the decision support system. This ability to link to resources outside the system to widen the knowledge base available to the end users (health-care personnel, in this case) is seen as one of the advantages of the system.
- To build rules for inferencing which could include: (1) making a reasonable and intelligent guess of the probable cause or condition of a patient based on values of certain clinical parameters, (2) making a decision regarding the possible course of action, (3) defining parameters that could generate an alert message.

While there is no question that patient records serve as an excellent source of concepts for populating domain ontology in view of the fact that they are continuously updated and contain the most current thinking of health care personnel, to be effective a decision-support system needs a wider knowledge base. Some of the issues that were encountered during the process of ontology building are not unique to this domain ontology. For example, decisions that had to be made regarding definition of classes, subclasses, properties and relation types are issues that one has to encounter in any ontology. There were, however, some other issues that had to with the domain under consideration:
- the complexity of medical terminology; non-availability of exact equivalents for some of the terms used by the physicians/surgeons (in patient records) in the standard medical terminologies (like MeSH and SNOMED CT);
- ethical issues that relate to the information/data in patient records vis-à-vis making the information data widely accessible even within the institution.

It is hoped that this domain ontology will emerge as a valuable tool to improve and enhance the quality of health care.
References
Na J.-C., Neoh H., 2008, Effectiveness of UMLS semantic network as a seed ontology for building a medical domain ontology: [editorial], Aslib proceedings, 60, n.1, p. 32-46.

Web documents have been accessed 23 September 2009.