A Consumer Lexicons Connector for the Domains of Medical Specialties

Stefano Bonacina¹, Francesco Pinciroli¹

1) Dipartimento di Bioingegneria, Politecnico di Milano

Abstract — Healthcare is a complex environment where the communication occurs among different actors, such as clinicians, doctors, nurses, and patients and their relatives. The communication deals with complex and interconnecting concepts and is often affected by uncertainty. Consequently, the sender should tailor the communication having in mind the receivers' understanding abilities. In particular, the “vocabulary gap” between language familiar to both patients and consumers and terminology used in medical practice and research influences the effectiveness of health communication. In addition, low health literacy of patients leads to discrepancy in understanding medical concepts and affects therapy adherence, even more when patients are elders. However, applications for consumer health informatics (e.g. personal health records) are on the increase. So, to present health information and medical terms to consumers and patients in a comprehensible way becomes crucial.

We designed and implemented a connector of lexicons based on ontology-like taxonomic structures we developed. It is a software tool for carrying out the matching between the medical vocabulary of the doctor and that of the consumer. It relates on several domains of medical specialties within the family: “Pregnancy and the Newborn”, “Emergency care at home”, “Pediatrics”, “Geriatrics ward (for the outpatient)”, and “Hemato-oncology”.

This education tool devoted to patients and consumers can help in bridging the “vocabulary gap” during patient-doctor interactions or communications.

I. INTRODUCTION

Recently developed patient- and consumer-centered health record systems allow the patients to participate actively in their healthcare process by collecting and managing their digital health documents and information (Kupchunas, 2007). However, the “vocabulary gap” between language familiar to both patients and consumers and terminology used in medical practice and research influences the effectiveness of health communication. In particular, limited health literacy has becoming an obstacle to overtake (Sudore and Schillinger, 2009). Health literacy has been defined as “the degree to which individuals have the capacity to obtain, process and understand basic health information and services needed to make appropriate health decisions” by the US Institute of Medicine - IOM (IOM, 2004). First, limited health literacy leads to discrepancy in understanding medical concepts and affects therapy adherence (Smith and Duman, 2009), even more when patients are elders (Kripalani et al., 2007). Second, health literacy affects both clinician-patient and clinician-parents communication for pediatric patients, as it is very important to understand and meet children needs in pediatric safety and chronic illness care (Rothman et al., 2009). Consequently, explaining parents how to care their children following the physician recommendations and instructions becomes fundamental (Sanders et al., 2009; Turner et al., 2009).

Therefore, while it is important to present health information using consumer-familiar terms in consumer applications, consistently converting medical terms to consumer-familiar ones is a challenging task (Zeng QT et al., 2007).

We designed and implemented a connector of lexicons - based on ontology-like taxonomic structures we developed - relating on the following domains of medical specialties: “Pregnancy and the Newborn”, “Pediatrics”, “Emergency care at home”, “Geriatrics ward (for the outpatient)”, and “Hemato-oncology”. The connector is a software tool for carrying out the matching between the medical vocabulary of the doctor and that of the consumer.

II. METHODS

A. The Contexts of Patient-Clinician Communication

Is a fact that, Healthcare is a complex environment where the communication occurs between different actors, such as clinicians, doctors, nurses, and patients and their relatives. The communication deals with complex and interconnecting concepts, and is often affected by uncertainty, in the sense that the education degree, domain knowledge, environment, held by the involved actors - senders and receivers - affect the message tailoring (Cantoni et al., 2008). Consequently, the sender should tailor the communication having in mind the receivers' understanding abilities. From the perspective of involved actors in the Healthcare domain, we identified some language contexts (Gould, 2008; Carthey, 2006). The language contexts (layers) form something like a hierarchy starting with the Specialist Medical Lexicon ending with the Consumer Medical Lexicon (Figure 1). Each language context (layer) receives “messages” from the layer above, adapts them depending on the layer and passes them down to the next language context (Gould, 2008; Carthey, 2006). In our approach, we considered the adaptation between the most distant layers: the “Specialist” context and the “Familiar” one. The “Familiar” context includes the lexicons related to
the patient, their relatives, and the consumer; while the “Specialist” context includes the lexicons of all the healthcare professionals, such as specialists, hospital physicians, general practitioners, and nurses. We named that adaptation “Case Conceptual Correspondence” (arrow in Figure 1).

**B. Modeling the Approach**

In this paragraph, we explained the approach we designed and followed, starting from some observations on medical dictionaries devoted to consumer. There are many biomedical dictionaries developed by professionals and devoted to professionals - all of which are gathered by the US NLM Unified Medical Language System (UMLS) (Bodenreider, 2004)-, whereas the number of the electronic medical dictionaries devoted to consumers is small (Zielstorff, 2003). One of them is the open access Consumer Health Vocabulary (CHV) (CHV initiative, 2007). It is an electronic medical dictionary devoted to consumers and whose terms and concepts are mapped to the UMLS Metathesaurus concept terms. However, the CHV is in English language only; therefore, we design a pattern to obtain the matching between the medical vocabulary of the consumer and that of the doctor in Italian language for each domain of the considered medical specialities, too.

The Figure 2 displays the conceptual approach of our project, considering the CHV vocabulary and UMLS knowledge base as available resources. “Balloon a” is the “Specialist Medical Sources/Textbooks in Italian”, i.e. specialist medical sources. The Italian Medical Specialist Lexicon (Balloon b) represents the healthcare professional’s lexicon. The Specialist Medical Lexicons in Italian have been built first extracting the terminology from some specialist textbooks on “Pregnancy and the Newborn”, “Pediatrics”, “Emergency Care at home”, “Geriatrics”, and “Hemato-oncology”, then performing a human review of the extracted terms, phrases and relationships. In addition, for the hemato-oncology domain, the review has been carried out by hemato-oncology specialists, too. Specialist lexicon is connected to the Italian Consumer Medical Lexicon (Balloon c). The Italian consumer medical lexicon has been built first, asking relatives of people acknowledged and authors, terms and concepts (for the “Pregnancy”, “Pediatrics”, “Geriatrics”, “Emergency care at home” domains); then, adapting gathered information and definitions from Italian Medical textbooks to a lower level and merging the two sets of terms.

The bolded arrow indicates the adaptation of information to a lower level for helping the consumer to comprehend (Figure 2). We named this process “Case Conceptual Correspondence”. In addition, the open access CHV (Balloon e) helped us to enrich the Italian consumer medical lexicons, by translating the original terms. At present, an Italian version of the CHV is not available (lined Balloon d), however we considered it within the conceptual schema. In fact, while the Balloon c represents the consumer/family lexicon for a single domain - coming from unstructured knowledge too -, the Balloon d has to be considered as relating to all the medical specialities, like the CHV in English language. In addition, Balloon c and Balloon d are connected each other (Figure 2). The connection starting from Balloon c and ending in balloon d describes an “Extension” action: terms originated from and used within the families are coded, organized, and added to the Italian CHV. The connection starting from balloon d and ending in balloon c describes a “Selection” action: originated from Specialist Lexicon of a Specialist Domain, terms become common within the family/consumer environment.

**C. Designing the Ontology-like Taxonomic Structures**

In the field of computer science: “An ontology defines (specifies) the concepts, relationships, and other distinctions, that are relevant for modeling a certain domain” (Gruber, 1993).

Within the development process of Information Technology projects, usually a preliminary phase concerns with whether buying something suitable available on the
shelf or building it (Ghezzi et al., 1991). Therefore, we carried out a white and grey literature researches; however, an ontology devoted to the considered domains and devoted to patients/consumer had not been developed yet. Consequently, we decided to model the application knowledge domain by developing ontology-like taxonomic structures from the beginning, following the step for building an ontology, as it follows:

1. Considering Italian Medical Specialist textbooks and the pair CHV-UMLS, in English, as specialist medical sources;
2. Apply human term recognition techniques to those sources;
3. Extract concepts, terms and their relationships;
4. Assemble those terms into ontology-like taxonomic structures considering both Italian and English.

We used the Protégé software tool (Protégé, 2000) that is a free, open source ontology editor and knowledge-base framework. The extracted concepts were grouped in two object categories within Protégé. First category is “Endurant” i.e. continuants, are “entities that endure through time while maintaining their identity and that are fully present at one single point in time” (Schulz and Hahn, 2007). The second is “Perdurant” i.e. occurrences, which are like processes, are “entities that never fully present at anyone given moment in time, but unfold themselves in successive phases, or temporal parts” (Schulz and Hahn, 2007). Examples of “Endurant” are “Body Part”, “Exam”, “Therapy” and “Pathology”. Examples of “Perdurant” are diagnostic procedures such as “Encounter” and “Exam execution”. Then, the existing relationships between concepts were defined and inserted as concept’s properties into two Protégé project files for each domain: one in Italian and another in English.

D. Designing a Relational Database and a Graphical User Interface

For each considered medical domain, we designed a relational database by Entity – Relationship (E-R) diagrams (conceptual model). The significant entities we defined are:

1. The “Familiar Medical Term” which is described by the attributes “Term Description”, “Term Definition” and “Source”;
2. The “Specialist Medical Term” which is described by the attributes “Term Description”, “Term Definition”, “Source” and “Characteristic”;
3. The “CHV Term” which is described by the attributes “Term Description”, “UMLS CUI Code” and the Boolean “Preferred” – that is “true” when a term is the preferred one to indicate a concept (Zeng QT et al., 2007) (CUI stands for Concept Unique Identifier);
4. The “UMLS Term” which is described by the attributes “Term Description”, “UMLS CUI Code”, “Term Definition” and “Source”.

To allow the storage of relationships, one or more, between concept terms, we defined E-R recursive binary relationships for the “Specialist Medical Term” and “Familiar Medical Term” entities. In addition, there are relationships among the entities to allow the construction of “Case Conceptual Correspondences”. Then, we transformed the E-R model to Structured Query Language (SQL) and implemented a database using MS Access® or MySQL database management systems; finally, we populated the database.

Finally, we designed a graphical user interface (GUI) able to make the database inquiring easy by using buttons instead of SQL scripts.

III. RESULTS

The results are presented in two sub-sections following the order of the Methods section: the first is related to the ontologies and the second to the interface of the software tool.

A. The Ontology-like Taxonomic Structures

Figure 3 displays a snapshot from the Protégé editor representing on the left, the first level of the developed ontology in English, on the right some details about the “Amnesia” concept for the Geriatrics Domain. The upper part contains the definition and the CUI code, while the lower part states that Amnesia belongs to Symptomatology class and has some relationships (for example, it is a symptom of Alzheimer’s disease).

B. The Tool and the Graphical User Interface

We developed a tool - based on a relational databases - able to show the defined correspondences between the family medical lexicon (Balloon c, Figure 2) and that of the doctor (Balloon b). In addition, the UMLS concepts and their definitions (Balloon f, Figure 2) are available for browsing. The text area ‘Object to search’ is where to insert the query string, e.g. a term, a concept, or a UMLS Concept Unique Identifier (CUI) (Figure 4). Clicking on one button...
launches the query within a specific lexicon, while clicking on one of the arrows shows the correspondences between two lexicons, e.g. from "Specialist Medical Lexicon" to "Consumer Medical Lexicon". This is what we named "Case Conceptual Correspondence".

When a term is inserted as 'Object to search' clicking on the ‘Wide search’ button returns the complete cycle of correspondences, not only the term matching between two lexicons for a specific medical domain.

IV. DISCUSSION

The built ontologies do not exhaust the considered specialist domains, and need to be updated and maintained; however, they were built considering both the medical vocabulary at the family level (i.e. the consumer as main beneficiary) and the first of well-formed ontology principles, that is “endurant”, “perdurant” as subclasses of “thing” class (Schulz and Hahn, 2007). While adopting automatic knowledge acquisition methods can decrease the costs and time of ontology development (Shamsfard et al., 2004), a human review of results must be taken into account (Brewster et al., 2009).

V. CONCLUSION

In this project, we considered the family environment that is characterized by verbal communication and not well-bounded knowledge domain in medical area. The medical language used by laypeople is very changeable and is affected by some factors: environment, skills, education. Attempting to represent such a language and matching professional medical language with it is a challenging task.

We designed and developed both ontologies devoted to specific medical domains (in Italian and in English) and a tool, based on relational databases, for carrying out the matching between the medical vocabulary of the consumer and that of the doctor. This patient education tool can help in bridging the “vocabulary gap” during patient-doctor interactions or communications.

ACKNOWLEDGMENT

The authors thank the “Division of Haemato-Oncology” of the “European Institute of Oncology” of Milan for the cooperation in developing the hemato-oncology ontology.

The authors would like to acknowledge the Biomedical Engineering students and Mathematical Engineering students of the ‘E-Health Methods and Applications’ course who contributed to develop the project.

REFERENCES

- Protégé Ontology Editor, 2000 Available: http://protege.stanford.edu/
- Zielstorff RD, Controlled vocabularies for consumer health, *Journal of Biomedical Informatics*, 36(4-5); 326-333, 2003