An ontology for computer-based decision support in rehabilitation

Laia Subirats and Luigi Ceccaroni

Barcelona Digital Technology Centre
Roc Boronat 177, 5th floor, 08018 Barcelona, Spain
{lsubirats, lceccaroni}@bdigital.org

Abstract. Although functionality and disease classifications are available thanks to initiatives such as the "international classification of functioning, disability and health", the "systematized nomenclature of medicine - clinical terms" and the "international classification of diseases", a formal model of rehabilitation interventions has not been defined yet. This model can have a fundamental role in the design of computer-based decision support in rehabilitation. Some initiatives such as the "international classification of health interventions" are in development, but their scope is overly general to cope with the specificities that characterize rehabilitation. The aim of this work is to represent knowledge in order to carry out diagnosis and personalization of activities in cases of people with functional diversity. To define the diagnosis and activity personalization, a methodology has been developed to extract standardized concepts from clinical scales and the literature.

Keywords: knowledge representation, rehabilitation, functional diversity, personalized medicine, evidence-based medicine.

1 Introduction

Until recently, progress in health and rehabilitation was hit-or-miss. We would find something without having a good understanding of how it worked. *Oh, here is an activity that improves quality of life. We have no idea why it works*. We would *discover* activities to perform desirable functions, often with many severe side effects, but we lacked the means to *design* medical interventions for a carefully targeted purpose. Such a random and irregular approach to medical discovery is typical. But now this situation is changing, and very rapidly. We have moved from the old paradigm, in which the progress in health and rehabilitation has been unpredictable, to a new era in which healthcare has now become an information technology.

Unfortunately, most healthcare practitioners actuate according to the old paradigm and still do not practice rehabilitation as an information technology; do not make maximal use of the latest rehabilitation knowledge that is already available today; and do not take full advantage of the available information and of simulation capabilities. We have the means of simulating biology, physiology and interventions on computers so that we can try out new rehabilitation interventions and drugs on simulators, a process dramatically faster than human testing. The point is that health and rehabilitation are now information technologies, and that represents a new frontier. As a result, our health technologies are subject to what is sometimes called the law of accelerating returns, an exponential improvement in the ability to understand, model, simulate and reprogram the information processes underlying disease and functioning. Yet, the way rehabilitation is currently being done is not entirely satisfactory. Therapies are based on therapists' experience and not necessarily on unbiased knowledge accumulated through exhaustive studies of a wide range of cases. This leads to frequent errors, which slow down recuperation, harm people and increase costs.

Research has been carried out to study the evolution of people undergoing neuropsychological rehabilitation. Nevertheless, in spite of their importance, psychological, social and environmental factors have been scarcely taken into account and, furthermore, a deep study of how interventions influence prognosis has not been done. The purpose of the work proposed here is to contribute to fill this gap. In particular, the objective of this work is to represent knowledge used in clinical decision support systems and study how interventions as well as the psychological, environmental and social issues influence the prognosis and the quality of life of people with functional diversity. This is done using several *artificial intelligence* (AI) techniques, with which knowledge will be used to build a decision support system in the field of rehabilitation and tools for prognosis assessment.

1.1 Computer-based decision support

Computer-based clinical decision support is an area between *health informatics* and AI. *Clinical decision support systems* (CDSSs) have a fundamental role in improving people safety and healthcare quality and efficiency (and their design and theoretical foundations are object of much research) when clinicians:

- deal with complex cases;
- are prone to making errors;
- cannot keep up with the ever increasing medical knowledge;
- deal with large numbers of routine decisions.

CDSSs are computer systems designed to impact clinician decision making (e.g., prognosis) about individual persons before, during or after (but ideally at the point in time in which) these decisions are made [1]. With the increased focus on the prevention of medical errors, *computer-based physician order entry* (CPOE) systems and CDSSs have been proposed as a key element in improving people safety [2].

There is a variety of systems that can potentially support clinical decisions. Decision support systems have been incorporated in healthcare information systems for a long time, but these systems usually have supported retrospective analyses of financial and administrative data. Recently, sophisticated data mining approaches have been proposed for similar retrospective analyses of both administrative and clinical data [3]. Although these retrospective approaches can be used to develop guidelines, critical pathways or protocols to guide decision making at the point of care, such retrospective analyses are not usually considered to be CDSSs. Perreault and Metzger [4] have described CDSSs using several dimensions. According to their framework, CDSSs differ in:

- the timing at which they provide support (before, during or after the clinical decision is made);
- how active or passive the support is, that is, whether the CDSS actively provides alerts and knowledge or passively responds to physician input or patient-specific information;
- how easy they are for clinicians to access.

Also, in principle, there are two types of clinical decisions:

- related to the diagnosis, in which computers may assist in diagnosing a disease on the basis of available patient data;
- related to the therapy, in which the best next test or therapy is determined on the basis of evidence or other knowledge.

Once large collections of patient data (e.g., patient history data, laboratory data, drug data, and patient outcomes) are available, new relations among data may be found. This will give rise to new insight and new decision rules, to be implemented in CDSSs. Most often, however, the rules to be implemented in a CDSS are derived from clinical evidence, i.e., from the medical literature and clinical experience.

CDSSs generally use one of the following paradigms to provide support: workflow-driven, production-rule—based, and predictive analytics. Some predictive-analytics systems use case-based reasoning (CBR) [5] to provide support in several stages of medical professionals' interventions [6]: diagnosis, treatment procedure, daily life management for people with chronic or degenerative conditions [7], [8], prognosis to reduce possible risks [9], and people's classification. This paper focuses on issues in clinical vocabularies, user modeling and reasoning.

1.2 Role of ontologies in clinical decision support system

In rehabilitation and related medical domains there is a lack of formalization and decisions are usually based on therapists' experience and not necessarily on evidence. Furthermore, there is a lack of interoperability among knowledge-management systems. The aim of the introduction of ontologies is to automatically share and reuse knowledge when building CDSSs. This will facilitate the adaptation of knowledge to provide evidence-based and personalized therapies, and the evaluation of the effectiveness of a treatment.

Interoperability is very important to be able to share and reuse knowledge. To provide interoperability, both the syntax of message exchange and the semantics of concepts (e.g., person's profile, scientific evidence, diagnosis, treatment, disease progression, and natural history) should be standardized. In the domain of healthcare services, to identify clinical activities and concepts, nomenclatures exist, such as: the *systematized nomenclature of medicine - clinical terms* (SNOMED CT), the *international classification of functioning, disability and health* (ICF), the *international classification of diseases* (ICD), the *international classification of health interventions* (ICHI), or the *diagnosis-related groups* (DRG) that allow classifying persons using few variables. As shown in Figure 1, mapping *electronic health records* (EHR), or *patient health records* (PHR), to terminologies and classifications helps to share and reuse knowledge in the domains of *population health*, *clinical environment*, *administration* and *report presentation*.

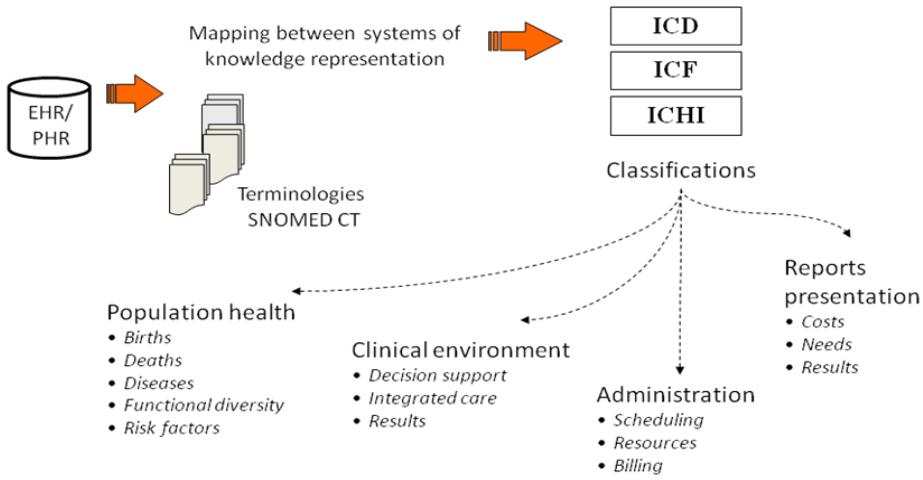

Figure 1. Role of classifications and terminologies in healthcare services

If ontologies are used for knowledge representation, the relationships in them (see sections 2.2, 3.2 and 3.3) can be exploited to automatically infer implicit knowledge. This kind of inference can be performed by reasoners such as Pellet [10]. To evaluate the consistency and check the validity of data, other ontologies can be used, such as the ICF [11], which specifies constraints for the values of concepts (classes, properties and instances). Queries and assignments in ontologies can be performed by means of *SPARQL protocol and RDF query language* (SPARQL) and Jena [12]**.**

2 Methodologies to represent knowledge

Methodologies to encode to international standards, reference models, guidelines and patterns have been designed to represent knowledge. This work is based on Beale and

Heard, who provide the reference model for the management of health information called *open*EHR [13]; and also on Cieza et al. [14], who provide a methodology to encode to the ICF standard. Regarding methodologies, Cimino [15] provides some guidelines for the design and use of medical ontologies, and differentiation between concept and context; while Sowa [16] offers best practices for building bio-ontologies and provides ontology design patterns (ODP).

To represent the status of a person in rehabilitation, indicators can be used that characterize the person's body, the rehabilitation processes and as many pathologies as possible. An *indicator* is a parameter or descriptor used to measure or compare *processes*, *results* (obtained in the execution of a rehabilitation task or activity), *body functions*, *body structures*, *activities and participation* and *environment factors*. Body functions, body structures, activities and participation and environmental factors are ICF's branches. A *process indicator* is used to assess whether a task is being performed correctly; and a *result indicator* is used to assess the performance in carrying out an activity or whether the objectives of the activity have been achieved. These indicators can be extracted from clinical scales and then encoded into international standards. If it is necessary to combine indicators and their values, a methodology might be needed to carry out this combination.

Indicators can be grouped into *core sets* to facilitate daily practice. Core sets can be formed according to functionality, pathology or rehabilitation process. (We consider four classes of rehabilitation processes: cognitive [228553007], functional [229594008], respiratory [108228000] and cardiac [313395003]). Core sets are useful because, in daily practice, clinicians and other professionals need only a fraction of the categories found in ontologies such as ICF and SNOMED CT. Several core sets already exist of different pathologies, such as multiple sclerosis, spinal cord injury or traumatic brain injury [17], however, finding the core categories for rehabilitation processes and moving from a pathology-based approach to one based on functionality and rehabilitation is needed.

A methodology to extract standard-based indicators from existing scales and parameters would include at least the following elements: a search for scales and parameters of the mentioned rehabilitation types (cognitive, functional, respiratory and cardiac rehabilitation); a prioritization and selection of scales and parameters based on literature, and coverage of rehabilitation processes and indicator types; an aggregation of indicators according to the process taxonomy.

2.1 Methodology to encode indicators into international standards

To encode indicators into international standards, ICF is considered first because its domain is closer to the one of rehabilitation and, if no category is found to define a concept, SNOMED CT is considered, which is less specific and includes top-level categories. The methodology to encode into ICF can be found in Cieza et al. [14], [18], while the methodology to translate to SNOMED CT is as follows:

1. Using any search-capable SNOMED CT interface (e.g., [19]), search for the concept that you want to encode.

   Example: "Infiltration of local anesthetic and corticoid".
2. If there is no exact match, search for a synonym.

   Example: "Infiltration of local anesthetic and corticosteroid".
3. If there are no synonyms, use a combination of hypernyms and hyponyms to find concepts that are modeled in SNOMED CT.

   Examples: "Skin infiltration of local anesthetic and steroid", "Intramuscular infiltration of local anesthetic and steroid", "Infiltration of local anesthetic and steroid to subcutaneous tissue".
4. Check if the type of the concepts found in SNOMED CT properly models the concept to be encoded. There are 19 types of concepts in SNOMED CT, such as clinical finding, physical object, social context, physical force, substance or procedure. In the previous examples, the type should be "procedure" in all cases.

2.2 Methodology to combine indicators and their values

Each indicator has a type and a value. When encoding parameters into indicators and combining several parameters into one indicator the following methodology is used:

1. <u>Type of value of the indicator</u>: If the indicator is encoded as ICF, its values are the ones specified by the ICF standard (five qualitative, ordered values, plus *not specified* and *not applicable*: 0, 1, 2, 3, 4, 8, 9). If the indicator is not encoded as ICF and there is only one parameter the indicator is derived from, the type of value is preserved. If the indicator is not encoded as ICF and there is more than one parameter the indicator is derived from and the type of these parameters is different, the type of these parameters is previously translated into five qualitative, ordered values, plus *not specified* and *not applicable*.
2. <u>Values of the indicator</u>: Depending on the specific indicator and on the type of its value, the function to aggregate values of several parameters can be the average, the maximum, the minimum or the median. For example, for *blood pressure* (b420)[1] the maximum or the minimum are used as aggregation functions, while for *dressing* (d540) average or median aggregation functions are used.

3. Representation of the health status of a person

For the representation of the health status of a person in rehabilitation, an ontology based on international standards is proposed. The person's information is based on Beale and Heard [20], who provide the reference model for the management of health information called *open*EHR [13]. External ontologies are used to define metadata [21], interventions [22], functionality [11] and diseases [23]. The proposed ontology, encoded in the OWL format [24], is summarized in the following subsections.

1 ICF and SNOMED CT codes are written in round and square brackets respectively.

3.1 Summary of the health information ontology

The ontology is composed of 77 classes. There are 2 classes in the first level of the hierarchy, 7 classes in the second, 12 classes in the third and 9 in the forth. The maximum depth is 6 and the maximum number of siblings is 9. Furthermore, there are 5 classes with a single subclass, and 8 properties.

3.2 Classes

The classes of the ontology are summarized in Figure 2. *Health information* has three subclasses: *clinical record* (referring to the past of the person), *opinion* (present) and *therapeutic process* (future). The class **clinical record** includes concepts such as:

Demographic_data (*Demographic_history_detail [302147001]*)

Observation (*Personal_health_status [405157008]*)

- Health_condition:
  - Indicator: *Body_structures (s), Body_functions (b), Activities_and_participation (d), Process [415178003]* and Result
  - *Assessment_scales [273249006]* and *Observation_ parameter [252116004]*
- Contextual_factors:
  - *Environmental_factors (e)*:
    - *Natural_environment_and_human-made_changes_to_environment (e2)*: *Air_quality (e260), Pollen_concentration [256259004], Humidity (e2251)* and *Temperature (e2250)*
    - *Residential_environment [272497004]*: *Ambulatory_care_site [35971002]*
    - *Hospital_AND/OR_institution [108343000]*: *Hospital [22232009]*
    - *Products_or_substances_of_personal_consumption (e110)*: *Drug_Aerosol [52262001], Drug_dose [398232005], Nebulizer [334947002]* and *Drugs (e1101)*
  - Personal_factors: *History_of_present_illness_section [422625006]* and *Traumatic_AND/OR_non-traumatic_injury [417163006]* with subclasses *Non-traumatic* and *Traumatic_abnormality [19130008]*
- **Therapeutic_process** *Past_history_of_procedure [416940007]*: *Activity [257733005]* and *Prescription_of_therapeutic_regimen [55053003]*

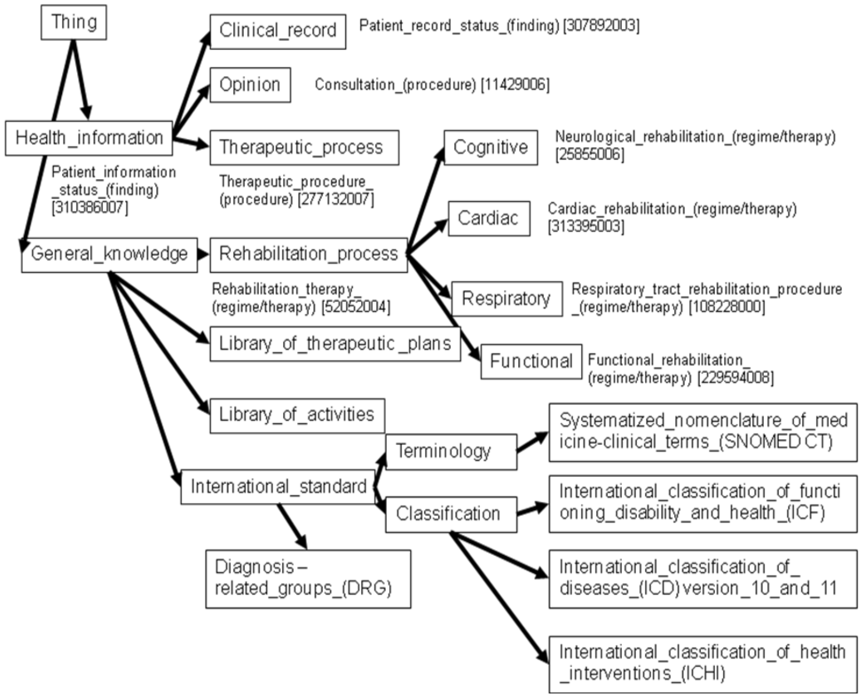

Figure 2 Class diagram of *health information*

The class **opinion** (*Consultation [11429006]*) includes concepts such as:

Assessment (*Assessment_section [424836000]*)

- *Diagnosis [439401001]: Indicator*
- *Prognosis/outlook [170967006]: Indicator*
- *Risk_factor [80943009]*: Family risk (*Familial_risk_factor [102486008]*, *Family_history_of_disorder [281666001]* and *Family_history_section [422432008]*) and Indicator

Proposal (*Plan_section [423134005]*): *Goal_context [410518001]* and *Recommendation_to [420227002]* (that has *Prescription_of_therapeutic_regimen [55053003]*)

Finally, the class **therapeutic process** includes concepts such as *Plan_section [423134005]* and *Prescription [16076005]*.

3.3 Properties

Object properties represent relationships between two individuals (classes or instances of classes). Data properties describe relationships between an individual and data values. In the ontology, some properties are semantic relations based on the current proposal of the ICD version 11 [25], for instance:

- Has_disease;
- Has_localization: e.g., an *observation* has as localization a *Body_structure*;
- *Is_manifestation_of [417318003]*: e.g., a *contextual factor* has as manifestation a *Health_condition*; a *Health_condition* is manifestation of a *contextual factor*.

Some object properties are related to activities:

- Has_recipient;
- Has_manager;
- Has_technology;
- Has_process_scale: *Assessment_scales [273249006]* and *Process [415178003]*;
- Has_result_scale: *Assessment_scales [273249006]* and *Result_comments [281296001]*.

For the main classes, the following properties are defined:

Activity_ *[257733005]*: *Medical_contraindication_(finding) [397745006]*, Recipient, Manager, *Instrument, device_ (physical object) [57134006]*, Identifier, Indications, Tasks, Title, Protocol, Scientific_evidence, *Procedure_milestone [397788003]*, Indicators (equivalent to *Indicator* that is subclass of *opinion*), *Goal_context_ (qualifier_value) [410518001]*.

Goal_context has the following sub-properties:

- *Activity_of_daily_living [129025006] (d6)*
- Participation: *Finding_related_to_ability_to_perform_community_living activities_(finding) [365341008] (d7) (d8)*
- Therapeutic: Rehabilitation of *Body_functions (b)* and *Body_structures (s)*

Demographic_history_detail *[302147001]*: *Surname_ [397678008], City_of_residence_ [433178008], Carer's_details_ [184140000], Date_of_birth_ [184099003], Patient_sex_[184100006], Patient_name_ [371484003], Social_security_number_ [398093005], Occupation_(occupation) [14679004]* and *County_of_residence_ [432407003]*.

Event *(event) [272379006]*: *Cause_of_accident_type_(qualifier_value) [278443006]* and *Origin* and *Type (attribute) [410657003]*

Therapeutic plan (*Prescription_of_therapeutic_regimen_(procedure) [55053003]*): Scene, stimulus (image, text and audio), critic point, screen and trajectory

Temporal_observable_ *[364713004]: Date_of_diagnosis_ [432213005], Date_of_onset_ [298059007], Date_of_report_ [399651003] and* Frequency.

Frequency has the subclasses of Occurrence, *Date_of_onset_ [298059007], Date_of_report_ [399651003]* and *Time_of_onset_ [263501003]*.

Data is introduced in the ontology by means of instances of classes. In Figure 3, a person is shown with user id 1, who suffers from a traumatic brain injury (TBI) and follows therapeutic plan 1. This therapeutic plan is composed of the *eating* and *dressing* activities of daily living. In particular, the *eating* activity is managed by an occupational therapist, and its result is evaluated by the *functional independence measure* (FIM).

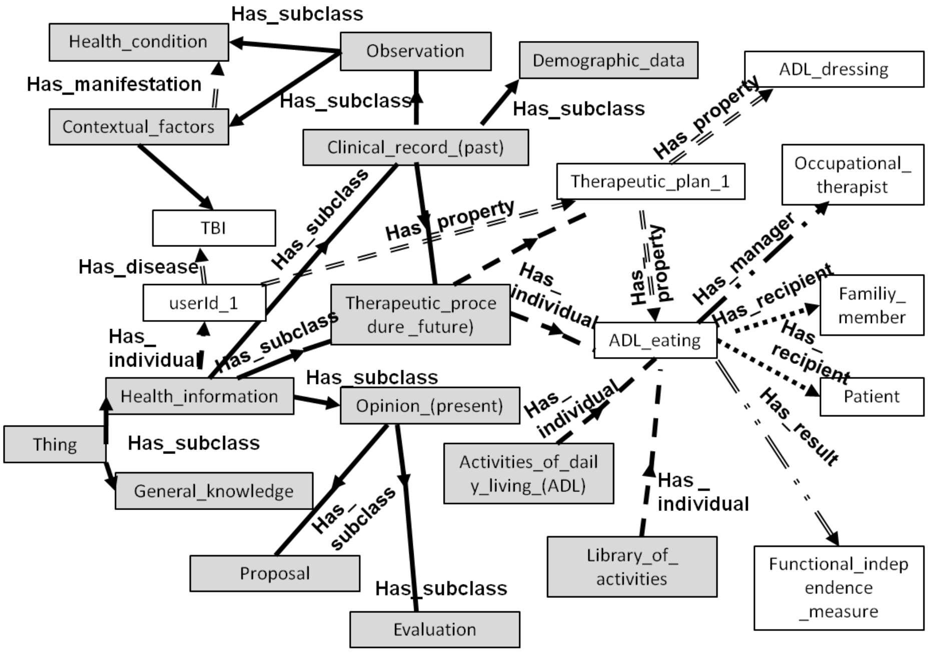

Figure 3 Summary of relationships of classes and instances

4. Conclusions and future work

The use of ontologies and international standards allows to automatically share and reuse knowledge and to build more robust clinical decision support systems, which, in turn, can provide personalized therapies and evaluate the effectiveness of a treatment. Usually, in rehabilitation, these systems have, as objectives, diagnosis and personalization of therapeutic plans.

To represent a person's medical information, a methodology has been used, according to which concepts are encoded into international standards, and indicators and their values can be combined. The representation of a person's medical information is based on the reference model called *open*EHR. Furthermore, standard ontologies are used, such as ICF, ICD and SNOMED CT.

As future work, the ontology will be extended to other types of rehabilitation and functional diversities. Furthermore, a decision support system using the ontology will be implemented and evaluated.

Acknowledgements

The research described in this paper arises from a Spanish research project called Rehabilita (Disruptive technologies for the rehabilitation of the future), which is funded by CDTI, under the CENIT program, in the framework of the Spanish government's INGENIO 2010 initiative. This work is also partly supported by the Catalonia Competitiveness Agency (ACC1Ó). The opinions expressed in this paper are those of the authors and are not necessarily those of Rehabilita project's partners, CDTI or ACC1Ó.

References

[1] Berner, E. S and La Lande, T. J., "Overview of Clinical Decision Support Systems," in *Comp. in Health Care*, Springer New York, 2007, pp. 3-22.

[2] Bates, D. W. et al., "Effect of Computerized Physician Order Entry and a Team Intervention on Prevention of Serious Medication Errors," *American Med. Association*, vol. 280, no. 15, pp. 1311-6, Oct-1998.

[3] DeGruy, K. B., "Healthcare applications of knowledge discovery in databases," *Health. Inf. Manag.*, vol. 14, no. 2, pp. 59-69, 2000.

[4] Perreault, L. E. and Metzger, J. B., "A pragmatic framework for understanding clinical decision support," *Healthcare Inf. Manag.*, vol. 13, no. 2, pp. 5-21, 1999.

[5] Aamodt, A. and Plaza, E., "Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches," *AI Commun.*, vol. 7, no. 1, pp. 39-59, 1994.

[6] Holt, A., Bichindaritz, I., Schmidt, R., and Perner, P., "Medical applications in case-based reasoning," *Knowl. Eng. Review*, vol. 20, no. 3, pp. 289-292, 2005.

[7] Marling, C.R. and Whitehouse, P.J., "Case-Based Reasoning in the Care of Alzheimer's Disease Patients," presented at the Int. Conf. of case-base reasoning, vol. 2080, pp. 702-715.

[8] Montani, S., Bellazzi, R., Portinale, L., Riva, A., and Stefanelli, M., "Multi-modal Reasoning in Diabetic Patient Management," *Art. Intell. in Med.*, vol. 1620, pp. 113-123, 1999.

[9] Palaudàries, A., Armengol, E., and Plaza, E., "Individual prognosis of diabetes long-term risks: A CBR approach," *Methods of Inf. in Med.*, 2011.

[10] Sirin, E., Parsia, B., Grau, B. C., Kalyanpur, A., and Katz, Y., "Pellet: A practical OWL-DL reasoner," *J. of Web Sem.*, vol. 5, no. 2, pp. 51-53, 2007.

[11] World Health Organization, "International Classification of Functioning, Disability and Health (ICF)," 2011. [Online]. Available:

http://apps.who.int/classifications/icfbrowser/ and http://bioportal.bioontology.org/ontologies/1411. [Accessed: 15-Jun-2011].

[12] HPLabs, “Jena, A Semantic Web Framework for Java,” 2010. [Online]. Available: http://openjena.org/index.html. [Accessed: 04-Jul-2011].

[13] OpenEHR Foundation, “The openEHR Reference Model,” 2008. [Online]. Available: http://www.openehr.org/releases/1.0.2/architecture/rm/ehr_im.pdf. [Accessed: 04-Jul-2011].

[14] Cieza, A., Geyh, S., Chatterji, S., and Konstanjsek, N., “ICF linking rules: an update based on lessons learned.,” *J Rehabil Med.*, vol. 37(4), pp. 212-218, 2005.

[15] Cimino, J. J., “Desiderata for Controlled Medical Vocabularies,” presented at the Methods Inf. Med., 1998, vol. 37, pp. 394-403.

[16] Sowa, J. F., *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Pacific Grove, Ca, USA: Brooks Cole Publishing, 2000.

[17] ICF Research Branch, “ICF Core Sets Projects.” [Online]. Available: http://www.icf-research-branch.org/icf-core-sets-projects.html. [Accessed: 04-Jul-2011].

[18] Cieza, A., Brockow, T., Ewert, T., and Amman, E., “Linking health-status measurements to the international classification of functioning, disability and health.,” *J Rehabil Med.*, vol. 34(5), pp. 205-210, Sep. 2002.

[19] College of American Pathologists y National Health Service, “Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT).” [Online]. Available: http://bioportal.bioontology.org/ontologies/42157?p=terms&conceptid=362981000#details. [Accessed: 15-Jun-2011].

[20] Beale, T. and Heard, S., “An Ontology-based Model of Clinical Information,” presented at the MedInfo, Brisbane, Australia, 2007, pp. 760-764.

[21] Dublin Core, “Dublin Core Metadata Element Set, Version 1.1.” [Online]. Available: http://protege.stanford.edu/plugins/owl/dc/protege-dc.owl. [Accessed: 15-Jun-2011].

[22] National Centre for Classification in Health, “Condensed Classification of Health Development,” 2004. [Online]. Available: http://sydney.edu.au/health_sciences/ncch/condensed_classifications.shtml. [Accessed: 04-Jul-2011].

[23] World Health Organization, “International Classification of Diseases (ICD-10),” 2008. [Online]. Available: http:// apps.who.int/classifications/apps/icd/icd10online and https://dkm.fbk.eu/index.php/ICD-10_Ontology. [Accessed: 15-Jun-2011].

[24] Cuenca Grau, B., Horrocks, I., Motik, B., Parsia, B., Patel-Schneider, P., and Sattler, U., “OWL 2: The next step for OWL,” *J. of Web Sem.: Science, Services and Agents on the WWW*, vol. 4, pp. 309–322, 2008.

[25] Rodrigues, J. M., Kumar, A., Bousquet, C., and Trombert, B., “Using the CEN/ISO Standard for Categorial Structure to Harmonise the Development of WHO International Terminologies,” presented at the Med. Inf. in a United and Healthy Europe, Sarajevo, Bosnia and Herzegovina, 2009.