

Article

## Knowledge Representation for Prognosis of Health Status in Rehabilitation

Laia Subirats \*, Luigi Ceccaroni and Felip Miralles

Barcelona Digital Technology Centre, Roc Boronat, 117, 5th floor, MediaTIC Building, 08018 Barcelona, Spain; E-Mails: lceccaroni@bdigital.org (L.C.); fmiralles@bdigital.org (F.M.)

\* Author to whom correspondence should be addressed; E-Mail: lsubirats@bdigital.org; Tel.: +34-93-553-4540; Fax: +34-93-553-4541.

Received: 27 April 2012; in revised form: 31 July 2012 / Accepted: 1 August 2012 /

Published: 20 August 2012

---

**Abstract:** In this article, key points are discussed concerning knowledge representation for clinical decision support systems in the domain of physical medicine and rehabilitation. Information models, classifications and terminologies, such as the “virtual medical record” (vMR), the “international classification of functioning, disability and health” (ICF), the “international classification of diseases” (ICD) and the “systematized nomenclature of medicine—clinical terms” (SNOMED CT), are used for knowledge integration and reasoning. A system is described that supports the measuring of functioning status, diversity, prognosis and similarity between patients in the post-acute stage, thus helping health professionals’ prescription of recommendations.

**Keywords:** representation methods; clinical decision support systems; knowledge systems; rehabilitation; information models; classifications; terminologies

---

### 1. Introduction

In the domain of medicine in general, and physical medicine and rehabilitation in particular, several standard terminologies and classifications exist [1] that can be used for knowledge representation and integration. Some examples are: the *systematized nomenclature of medicine—clinical terms* (SNOMED CT) [2]; the *unified medical language system* (UMLS); GALEN; the *international classification of diseases* version 10 (ICD-10) [3]; and the *international classification of functioning,*

*disability and health* (ICF) [4] defined by the *World Health Organization* (WHO). In particular, the use of the ICF for measuring the functioning status and diversity with *multidimensional* indicators, at both individual and population levels, can contribute to solve interoperability problems among health institutions that employ different measuring questionnaires. The ICF framework classifies concepts of functioning, disability and health and specifies their range of values. In rehabilitation, which is a multidimensional process, the ICF is useful to achieve comparable and interoperable data collections [5] and is suitable for knowledge representation in *clinical decision support systems* (CDSSs), together with information models such as the *virtual medical record* (vMR), to contribute to solve interoperability problems in the electronic exchange of clinical information.

We designed and developed a CDSS which uses standard concepts and relations to represent and reason about patients' health status. For the decision-making process, a *knowledge-based system* (KBS) is used with *case-based reasoning* (CBR). In CBR, the two main types of case representation are *points* and *series*. Two key issues when cases are time series are the choice of *representation methods* and *similarity measures*. Regarding the series-based representation, various types of data abstraction [6] and time granularities can be used. Although many similarity measures have been proposed for time series [7], only the Euclidean distance is used in this paper, this being the first attempt to apply CBR to the rehabilitation domain. Our goals are:

- to present an application in the rehabilitation domain, in particular, a CDSS for the prognosis of health status of chronic patients who suffer from neurological diseases in the extra-hospital stage;
- to describe how the ICF, ICD-10, SNOMED CT and vMR can be used to build interoperable CDSSs;
- to show the pros and cons of representations methods of cases; and
- to explain how the ICF and ICD-10 are used in similarity measures of functioning status and diversity.

In the rehabilitation of people who suffer from neurological diseases, there are two basic stages: *intra-hospital* and *extra-hospital*. In the intra-hospital (acute) phase, patients who suffered from a traumatic or non-traumatic injury stay in hospital undergoing rehabilitation. After typically a few months in hospital, they return home and the extra-hospital phase starts. Thereafter, they go once a year to the rehabilitation hospital for a *periodic integral evaluation* (PIE), when they are administered several questionnaires (depending on the disease they are suffering from) with the aim of measuring functioning independence, psychological and social variables. In Section 3.2 we present examples of questionnaires administered in the PIE. This study is focused on the extra-hospital or chronic phase.

## 2. Results and Discussion

### 2.1. Characterization of CDSS in Rehabilitation

There are two main categories of CDSSs: those oriented to *assessment* and those oriented to *proposal*. The objectives of the assessment-oriented ones are: assessment of a patient's past, current and future status (this includes prognosis, *i.e.*, the likely outcome of an illness); risk quantification; and

classification of patients according to their functional diversity [8]. The objectives of proposal-oriented ones are: risk prevention; and definition of therapeutic goals.

We designed and developed a CDSS oriented to *assessment* and, more specifically, to prognosis, in a clinical-system environment which requires reasoning under uncertainty, with standard indicators (e.g., *Ingestion functions*—b510) used in the representation of cases. The way of **giving advice** can be a *push* action, with warnings or alarms, or a *pull* action, giving support when explicitly requested. In the intra-hospital stage, support is pulled during all the rehabilitation process and pushed in the initial assessment and at the end of each rehabilitation activity. In the extra-hospital stage, advice is pulled in the prognosis support and pushed at the end of the PIE.

In the **style of communication and decision-making process** of CDSSs applied to rehabilitation, both a consulting model and a critiquing model are used. In our case, the consulting model is used; the system provides an assessment of health future status and risks; and the final decision is taken by the health professional. With respect to the CBR used for the decision-making process, the system proposed only includes the *retrieve* step of the typical *retrieve, reuse, revise* and *retain* phases shown in [9].

## 2.2. Use of Information Models, Classifications and Terminologies in Rehabilitation

Regarding the use of **information models** in rehabilitation CDSSs, vMR is still in process of improvement and evolution by VMR Project Team [10], and there are several CDSSs that use it [11]. Moreover, VMR is designed to reduce development costs and time responses in CDSSs. As a consequence, although it is not widely implemented in hospitals and there are not any tools available to facilitate its implementation (unlike the information model EN/ISO 13606 [12]), it is probably the most appropriate information model for CDSSs today.

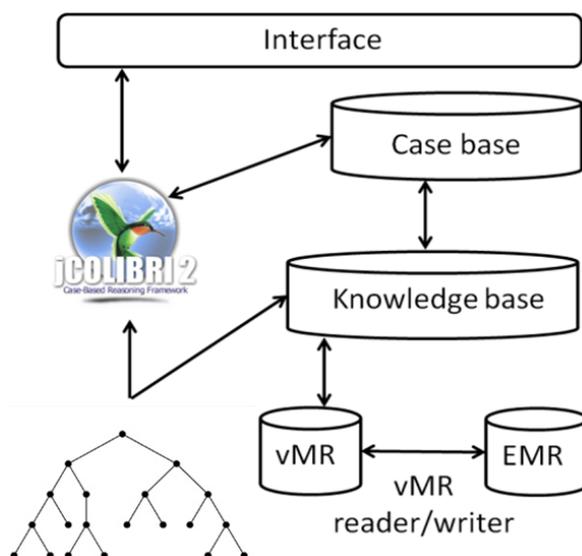
Several **classifications** and **terminologies** are used in CDSSs. In rehabilitation, ICF is used for encoding patients' health status and professionals' *recommendations*. Recommendations are activities and changes in environmental factors suggested by professionals to improve the quality of life of the patient. ICD-10 is used for representing diseases and SNOMED CT for other attributes; in particular, SNOMED CT has the potential of getting ICF-related terms mainstreamed in clinical systems, providing more context when necessary and including semantic relationships.

In rehabilitation processes, several questionnaires are used, which evaluate functioning, disability and health. In order to solve interoperability problems among questionnaires, items can be encoded to ICF concepts following the standardization methodology proposed by Cieza *et al.* (2005) [13]. Difficulties in mapping clinical questionnaires to standard terminologies and ontologies in the rehabilitation domain are summarized by Ceccaroni and Subirats (2012) [14]; e.g., data from questionnaires have usually a finer granularity than ICF *core set* categories. That is the reason why, to *summarize the health status* using ICF *core sets* from data obtained via questionnaires administered in the PIE (see examples of questionnaires in Section 3.2), the value of some (more general) nodes of the ICF taxonomy needs to be computed from the (more specific) node's children. This process is therefore performed in a bottom-up fashion and only the nodes with values are considered in the computation.

Several aggregating functions can be used depending on the attribute and the domain. In the physical rehabilitation domain, the average function is usually used. Another example is the *pessimistic function*, where if parents have an empty value, the parents' value is computed from the worst value of their children (ICF scale ranges from 4 = *complete problem* to 0 = *no problem*).

The CDSS, whose architecture is shown in Figure 1, uses standard information models, classifications and terminologies. Data from questionnaires administered in the PIE are stored in an *electronic medical record* (EMR) and a vMR reader/writer converts the EMR into vMR format. The health status obtained from vMR is represented in the knowledge base in the OWL ontology language [14]. A Java library converts original data into standardized data. A selection of instances of the knowledge base is stored in the case base, which, together with the ICF and ICD-10, is used to find similar patients via an open source CBR framework: jColibri [15]. Patients' health status information is summarized by another Java library which uses the ICF and is accessed through an Internet-based interface.

**Figure 1.** Clinical decision support systems (CDSS)' architecture.



### 2.3. Representation of Cases

The EN 12381:2005 “European norm of health informatics—Time standards for healthcare specific problems”, described by Ceusters *et al.* (1998) [16], is used to provide ontological and logical basis for describing temporal issues in healthcare. There are two types of patients' data: *static* and *dynamic*. Static variables are *absolute temporal expressions* as they do not change over time, for example *date of birth* and *date of diagnosis*. In the domain considered for this study, patient's *health status* and professionals' *recommendations* are dynamic attributes collected once a year during patients' rehabilitation.

Two representations of cases are implemented: a *point-based* representation, in which a case is the state of a patient in a given year; and a *series-based* one, in which a case includes several states. The temporal consistency of the case base must be ensured before the addition of a new case: if there is more than one state per year, the one with more available data is chosen; and alerts are activated when introducing temporally inconsistent states to the knowledge base.

### 2.3.1. Point-based Representation

In the point-based representation, health status and recommendations refer to patient encounters with yearly periodicity. The similarity measure for health status indicators and recommendations is

$$\text{interval}(i_1, i_2) = 1 - (|i_1 - i_2|/4) \quad (1)$$

where  $i_1$  and  $i_2$  are ICF values. To encode ICF values of difficulty, deficiency or barrier, the following code scheme is used:

- red/4: complete problem, difficulty, deficiency or environmental factors: 96%–100%;
- orange/3: severe problem, difficulty, deficiency or environmental factors: 50%–95%;
- yellow/2: moderate problem, difficulty, deficiency or environmental factors: 25%–49%;
- green/1: mild problem, difficulty, deficiency or environmental factors: 5%–24%;
- blue/0: no problem, difficulty, deficiency or environmental factors: 0%–4%.

The main drawback of this representation method is that, considering  $n$ -year time series of a patient represented as  $n$  cases, usually the most similar case belongs to the same patient in another year.

### 2.3.2. Series-based Representation

In the series-based representation, there are several structures which improve case performances in certain application domains such as *hierarchies* and *chains* or sequences. In this approach, *chains* are used to represent changes in values. This representation solves *incompleteness of data* in a year by considering the following year to compute changes over time.

In addition, in contrast with other domains, it is not possible to focus on a single case feature because rehabilitation is a *multidimensional* and holistic process. Effective and high quality rehabilitation of neurological diseases should cater for the physical, cognitive, psychological, social and cultural dimensions of the personality and lifestyle of patients and their families alike.

Figure 2 shows the symbol taxonomy used to encode changes in ICF values of barriers, facilitators, difficulties, capacity or magnitude of impairment and performance. The three **granularity levels** studied are:

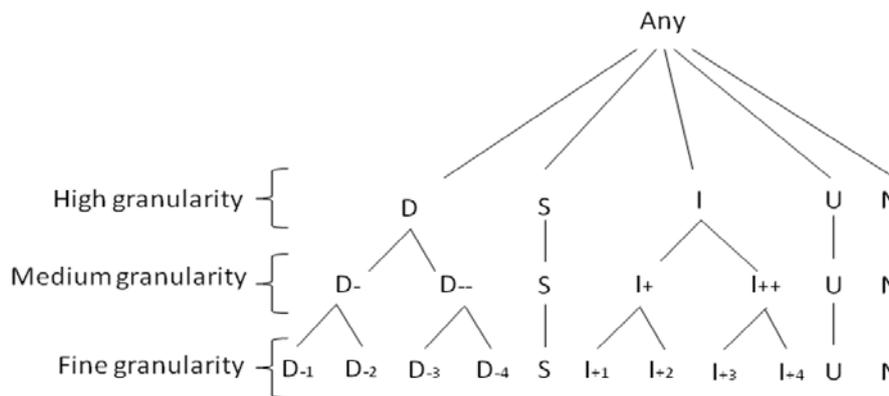
- fine granularity, which contains all possible changes in ICF values:  $D_{-n}$  means decreasing  $n$  levels;  $I_{+n}$  means increasing  $n$  levels; and  $S$  means remaining stationary;
- coarse granularity, which summarizes changes as decreasing (D), stationary (S) and increasing (I); and
- medium granularity, which is an intermediate representation between coarse and fine granularity.

Unspecified (U) and inapplicable (N) symbols correspond respectively to ICF values 8 and 9. It is important to consider that when an indicator decreases it improves according to the ICF standard because the lower the level, the less the difficulty, deficiency or barrier. Likewise, an increase indicates a deterioration of health status.

In a series-based representation, changes are considered instead of values. In order to also include values in series-based representations, the **first point** of the series is included. The similarity measure used is the Euclidean distance. The number of cases increases in fine granularities, especially in

large-time series and series with high rates of change, and sequences often only partially match. Therefore, finer granularities are the most suitable option for ICF categories with low rates of change and short time-series.

**Figure 2.** Taxonomy of the symbols used to encode changes in international classification of functioning, disability and health (ICF) values.



In the rehabilitation domain, there are no significantly **high rates of change** in the progression of most attributes (peaks are not frequent). The **frequency of time series** is *lower* than in other medical domains: usually it is yearly. Consequently, in physical rehabilitation, finer granularities are generally the most suitable option. Regarding the optimal length of time-series, it is not studied here as there are only four data points available per patient and they are all used.

#### 2.4. Patients' Similarity Measure

Patient similarity is actionable by the CDSS because the most similar patient's evolution is used for prognosis. The selection of the most similar case can be supervised by the professional. In that case, most relevant categories of k patients are shown to the professional. By default, relevant categories are ICF *core set* categories [17]. ICF *core sets* are subsets of ICF that can be formed according to functioning, pathology or rehabilitation process [18]. *Core sets* are useful because, in daily practice, clinicians and other professionals need only a fraction of the categories found in ICF.

To compute cases' similarity, the *k-nearest neighbor* (k-NN) algorithm is used. Similarity between patients can be calculated through  $Sim(X, Y) = \sum_{i=1}^j w_i Sim(X_i, Y_i)$  (2), in which X and Y are two patients,  $Sim(X_i, Y_i)$  is the similarity on the *i*<sup>th</sup> attribute with weight  $w_i$ . ICF core sets and ICF taxonomy are used for weight assignment. The weight given to level-4 and level-3 *core set* categories is 2 and 1, respectively. Therefore *core set* categories which are more specific in the ICF taxonomy (level 4 *core set* categories) carry more weight in the similarity calculation.

Diseases terms are encoded to ICD-10 following the Centers for Medicare and Medicaid Services (CMS) and the National Center for Health and Statistics (NCHS) (2012) [19] methodology. Problems have been found in the use of ICD-10 for similarity calculation between diseases and have been solved as follows:

- Disease terms are not encoded to a single ICD-10 category. In this case, the most representative ICD-10 category is chosen.

- Disease terms are encoded to the same ICD-10 category. In this case, both concepts cannot be distinguished if ICD-10 is not extended.
- ICD-10 contains catch-all categories. In this case, only categories that cannot be encoded to an ICD-10 category are encoded to “Other disease of type X”.
- ICD-10 contains scattered exclusions. In this case, the most representative ICD-10 category is chosen.
- Diseases terms are complications of other diseases. In this case, the most representative ICD-10 category is chosen.

The similarity function between diseases used in the physical rehabilitation domain is [14]:

$$\text{cosine}(i_1, i_2) = \text{sim}(t(i_1), t(i_2)) = \frac{\left| \left( \bigcup_{d_i \in t(i_1)} (\text{super}(d_i, CN)) \right) \cap \left( \bigcup_{d_i \in t(i_2)} (\text{super}(d_i, CN)) \right) \right|}{\sqrt{\left| \left( \bigcup_{d_i \in t(i_1)} (\text{super}(d_i, CN)) \right) \right|} * \sqrt{\left| \left( \bigcup_{d_i \in t(i_2)} (\text{super}(d_i, CN)) \right) \right|}} \quad (3)$$

where  $i_1$  and  $i_2$  are the diseases,  $CN$  is the set of all concepts in the current knowledge base,  $\text{super}(c; C)$  is the subset of concepts in  $C$  which are super concepts of  $c$ , and  $t(i)$  is the set of individual  $i$  concepts.

### 3. Experimental Section

#### 3.1. Physical Rehabilitation Scenario

Prognosis of patient’s functioning status, based on evidence or previous cases, help health professionals to analyze patients’ case and improve the efficacy of their recommendations. There have already been some initiatives for extracting knowledge patterns about the evolution over time of patients who suffer from *spinal cord injury* (SCI) [20]. In the following physical rehabilitation scenario, we describe a real life setting where the CDSS is used for research purposes.

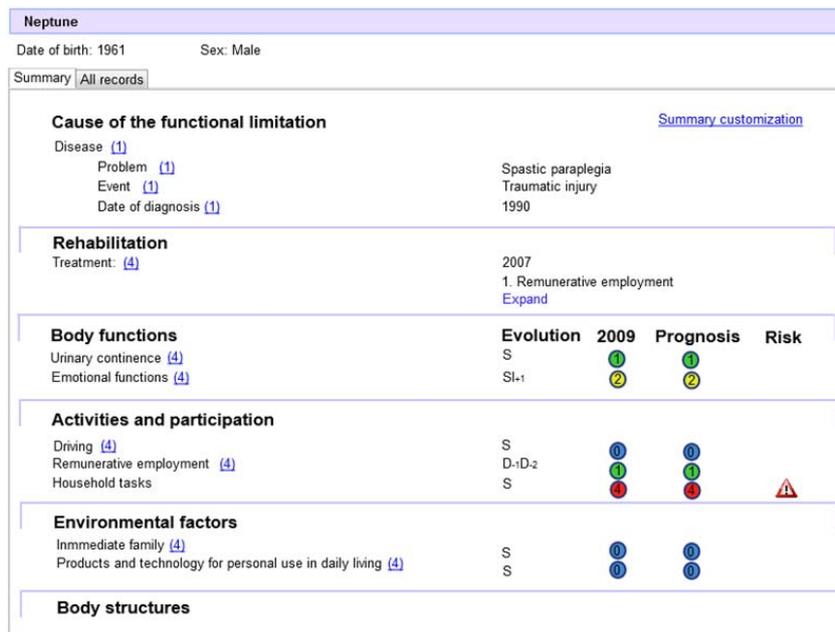
Chronic patients’ data in the extra-hospital phase used in the scenario are taken from a bio-psycho-social study which analyzed the evolution of patients with neurological disability [21]. This study contains data of 661 patients who suffer from SCI over a span of 4 years. In total there are 5995 series-based cases and 2943 point-based cases obtained from PIE measurements in the period 2007–2010.

As an example, Neptune is the (anonymized) name of a 51-year-old man from Barcelona who suffers from spastic paraplegia since he had a traumatic injury in 1990. In his last PIE in 2007 (2008 data are not available), his doctor recommended that he seek an employment. After completing his 2009 PIE, his doctor analyzed his health status and recommended for him a change in his activities or environmental factors. Figure 3 shows Neptune’s summary and prognosis during his PIE in 2009. The added value of using a CDSS in this scenario, apart from predicting a patient future status, is that the prognosis motivates Neptune to find an employment, as he can see the impact of this factor on his functioning. This prediction helps doctors to analyze his case and prescribe recommendations for changes in activities, participation and environmental factors.

The proposed CDSS predicts Neptune’s functioning status based on previous cases. His summary is composed of the cause of functioning limitation, his recommendations, and the four chapters of ICF: body functions, activities and participation, environmental factors and body structures. In Figure 3,

indicators of the ICF spinal cord injury (SCI) core set for which data is available appear in ICF sections. Risk attributes which have a severe or grave level, or decrease by 2 levels or more, are outlined.

**Figure 3.** Neptune’s summary and prognosis during his periodic integral evaluation (PIE) in 2009.



### 3.2. Use of Information Models, Classifications and Terminologies in Rehabilitation

Classes obtained from the EMR of the hospital are mapped to vMR classes, and standard classifications and terminologies, as shown in Table 1. In the physical rehabilitation scenario, the PIE assesses:

- functioning data, using the *functional independence measure* (FIM) and the *spinal cord injury measure* (SCIM);
- emotional status and well-being, using the *hospital anxiety and depression* (HAD) questionnaire and the *psychological well-being index* (PWBI);
- quality of life, using the short version of the *world health organization quality of life assessment instrument* (WHOQOL-BREF).

Due to the fact that the **recommendations** provided by professionals are not among available data, this study assumes that changes in part of the environment and activities arise as result of health professional’s prescriptions. In the extra-hospital phase, recommendations are activities (e.g., *remunerative employment*) and changes in the patients’ environment (e.g., *social support services*) to improve their quality of life. One criterion for selecting recommendations is to choose at least one item from each type of questions of the *socio-demographic questionnaire* (SDQ), namely: changes in coexistence, dependence care, housing, mobility, work activity, educational level, activities, economic benefits and health services.

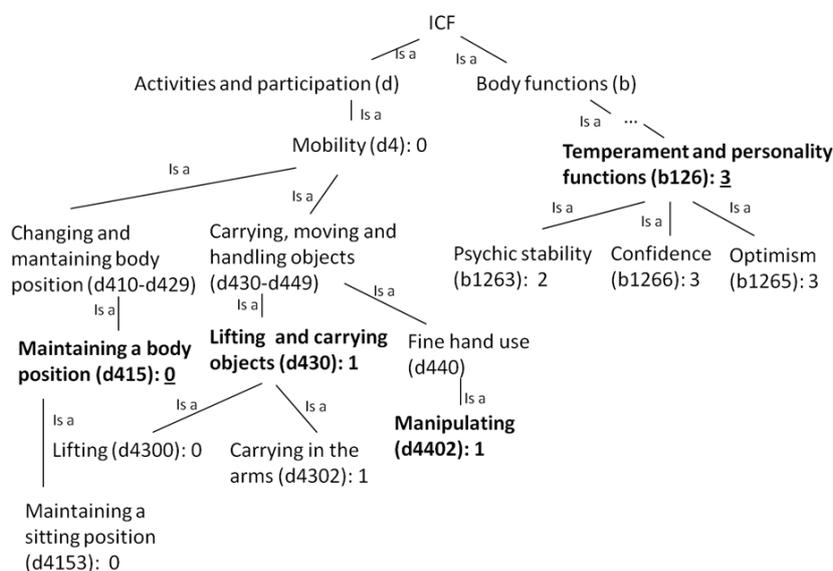
To standardize to ICF and summarize patient’s health status shown in Figure 4, a Java library was developed. Categories appearing in the SCI long-term context *core set* are marked in bold, computed values from lower level categories are underlined. Values which are not underlined are obtained from

PIE questionnaires. In this example, the result of computing average and pessimistic functions for *temperament and personality functions* coincide. In addition, *maintaining a body position* has the same value as its child (*maintaining a sitting position*) because there is only one child with value.

**Table 1.** Relationship between Virtual Medical Record (vMR), terminologies and Electronic Medical Record (EMR) classes.

vMR Class: attribute	Standard classification class	EMR reference
Person: name	Patient name	
EvaluatedPerson (subclass of Person): birthTime, gender	Date of birth, gender	
Problem: problemCode	ICD-10	Patient’s static information
ProblemBase (subclass of Problem): diagnosticEventTime, severity	Date of diagnosis, severity score	
AdverseEventBase: adverseEventcode	Event	
GoalBase: goalFocus, goalAchievementTarget Time	Therapy, timing milestone	SDQ
ObservationBase: observationFocus	ICF	FIM, SCIM, HAD, PWBI, WHOQOL-BREF

**Figure 4.** Summary of Neptune’s health status in 2007 using international classification of functioning, disability and health (ICF) taxonomy.



### 3.3. Use of Diseases to Measure Patients’ Similarity

Some examples of how problematic diseases (see Section 2.4) are encoded and used in similarity calculation are:

- Diseases not encoded to a single ICD-10 category. *Myopathy* is encoded as *primary disorders of muscles* (G71), *other myopathies* (G72) and *disorders of muscles* (M60-M63) and, if no additional information is available, the first option is chosen as it is the most representative according to an experts’ consensus.

- Diseases encoded to the same ICD-10 category. Both *complete paraplegia* and *incomplete paraplegia* are encoded as *spastic paraplegia* (G82.1), therefore their similarity is considered equal to 1.
- Diseases encoded to catch-all ICD-10 categories. *Other ischemic stroke* is encoded as *other cerebral infarction* (I63.8).
- Diseases encoded to ICD-10 categories with scattered exclusions. In this case, if no additional information is available, the most representative ICD-10 category is chosen according to an experts' consensus. For example, for paraplegia various categories exist which include the concept, such as *hereditary spastic paraplegia* (G11.4), but the *spastic paraplegia* (G82.1) category is chosen for being more representative.
- Diseases that may occur as a complication in the course of some other disease. This problem has not been studied in the scenario.

In this way it is possible to calculate the similarity between two diseases with ambiguous representation in the ICD, such as, for example, *Myopathy* and *Other ischemic stroke*; similarity which has a value in this case of 0.5, considering related classes G71 and I63.8 respectively.

Regarding patients' similarity calculation, a 10-NN classifier was developed using the jColibri framework.

### 3.4. Case Representation Methods

To compare and evaluate the different case representation methods, chronic patients' original data from the physical rehabilitation scenario are converted into the different representations. Point-based and series-based representations of Neptune are shown in Tables 2 and 3. In the point-based representation, only the prognosis of 30 patients out of 661 who suffered from SCI could be performed because of missing data. In Table 2, the values of *remunerative employment* in years 2006, 2007, 2009 and 2010 are 4, 3, 1 and 1, respectively. These values, encoded according to the *fine granularity* are represented as D<sub>1</sub>D<sub>2</sub>S (See Table 3.) because the attribute decreases one level, then two levels and finally remains stationary. The values, encoded to the *medium granularity*, are represented as D.S according to the correspondence between fine and medium symbol granularities of Figure 4 and considering that a sequence of same symbols is summarized using just one instance of these symbols; *i.e.*, D.D.S is written as D.S. Finally, in the *coarse granularity*, these values are encoded as DS.

The most similar patient to Neptune is dependent on the *symbol granularity* chosen. For example, the most similar patient to Neptune is Diana, Juno and Apollo according to the fine, medium and coarse symbol granularity, respectively. However, there are cases in which the most similar case according to the three granularities is the same. One example is the case of Venus and Minerva, which are the most similar patients according to all symbol granularities; both of them have stationary values in most of their attributes and only data from two years are available.

**Table 2.** Neptune’s recommendations of changes in activities, participation and environmental factors to improve his functioning in the point-based case representation. For more information about ICF codes and descriptions see [4].

Person, year\recommendations	Environmental factors		Activities and participation								
	Immediate family	General social support services	Acquiring a place to live	Driving	Using transportation	Informal education	Public economic entitlement	Sports	Preparing meals	Crafts	Remunerative employment
Neptune, 2006	0	8	0	0	8	2	0	4	4	4	4
Neptune, 2007	0	8	0	0	8	2	0	4	4	4	3
Neptune, 2009	0	8	0	0	8	2	0	4	4	4	1
Neptune, 2010	0	8	0	0	8	3	0	4	4	4	1

**Table 3.** Patients’ recommendations of changes in activities, participation and environmental factors to improve his functioning in the series-based case representation.

Symbol granularity	Recommendations\person	Neptune	Diana	Juno	Apollo	Venus	Minerva
<b>Fine</b>	Using transportation	M	<u>M</u>	SS	<u>M</u>	M	M
	Driving	SSS	<u>SSS</u>	M	S	S	S
	Acquiring a place to live	SSS	<u>SSS</u>	SS	<u>SSS</u>	S	S
	Preparing meals	SS	<u>SS</u>	S	<u>SS</u>	I <sub>+4</sub>	I <sub>+4</sub>
	Informal education	SSI <sub>+1</sub>	SSS	SS	I <sub>+1</sub> SS	S	S
	Remunerative employment	D <sub>-1</sub> D <sub>-2</sub> S	SSD <sub>-1</sub>	I <sub>+1</sub> I <sub>+2</sub>	I <sub>+1</sub> D <sub>-1</sub> S	S	S
	Public economic entitlement	SSS	<u>SSS</u>	SD <sub>-4</sub>	<u>SSS</u>	S	S
	Sports	SSS	D <sub>-4</sub> SS	SS	I <sub>+4</sub> D <sub>-4</sub> S	S	S
	Crafts	SSS	<u>SSS</u>	SS	SD <sub>-4</sub> I <sub>+4</sub>	S	S
	Immediate family	SSS	M	SS	<u>SSS</u>	S	S
General social support services	M	<u>M</u>	SI <sub>+1</sub>	<u>M</u>	M	M	
<b>Medium</b>	Preparing meals	S	<u>S</u>	<u>S</u>	<u>S</u>	I <sub>++</sub>	I <sub>++</sub>
	Remunerative employment	D-S	SD <sub>-</sub>	I <sub>+</sub>	I <sub>+</sub> D <sub>-</sub> S	S	S
	General social support services	M	<u>M</u>	SI <sub>+</sub>	<u>M</u>	M	M
<b>Coarse</b>	Preparing meals	S	<u>S</u>	<u>S</u>	<u>S</u>	I	I
	Remunerative employment	DS	SD	I	IDS	S	S
	General social support services	M	<u>M</u>	SI	<u>M</u>	M	M

A weakness of Table 3 is the fact that only changes are considered and no values. This is solved in Table 4 by adding the value of the initial point. For instance, although the attribute *preparing meals* is stationary for Neptune, Diana, Juno and Apollo; Neptune and Juno have complete difficulty while Diana and Apollo have no difficulty. In order to include values in series-based representation, the first point level precedes the series of interval symbols. In Table 4, Vesta is the most similar person to Neptune. Ceres is an example of a person with the recommendation of *general social support services*, and the most similar person to her is Mars.

**Table 4.** Patients’ recommendations of changes in activities, participation and environmental factors to improve functioning in the series-based case representation which includes the value of the initial point.

Recommendations/person	Neptune	Vesta	Diana	Ceres	Mars
Preparing meals	4 > S	0 > S	0 > S	4 > S > D <sub>-4</sub>	4 > S
Remunerative employment	4 > D <sub>-1</sub> > D <sub>-2</sub> > S	4 > S > D <sub>-1</sub>	4 > S > D <sub>-1</sub>	3 > S	4 > S
General social support services	8	8	0	1 > S > D <sub>-1</sub>	1 > S

When computing the most similar person among people who suffered from SCI, the number of **coincidences** between the coarse and medium granularities is 61%, and the coincidence in the coarse and fine granularities is 35%. This low value made us decide to generally use the fine granularity in the physical rehabilitation scenario to avoid loss of information. Furthermore, the coincidence between fine-granularity series with and without the value of the initial point is 5% only. Therefore, the inclusion of the value of the initial point has a high repercussion on the similarity measure, and should be included in series-based representation.

### 3.5. Limitations

An initial limitation of the system was the high response time in computing ICD-10 similarities and in the retrieval of the most similar case in large databases. To improve the response time, all ICD-10 similarities were computed once and then stored in a database, thus avoiding the need of loading and reasoning on the 14502 classes of the ICD-10 ontology [3] in real time. Another limitation is that the cosine similarity is an edge-based approach where diseases at the same level in the ontology have the same similarity, while their semantic distance is not the same. We will study other similarity functions proposed for biomedical ontologies [22]. Furthermore, patient’s similarity calculation is currently slow implementing the *cached linear case base* [15] on a typical desktop computer. In the future, an optimized organization of the case base will be studied. Finally, the optimal length of time-series to be used as cases is not studied in this experimental scenario in which time-series were only up to four points; therefore large time-series might also be a limitation of the current model if they are not dealt with adequately.

## 4. Conclusions

We presented a *clinical decision support system* for prognosis in the physical medicine and rehabilitation domain, which uses case-based reasoning and a standardized knowledge representation. The use of standard classifications, terminologies and information models, such as HL7’s *virtual medical record*, helps to integrate knowledge in *clinical decision support systems*. In particular, the use of the *international classification of functioning, disability and health* (ICF) contributes to solve interoperability problems among questionnaires used in the management of various diseases by different health institutions. The ICF is also used to summarize the patients’ health status and to calculate similarity among patients. The *international classification of diseases version 10* helps to calculate similarity between diseases (and patients). Finally, the *systematized nomenclature of medicine—clinical terms* gives more level of detail when necessary. Regarding the case representation,

a series-based representation method has been used because representing attribute's changes over time is important in the rehabilitation domain. We showed results for the prognosis of the health status of chronic patients who suffer from neurological diseases in the extra-hospital stage. Four-year time series are used with over a hundred attributes.

### Acknowledgments

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 and it is also partly supported by the Catalonia Competitiveness Agency (ACCÍÓ). The opinions expressed in this paper are those of the authors and are not necessarily those of Rehabilita project's partners, CDTI or ACCÍÓ.

### References

1. Carter, J.H. Design and Implementation Issues. In *Clinical Decision Support Systems*; Berner, E.S., Hannah, K.J., Ball, M., Eds.; Springer: New York, NY, USA, 2007; Section 1, pp. 64–98.
2. College of American Pathologists and National Health Service. Systematized Nomenclature of Medicine—Clinical Terms (SNOMED CT). Available online: <http://purl.bioontology.org/ontology/SNOMEDCT> (accessed on 31 July 2012).
3. World Health Organization. International Classification of Diseases (ICD-10). Available online: [https://dkm.fbk.eu/index.php/ICD-10\\_Ontology](https://dkm.fbk.eu/index.php/ICD-10_Ontology) (accessed on 30 July 2012).
4. World Health Organization. International Classification of Functioning, Disability and Health (ICF). Available online: <http://purl.bioontology.org/ontology/ICF> (accessed on 31 July 2012).
5. Steiner, W.A.; Ryser, L.; Huber, E.; Uebelhart, D.; Aeschlimann, A.; Stucki, G. Use of the ICF model as a clinical problem-solving tool in physical therapy and rehabilitation medicine. *J. Phys. Ther.* **2002**, *11*, 1098–1107.
6. Keravnou, E.; Shahar, Y. Temporal Reasoning in Medicine. In *Handbook of Temporal Reasoning in Artificial Intelligence*; Fisher, M.D., Gabbay, D.M., Vila, L., Eds.; Elsevier: Amsterdam, The Netherlands, 2005; pp. 587–653.
7. Ding, H.; Trajcevski, G.; Scheuermann, P.; Wang, X.; Keogh, E. Querying and mining of time series data: Experimental comparison of representations and distance measures. Available online: <http://www.vldb.org/pvldb/1/1454226.pdf> (accessed on 14 August 2012).
8. Romañach, J.; Lobato, M. Functional diversity, a new term in the struggle for dignity in the diversity of the human being. Available online: [http://www.leeds.ac.uk/disability-studies/archiveuk/zavier/Functional%20Diversity%20%20%20%20%20%20%20%20%20%20%20\\_fv%20Roma%Flach.pdf](http://www.leeds.ac.uk/disability-studies/archiveuk/zavier/Functional%20Diversity%20%20%20%20%20%20%20%20%20%20%20_fv%20Roma%Flach.pdf) (accessed on 14 August 2012).
9. Aamodt, A.; Plaza, E. Case-based reasoning: Foundational issues, methodological variations, and system approaches. *J. AI Commun.* **1994**, *7*, 39–59.
10. VMR Project Team. Virtual Medical Record (vMR) for Clinical Decision Support—Domain Analysis Model 2011. Available online: [http://wiki.hl7.org/index.php?title=Virtual\\_Medical\\_Record\\_\(vMR\)](http://wiki.hl7.org/index.php?title=Virtual_Medical_Record_(vMR)) (accessed on 31 July 2012).

11. Kashfi, H. The Intersection of Clinical Decision Support and Electronic Health Record: A Literature Review. In *Proceedings of the Federated Conference on Computer Science and Information System*, Szczecin, Poland, 18–21 September 2011; pp. 347–353.
12. Martínez-Costa, C.; Menárguez-Tortosa, M.; Fernández-Breis, J.T. An approach for the semantic interoperability of ISO/EN 13606 and *openEHR* archetypes. *J. Biomed. Informa.* **2010**, *43*, 736–746.
13. Cieza, A.; Geyh, S.; Chatterji, S.; Konstanjsek, N. ICF linking rules: An update based on lessons learned. *J. Rehabil. Med.* **2005**, *37*, 212–218.
14. Ceccaroni, L.; Subirats, L. Interoperable knowledge representation in clinical decision support systems for rehabilitation. *J. Appl. Comput. Math.* **2012**, *11*, 303–316.
15. Recio-García, J.A.; Díaz-Agudo, B.; González-Calero, P. *jColibri2 Tutorial*. Universidad Complutense de Madrid: Madrid, Spain, 2008.
16. Ceusters, W.; Steurs, F.; Zanstra, P.; van der Haring, E.; Rogers, J. From a time standard for medical informatics to a controlled language for health. *J. Med. Inf.* **1998**, *48*, 85–101.
17. Coenen, M. Developing a method to validate the WHO ICF Core Sets from a patient perspective: Rheumatoid arthritis as a case point. Ph.D. Thesis, Ludwig-Maximilians-Universität, München, Germany, January 2008.
18. Herrmann, K.H.; Kirchberger, I.; Stucki, G.; Cieza, A. The comprehensive ICF core sets for spinal cord injury from the perspective of occupational therapists: A worldwide validation study using the Delphi technique. *J. Spinal Cord.* **2011**, *49*, 600–613.
19. Centers for Medicare and Medicaid Services (CMS) and the National Center for Health and Statistics (NCHS). ICD-10-CM Official Guidelines for Coding and Reporting 2012. Available online: <http://www.cdc.gov/nchs/data/icd10/10cmguidelines2012.pdf> (accessed on 31 July 2012).
20. Gibert, K.; García-Rudolph, A.; Curcoll, L.; Soler, D.; Pla, L.; Tormos, J.M. Knowledge Discovery about Quality of Life Changes of Spinal Cord Injury Patients: Clustering based on Rules by States. In *Proceedings of the International Conference of the European Federation for Medical Informatics (MIE)*, Sarajevo, Bosnia and Herzegovina, 30 August–2 September 2009; pp. 579–583.
21. Subirats, L.; Orte, S.; López, R.; Torrellas, S.; García-Rudolph, A.; Ceccaroni, L.; Gil, A.; Miralles, F.; Tormos, J.M. Use of Bio-psycho-social Standards and User Experience in a Social Network Which is Focused on People with Neurological Disability. In *Proceedings of the International Congress on Design, Research Networks and Technology for All (DRT4ALL)*, Madrid, Spain, 27–29 June 2011; pp. 314–330.
22. Pesquita, C.; Faria, D.; Falcão, A.O.; Lord, P.; Couto, F.M. Semantic similarity in biomedical ontologies. *PLoS Comput. Biol.* **2009**, *5*, doi:10.1371/journal.pcbi.1000443.