

## ORIGINAL ARTICLE

# Integrating semantic and fuzzy dimensions into electronic medical records: Case of cerebral palsy information system

Hanen Ghorbel\*, Sirine Farjallah

*Higher Institute of Management, Sousse, Tunisia*

**Received:** February 16, 2018

**Accepted:** May 4, 2018

**Online Published:** June 20, 2018

**DOI:** 10.5430/jha.v7n4p60

**URL:** <https://doi.org/10.5430/jha.v7n4p60>

## ABSTRACT

The meta-modeling of medical records helps standardize and capitalize the expert's knowledge domain. It promotes the interoperability knowledge and the reuse of clinical concepts, i.e., archetypes. It also promotes high quality electronic medical record system (EMRS) design, which helps provide better care service delivery. As a result, different standards of medical informatics use the dual model to support interoperability between Medical Information Systems. We particularly quote ISO/EN 13606 and OpenEHR. However, the use of these standards still presents challenges. Apart from political reasons, the main obstacles to the adoption of these standards include: (1) a lack of guides and methodological tools to facilitate the construction of EMRS using two conceptual levels. Designers must have languages, approaches and tools to assist them in the modeling of archetypal EMRS; (2) a lack of methodologies for semantic activities on the content of electronic health records in the semantic web environment; (3) and a lack of management of uncertainties and inaccuracies that may exist in the medical field. The construction of an approach to modeling EMRS according to the dual model approach, considering the uncertainties, inaccuracies and semantics of these systems, is a difficult task, given the challenges to emancipate. In literature, we don't find such an approach. We, therefore, defined one in this paper. Our goal is to guide the designer in all stages of developing a new generation of EMRS, from analysis and specification of requirements to implementation. To achieve this goal, we have created an approach to support the following activities: (1) clinical concepts and information management and meta-modeling in accordance with the openEHR standard, (2) integration of the semantic dimension into EMRS considered to enable the execution of semantic activities in the semantic web environment; and (3) integration of the fuzzy dimension into electronic medical record data structures. As a contribution, we defined an approach called Fuzzy SemanticOpenEHR allowing the integration of semantic and fuzzy dimensions into EMRS modeled using the openEHR standard. Fuzzy SemanticOpenEHR intends to help and equip the designer during the different phases of creating a fuzzy ontology. Thanks to the mechanisms offered by this approach, we have been able to obtain a fuzzy ontological basis that can serve as a knowledge base that can support the semantic interoperability between EMRS, the deduction of new knowledge and the taking of knowledge's clinical decision. To test our contribution, we proceeded to the realization of a prototype of tools realized for the pediatric neurology service of the university hospital "Hédi Chaker Sfax - Tunisia" and the association of the handicapped persons safeguard of Sfax. This prototype is a framework called "XML 2 FuzzyOWL". Then, we tested this framework using a case of a disease which is "Cerebral Palsy".

**Key Words:** Electronic medical record, Semantic interoperability, Fuzzy ontology, Cerebral palsy information system

\*Correspondence: Hanen Ghorbel; Email: [hanenghorbel.fseg@gmail.com](mailto:hanenghorbel.fseg@gmail.com); Address: Higher Institute of Management, Sousse, Tunisia.

## 1. INTRODUCTION

The medical sector is a broad field of application for IT: from the management of healthcare facilities and medical practices, to the development of expert systems, to diagnostic support systems. The reluctance of the medical community to deal with computers in the sixties was understandable. Indeed, the big machines of the time required the permanent assistance of a computer scientist. A decisive turning point came in the mid-1970s, thanks to the appearance of micro-computers, which offer flexibility of use and a constantly decreasing price / quality ratio, thus encouraging the medical profession to become more interested in IT tools.

Medical informatics then asserted itself as a scientific discipline aiming to reinforce the capacities of health services with technologies of treatment and communication of the medical information, thanks to its contributions at the level of the techniques of collecting, memorizing, exchanging and interpreting this information. Its contribution to improving health systems and reducing costs has greatly promoted the adoption of computerized Medical Information Systems (MIS).

A MIS is an organized set of resources for grouping, classifying, processing and disseminating medical information.<sup>[1]</sup>

Our problem in this work is based on the modeling of information and medical knowledge. This modeling must ensure the semantic interoperability of heterogeneous Electronic EMRS. The European Commission's (Commission Recommendation of 2 July 2008 on cross-border interoperability of electronic health record systems; document number C [2008] 3282; 2008) recommendations have stated that the interoperability of these systems is necessary to improve the quality of care.<sup>[2]</sup> In fact, the information exchanged is of little use in the absence of clarity about its context and meaning. Semantic interoperability helps to clarify meaning by harmonizing data, ensuring that information is used in a sustainable and unambiguous way. However, the fundamental problem of the interoperability of EMRS is to see how complex "terms" (diagnostics or results) can be coded, in free texts, but also by an abstract formulation, to leave no room for free interpretation and ambiguity. This semantic interoperability is in fact ensured through ontologies. Actually, these ontologies model the medical field using their different components: concepts, relationships, axioms and instances. And since this field has inaccuracies and imperfections, it cannot be conceived using conventional ontologies. To fill this gap, the researchers thought of applying the notions of fuzzy logic to ontologies, which results in fuzzy ontologies. Indeed, these ontologies now represent a great interest for various domains, including that of the semantic Web that

makes semantic contents of Web resources interpretable not only by humans, but also by the machine. Fuzzy ontologies are a promising area of research on which we have located our work. Indeed, our research work aims to contribute to the promotion of the efficiency of MIS with means to ensure a better treatment of the patient's therapeutic state; we want to improve MIS by using fuzzy ontologies to ensure semantic interoperability.

In fact, the meta-modeling of medical records helps standardize and capitalize the knowledge of the experts' domain. It promotes the interoperability of knowledge and the reuse of clinical concepts, i.e., archetypes. It also promotes high quality EMR systems design, which helps provide better care service delivery. As a result, different standards of medical informatics use the dual model to support interoperability between MIS. ISO/EN13606<sup>[3]</sup> and OpenEHR are particularly cited.

However, the use of these standards still presents challenges. Key barriers to adopting these standards include: (1) a lack of tools and guidelines to guide the construction of EMRS. Indeed, designers must have languages, approaches and tools to assist them in modeling EMRS based on archetypes; (2) a lack of methodologies for semantic activities on the content of EMRS in the semantic web environment.

The construction of a methodology for modeling EMRS according to the dual model approach, considering the evolution, adaptation and semantics aspects of these systems, is a laborious task, given the challenges to overcome. In the literature, we find in the "Semantic OpenEHR" approach.<sup>[4]</sup> This approach concerns the technological space of the semantic web. It allows the integration of the semantic dimension into the EMRS modeled using the openEHR standard. Their adoption of this standard is argued by the fact that it has a large community of users and developers and is used in several countries such as Australia and Holland. This allowed us to situate our intention of contribution. We defined an approach called Fuzzy SemanticOpenEHR allowing the integration of semantic and fuzzy dimensions into electronic health record systems modeled using the openEHR standard. Fuzzy SemanticOpenEHR is intended to help and equip the designer during the different phases of creating a fuzzy ontology. Thanks to the mechanisms offered by this approach, we have been able to obtain a fuzzy ontological source that can serve as a knowledge base that can support the semantic interoperability between electronic medical record systems, the deduction of new knowledge and the taking of knowledge's clinical decision. This approach is defined as an extension of the SemanticOpenEHR approach presented.<sup>[4]</sup>

To test our contribution, we proceeded to the realization of a

prototype of tools realized for the pediatric neurology service of the university hospital “Hédi Chaker Sfax - Tunisia” and the association of the handicapped persons safeguard of Sfax. This prototype is a framework called “XML 2 FuzzyOWL”. Then, we tested this framework using a case of a disease that is “Cerebral Palsy” (CP).

The present paper is composed of five sections. It starts with an introductory section. Section 2 provides some background that represents CP disease modeling in accordance with the archetype approach and describes some previous works which focused on providing an ontological representation of archetypes and on their semantic management. Section 3 describes our suggested solution for supporting semantic activities and fuzzy dimension on CP EMRS. The feasibility of our solution is shown in Section 4. Finally, some discussions and conclusions are put forward in Section 5.

## 2. BACKGROUND

### 2.1 Information Medical System (MIS)

An MIS is an organized set of resources for grouping, classifying, processing and disseminating medical information.<sup>[1]</sup> It includes all the information necessary for the accomplishment of the activities belonging to the therapeutic course of the patients.

A MIS can be considered as a type of the Hospital Information System (HIS).<sup>[5]</sup> In fact, there are two types of HIS:

- The administrative information system including activities that support the patient’s care process, but do not participate directly. The information generated by this type of systems focuses on the personnel’s management, financial management, and all so-called stewardship activities (e.g., meal management, lingerie management ...).
- The medical information system including all the information that are directly related to the accomplishment of the acts of the production process of patients’ care (we are interested in the second type of HIS).

### Objectives

A MIS has the following objectives:

- Improving the quality and continuity of care: the sharing of medical information, as well as the interoperability between health systems, are two major contributing factors in improving the quality and continuity of care.
- Cost control: The optimization of medical processes, the reduction of administrative tasks, the reduction of the length of stays and the provision of medico-economic management tools are factors that influence

cost control. Proper modeling of medical information influences these different factors and thus leads to cost control.

Cost control: The optimization of medical processes, the reduction of administrative tasks, the reduction of the length of stays and the provision of medico-economic management tools are factors that influence cost control. Proper modeling of medical information influences these different factors and thus leads to cost control.

- The patient and one or more health professionals, including at least one medical professional;
- More health professionals, including at least one medical professional.

It is possible to establish a remote diagnosis to : require the opinion of a specialist, provide a preventive follow-up for a patient at risk or a post-therapeutic follow-up, prescribe therapies, prescribe or carry out services or acts, or to monitor the patient’s condition. However, telemedicine has been integrated into a broader framework, called e-health, which refers to all technologies for health. E-health has promoted coordination and cooperation between care providers, where the doctors’ and all health professionals’ actions on the patient can be treated. This makes the possibility of having a patient’s centered view where a caregiver can access all the medical information related to a patient, no matter the place and the time of collecting this information, to ensure a real coordination of multi-criteria and multidisciplinary care and decision-making.

### 2.2 Medical file: Crucial component of a MIS

It allows to progressively saving many data related to his health.

Any patient admitted to hospital or consultation, in a public or private care facility, must be assigned by a medical file. The latter includes all the information necessary for the care and monitoring of a patient. This document makes it possible to record the trace of any diagnostic, therapeutic and preventive action applied on a patient. The content of the medical file is not limited to the written observation of the doctor or the notes of the nurse but earlier, it includes all types of data related to a patient: administrative data, clinical, paraclinical, diagnostic, therapeutic and preventive.

It comprises:

- The administrative file: The hospital administration constitutes an administrative file which includes all the elements enabling the patient to be identified: his administrative position, his social security cover, his date of entry into the hospital and his date of release.

- The professional medical file: Every doctor is concerned by keeping this file; he must mention all his observations, his interventions and the hypotheses he formulates in conclusion.
- Nursing record: It is defined as “a unique and individualized document gathering all the information concerning the person being cared about. It considers the preventive, curative, educational and relational aspect of the treatment. It also includes the care plan that should be established with the caregiver and contains information specific to nursing practice”.<sup>[6]</sup>

Now, to increase the quality of care and reduce costs, we have moved from paper-based medical records to computerized medical records called EMR while ensuring a gain for health professionals and patients, a gain in terms of patient service, coordination and comfort exercise for health professionals.

According to ISO/TS183086, the primary purpose of an EMR is to provide a documented record, developed by physicians or others, that supports current care and history. It is also a means of communication between physicians who contribute to patient care.

Since traditional information models do not meet the requirements of MIS and their continuing evolution, advanced standards for representation and communication of EMRs<sup>[7,8]</sup> propose to use an architecture based on the dual model approach. This architecture defines two conceptual levels: (1) the reference model and (2) the archetype model.<sup>[4]</sup>

As well as, different medical informatics standards use the dual model to support interoperability between MIS such as ISO 21731 “Health informatics - HL7 version 3 Reference Information Model” 9, ISO/EN 13606 “Health Informatics - EHR communication”<sup>[3]</sup> and OpenEHR, but they are not adopted because of the lack of tools, methodologies and methods to carry out semantic activities on the content of EMRs in the semantic web environment and incompleteness of standards.

### 2.3 Interest of medical informatics

According to Resnik P,<sup>[9]</sup> medical informatics feeds on research from various fields such as knowledge engineering, model engineering and artificial intelligence, because it is based on formal representations of knowledge, in the form of symbols that the system can store and manipulate. Each of these areas of research provides medical informatics with methods, techniques and tools to improve the formalization of data and knowledge in MIS for better patients’ management.

Knowledge engineering is used to organize medical knowl-

edge. Therefore, systems of organization of knowledge, in these case terminologies, thesauri and ontologies, are products of this engineering.

Vandenbussche P.<sup>[10]</sup> states that these systems make it easier to access and share medical information. They serve as reference resources, which facilitate the understanding and allow a common and unique interpretation of medical information. We use these systems to represent, organize and model medical information.

Vandenbussche P.<sup>[10]</sup> also reports that the models built by the knowledge engineering work contribute to providing the MIS with a behavior and making it relevant and effective in the intended task.

However, patient records as well as medical courses or opposable medical references are written in natural language. The power of the natural language at the same time creates an obstacle to its use for the processing of information. In fact, computer systems have difficulties in the presence of the construction of new concepts. They tend to interfere with the ambiguity of some information, especially that MIS do not deal with the semantic side of information from where it comes from the notion of ontologies.

### 2.4 Use of ontologies in the medical field

Most of the use of ontologies was in the field of information retrieval; Jean Charlet<sup>[11]</sup> with his partners developed ON-TOLURGENCES under 6 stages, a termino-ontological resource that ensures:

- the role of the domain model listing all relevant concepts;
- the link between the concepts and how they are named in the Computerized Patient Record documents.

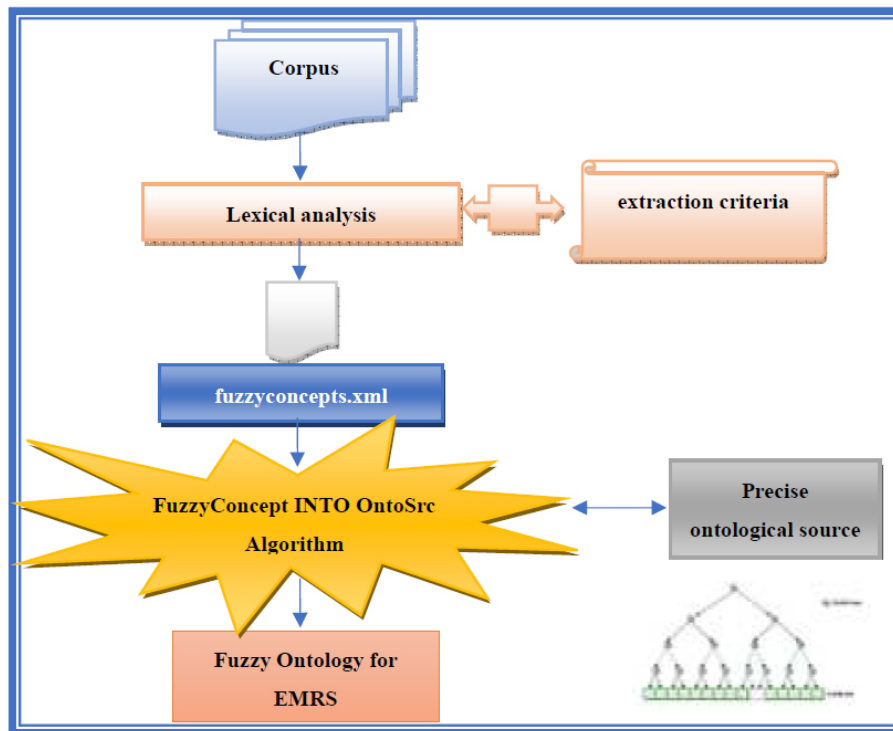
This dual feature allows annotation and indexing of patient records and searching for information in indexed folders. Sonia and her partners, in their conference paper,<sup>[12]</sup> built, using knowledge engineering tools, an ontology using the textual resources found in the corpora, via the automatic processing of the natural language. This is to exploit the database, already created by a regional network of coordination of pathways of patients with Amyotrophic Lateral Sclerosis, which traces the requests, the needs of the patients and the actions put in place to answer those needs. The aim was to understand, describe the health pathways and identify points of weakness and levers to improve to provide the continuity of routes and prevent breaks.

The starting point of this work was the work done by SAMET ELLOUZE et al.<sup>[4]</sup> who proposed a methodology for modeling EMRS in accordance with the openEHR standard, a

methodology for integrating the semantic dimension into these principles. DME, a time dimension integration approach in EMRS as well as an approach for generating medical interfaces that are sensitive to user contexts. Their work aims to provide levers to build efficient SIMs for better care quality. They ensured these levers through the development of an original methodology, called openEHR-MM, within a uniform framework of cooperation between the four technological spaces: the MDA space, the syntactic space, the semantic web space and the XML space. What we are interested in is the semantic web space ie, ontology, to promote

semantic interoperability between EMR systems. To govern this space, the authors have defined an original methodology, called “semantic OpenEHR”, allowing the integration of the semantic dimension into EMRS modeled according to the openEHR standard.

Thanks to the mechanisms offered by SemanticOpenEHR, they have been able to obtain an ontological source (see Figure 1) that can serve as a knowledge base that can support the semantic interoperability between EMRS, the deduction of new data knowledge and clinical decision-making.



**Figure 1.** Fuzzy Semantic OpenEHR Overall Scheme

As part of this work, we will benefit from the work of Samet Ellouze et al.<sup>[4]</sup> However, his work has, among other things, limitations. For example, the medical file mentioned in it contains inaccuracies and uncertainties in the information it contains. This problem of uncertainty is subsequently resolved by applying fuzzy logic to ontologies. That is why we thought about extending this work.

**2.5 Fuzzy ontologies**

Fuzzy ontologies are an extension of the domain of crisp ontologies for solving the uncertainty problems. Current fuzzy ontological models do not focus on essential semantic relationships between fuzzy concepts, which lead to difficulty in ontology integrating. To represent formally the fuzzy knowledge, we applied the fuzzy logic<sup>[13]</sup> and we proposed

an appropriate model.<sup>[14]</sup>

Given its promising nature, fuzzy ontology has been used in various fields of computer science research, such as knowledge management, data integration and information retrieval as well as the medical field since this field presents uncertain and subjective data, they integrated the fuzzy logic to give better results.

The use of fuzzy logic and fuzzy ontologies is attracting more and more interest from the scientific community. Their contributions are shown by works in the medical field such as:

- Fuzzy ontologies are used for the automatic generation of medical diagnoses. The input of this system are medical documents, it consults a fuzzy ontology to

understand the results, and automatically generates a pre-diagnosis.<sup>[15]</sup>

- The paper describes the rational and initial testing of a system for collaborative research and ontology construction for occupational groups in the health sector. The approach is based on the use of a browser using a fuzzy ontology based on the Unified Medical Language System (UMLS) of the National Library of Medicine. This approach can provide high quality information for professionals in the future since evidence-based medicine requires appropriate information to be available to clinicians at the point of care.<sup>[16]</sup>
- To represent semantically medical data relating to the diagnosis and management of alzheimer's disease, the authors proposed a fuzzy ontology "AlzFuzzy-Onto" relating to the specific concepts of the disease of alzheimer.

Fuzzy ontologies can resolve many problems in Semantic Web, but it has a complex model. As far as that goes, we must then find the means to ensure their correct building. We find a methodology for constructing fuzzy ontologies called "FuzzyOnto Methodology".<sup>[14]</sup> However, we will not use the latter because, in this methodology, we must begin with the phase of constitution of the corpus and specification of the field of knowledge. Whereas in our case, our starting point is ontology i.e., we already have an ontological source and we are going to enrich it with new fuzzy concepts. This is the reason for which this methodology seems to be useless.

### 3. METHODS

Our starting point is a corpus of semantic data containing medical information. We can find, in such a medical document, different specialties involved to communicate in the treatment of the corresponding disease. For example, in the treatment of the disease "cerebrovascular accident", the cardiologist, the neurologist, and the angiologist (vascular doctor) intervene together. Also, when treating the disease "Venous thrombosis", the pulmonologist, the radiologist, the Doppler ultrasonographer and the cardiologist must intervene. The diversity of specialties in the same medical document requires standardized vocabulary between them, given the uncertainties and inaccuracies that may exist in terms of its medical information. Our approach, titled Fuzzy Semantic OpenEHR, is divided into three stages as follows (see Figure 1):

- Lexical analysis of the semantic data corpus (extraction of fuzzy concepts);
- Processing the XML file;
- Insertion of fuzzy concepts into the precise ontological source.

In the first step, it is a matter of a lexical analysis of the body of data that we will enter under a parser. This analysis is based on a grammar made up of a set of rules.

In the second step, we process the XML file resulting from the previous step, using an algorithm to read this file and display the fuzzy extracted concepts.

As for the third step, we will fuzzify the already existing ontological source<sup>[4]</sup> by enriching it with the new fuzzy concepts that we extracted during the first and second stages.

#### 3.1 Lexical analysis of the semantic data corpus

The first step is to process this corpus following a few steps using the experts in the field. First, we will introduce it under a lexical analyzer "NooJ" (linguistic corpus tool; downloadable from the website: <http://www.nooj-association.org/>)<sup>[18]</sup> for the processing of corpora of fuzzy semantic data. Indeed, NooJ is a system of automatic processing of natural languages. Specifically, it is a language development environment that includes dictionaries, grammars, and corpus analyzes. This tool allows users to process large sets of text in real time. Users can build, accumulate and control sophisticated matches that match morphological and syntactic grammars organized in reusable libraries.

The rationale for choosing NooJ lies in the following characteristics that should rather interest to meet the needs of uses:<sup>[14]</sup>

- Its corpus processing engine uses high lexical and syntactic language resources. This allows its users to perform sophisticated queries that include one of the morphological, lexical, or syntactic properties available.
- NooJ can directly process potentially large sets of documents, which can be text, in any format: Web pages (HTML), XML documents, Microsoft Word, etc.
- NooJ's graphical editor contains a dozen of development tools for editing, testing and debugging local grammars, organizing them into libraries and applying them to texts, either as queries or to add (or filter) annotations.
- Project concept: A project is a given state of the NooJ environment configuration which can be saved and recalled at will.

There are other analyzers like:

- Unitex (downloadable from the website: <http://unitexgramlab.org/>), developed by the LADL, Maurice Gross Laboratory: it is an automatic language processing software using many linguistic resources (dictionaries) allowing the processing of large corpora.

It allows the matching of terms from regular expressions or graphs (also called local grammars).

- INTEX (downloadable from the website: <http://intex.univ-fcomte.fr/>) is a language development environment that includes large-coverage dictionaries and grammars and analyzes texts of millions of words in real time. It includes tools to create and maintain lexical resources with wide coverage, as well as morphological and syntactic grammars. Dictionaries and grammars are applied to texts to locate morphological, lexical and syntactic profiles, eliminate ambiguities and mark simple and compound words. It is used by several research centers to quickly build extractors to identify semantic units in large texts, such as the proper names of people, locations, technical expressions of finance, and so on.<sup>[18]</sup>

Compared to INTEX, NooJ uses new technology (.NET), a new language engine, and has been designed with a new range of applications in mind. NooJ's architecture is based on the .NET "Component programming" technology, which goes beyond the object-oriented approach.<sup>[18]</sup> This architecture offers several advantages: It allows NooJ to read any document that can be managed on the user's computer. For example, on a typical MS-Windows computer, NooJ can handle bodies in more than 100 file formats, including all variants of ASCII, ISO and Unicode, HTML, RTF, XML, MS-WORD, and so on. However, NooJ does not deal with the semantic part of words; he then makes errors in extracting words from the data corpora. To solve this problem, we have created our own data dictionary where we associate with each word one and only one grammatical role: noun, adjective or adverb. Indeed, we can determine the basic criteria (i.e., which represent the inputs of the parser's grammar) of the concepts from their definition according to the language of the parser. Based on these criteria, the parser automatically extracts concepts from corpora.

We have built a textual grammar that brings together a set of rules that we will apply to the corpus. Just go to the NooJ menu and select File -> New -> Grammar then Inflection & Derivation, we will have an interface to type our grammar and we introduce our set of rules. In what follows, we will present, explain and select the desired language, click more on these fuzzy concepts extraction rules. In fact, fuzzy concepts can be defined from linguistic variables (example: disease) and the fuzzy values (example: "normal", "deadly" ...). These concepts are expressed by adjectives and adverbs (example: mortal disease) indicating inaccuracy in the margin corresponding to a precision (having a life expectancy of 2 years, 3 years, 5 years?). Subsequently, we can deter-

mine the basic criteria of the concepts (i.e., which present the entries of the analyzer's grammar) from their definition according to the language of the analyzer. We propose the "Name" as a first input of the grammar of the parser. The second entry must represent the vagueness and uncertainty of its fuzzy property that are generally expressed by adjectives or adverbs, ie fuzzy sets will be represented by adverbs and adjectives. Now we can define the extraction criteria of fuzzy concepts as follows: "Name + Adjective"; "Adjective + Name"; "Name + Adverb + Adjective" and "Adverb + Adjective + Name" (as for example the fuzzy concept: "very high cholesterol").

In fact, the grammar is a set of graphs and there are three different types of graphs under NooJ:

- Morphological: to analyze the morphology of a word (with the suffix. nom).
- Inflection: to analyze inflections of a word (with the suffix. nof).
- Grammar: to analyze the texts (with the suffix. nog).

We will then apply our grammar to the corpus to extract all the fuzzy concepts that exist in this corpus. Just follow the following steps of our approach: (1) open the corpus, (2) click on the right button, (3) click on "Linguistic Analysis". Subsequently, the launch of the grammar is done by the option "Locate Pattern" from the drop-down menu obtained by a right click on the text window.

We can extract the resulting result as an XML document. Subsequently, using an algorithm, we treat this document to be able to insert these new fuzzy concepts identified in the ontological source already existing, that were created during the methodology "Semantic OpenEHR".

### 3.2 Algorithm for processing the XML file

After extracting fuzzy concepts from an XML file, we defined an algorithm called "FuzzyConcept INTO OntoSrc" with the following steps:

- S1) Analyze the XML file generated by the lexical analyzer;
- S2) Validate extracted concepts;
- S3) Search for synonyms of fuzzy concepts;
- S4) enrich the ontological source.

#### 3.2.1 Analysis of the generated XML file

The first step (S1) consists in processing the XML file generated by the NooJ analyzer in order to display the fuzzy concepts in a table. Once the fuzzy concepts are presented in the list, the ontologist can perform the following actions for each concept:

- Addition of a concept that was not recorded in the XML file during the extraction step,
- The deletion of a concept deemed unnecessary (or invalid) for the fuzzy ontology,
- Or the modification of a concept.
- Represent entities in expert systems and decision support systems;
- Index the documents, especially the medical literature;
- Serve as an “interface” support for the collection of information.

### 3.2.2 Validation of extracted concepts

Once the fuzzy concepts are validated, the ontologist must specify for each concept: (1) the type of its membership function and (2) indicate the attributes of the linguistic variable associated with it. Indeed, each fuzzy concept is expressed through its proper membership function. We find the following most used membership functions: the triangular function (with 3 variables), the trapezoidal function (with 4 variables), the increasing monotonic function (with 2 variables), and the decreasing monotonic function (also with 2 variables). In fact, the values are to be taken into consideration when calculating the degree of belonging.

For instance: the concept of “life-threatening illness” is defined using a membership function defined on the attribute “life expectancy”. The term “life expectancy” is a vague attribute that has the potential value of the “deadly” linguistic term. We then associate a function of belonging, of trapezoidal type to the linguistic value “mortal”.

### 3.2.3 Inserting fuzzy concepts into the ontological source

The next step of the algorithm (S3) concerns the search for synonyms of the selected fuzzy concept, this is through the consultation of a terminological system and it is to further explain the meaning of the concept. It is therefore sufficient to write in the search area of this browser the fuzzy concept that we wanted to find its synonyms, then a list of concepts is displayed as a result (the synonyms of the concept). The last step of the algorithm is to find the exact place of the concept through one of the synonyms found and therefore the ontologist offers a better position relevant to insert this concept into the ontological source.

It is known that medical language is characterized by an extremely rich vocabulary that is difficult to manipulate and in which the terms used are often imprecise and rarely subject to rigorous definitions. In this type of language, there are several ways to express the same thing (synonyms). As a result, the medical community has been interested in structuring clinical knowledge and developing terminological systems. These systems are the fruit of the work of knowledge engineering, and more precisely knowledge organization systems. Many medical terminologies exist and each of them has been created to meet a need. These needs are concretized mainly in the following points:<sup>[19]</sup>

- Code information, especially that relating to a patient, for care or public health;

A terminology system links the concepts of a domain and provides terms, definitions and codes. It can take the name of: terminology, thesaurus, vocabulary, nomenclature, classification, taxonomy or ontology.<sup>[19]</sup> There are several terminologies such as: ICD-10 (International Classification of Diseases, available at: <http://www.who.ch/hst/icd-10/icd-10.htm>), SNOMED-CT, MeSH (Medical Subject Headings, available from: <http://www.ncbi.nlm.nih.gov/mesh>), and LOINC (Logical Observation Identifiers Names and Codes, available from: <http://loinc.org/>). We chose to work with standard health terminology that contains more than 350,000 “Systemized Nomenclature of Medical - Clinical Terms” (SNOMED-CT) concepts because it is the most detailed health terminology to describe an EMR to date.<sup>[19]</sup> It combines a hierarchical organization and a compositional representation of concepts that are of two types: “primitive” or “completely defined”.<sup>[20]</sup> The concepts “completely defined” represent all the concepts that can be differentiated from their concepts parents and brothers by the relations with other concepts. It provides formal definitions for the different concepts using the “is-a” relationships and assigned relationships. These relationships have three main objectives: to clarify the semantics, to automate the classification and to allow the postcoordination. Relationships formally reflect the semantics of the concept. Indeed, SNOMED-CT does not give a textual definition of concepts, but rather aims to formally specify the properties of the concept. For example, in SNOMED-CT, “Pneumonia” is defined as a “lung disease”.

### 3.2.4 Enrichment of the ontological source

This last step consists in adding the fuzzy concepts after their validations one by one. The ontologist at this stage, either knows the meaning of the concept and goes directly to find its exact place between the displayed concepts of the ontological source (that he should import it already), or if he wanted to have further explanation of this concept, he captures this concept in the browser in English (SNOMED-CT) and in this way, a list of synonyms appears so that he can better understand the meaning of the concept. Then the ontologist will complete the selection of the exact position of this concept according to the concepts displayed (that of the ontological source). Subsequently, the ontologist records this addition and in this way the OWL file of the ontological source will be modified by the addition of these new fuzzy

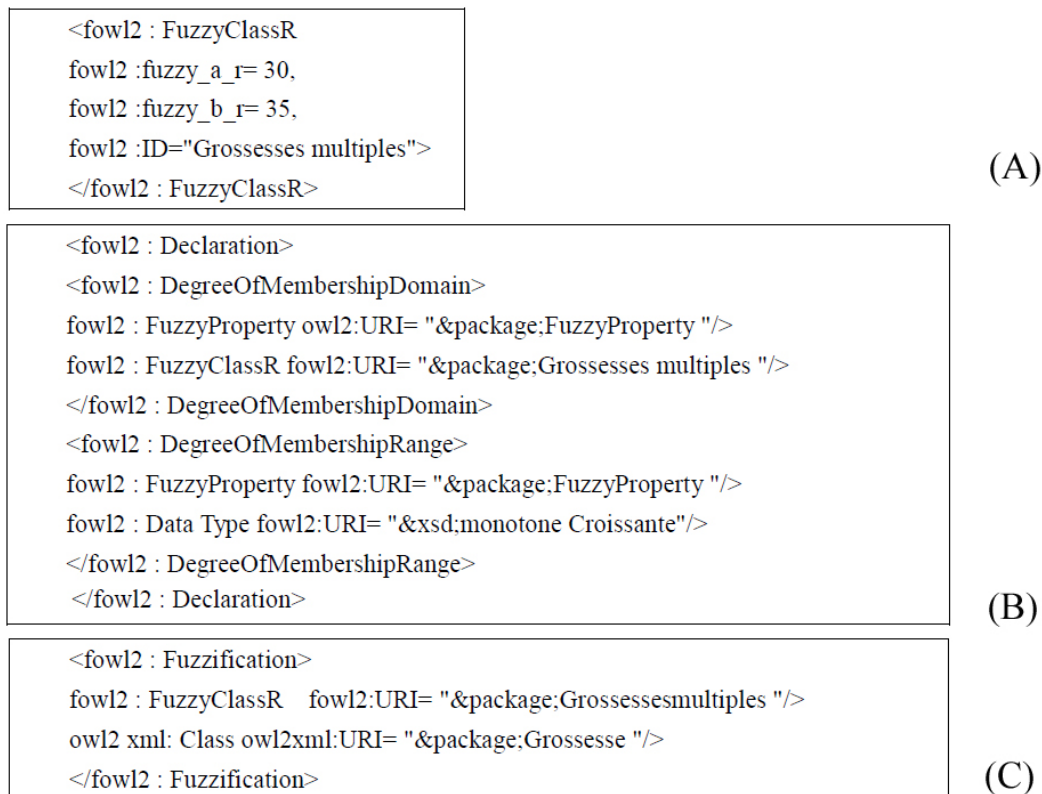


concepts through the Fuzzy OWL 2 language tags.<sup>[14]</sup>

### 3.3 Validation of the ontological source enriched by fuzzy concepts

When we add a new fuzzy concept in the ontological source, this file will be modified through the Fuzzy OWL 2 language, adding: the <Declaration> tag to declare the name of the class (i.e., name of the concept), and the tag <SubClassOf> to declare the relationship between the concept that will be added and the parent concept. As we see in Figure 2, the prologue in the first line of the document then the classic

namespaces (xmlns) and standard prefix names (rdf, rdfs, xsd, owl) of an XML file as well as the identifiers internationalized resources (IRI) to identify ontologies and their elements. We then find the declarations of each element found in the ontological source beginning with the <Declaration> tag followed by the <Class IRI> tag to write the name of the concept preceded by a # such as the declaration of the concept “Admission” is as follows: <Declaration> <Class IRI = “# Admission”/> </Declaration>. Likewise, for the remains of the concepts of the ontological source.



**Figure 2.** (A) Fuzzy OWL 2 code relating to the “multiple pregnancies” fuzziness concept, added in the ontological source, having a growing monotonic membership function; (B) Fuzzy OWL 2 code relating to the “multiple pregnancies” fuzziness concept having a monotonically increasing ownership property; (C) Fuzzy OWL 2 code relating to the fuzzy concept “multiple pregnancies” having a fuzzyfiant specification relationship

## 4. XML TO FUZZY OWL FRAMEWORK

Toward validating the contribution of our proposed approach throughout this paper, we have implemented a framework integrating different modules implementing each of the aspects mentioned in this memory. This platform implements our “FuzzyConcept INTO OntoSrc” algorithm. It processes the XML file resulting from the step of extracting the fuzzy concepts of a semantic data corpus through the NooJ analyzer, then enriches the ontological source with these extracted concepts.

Our platform is integrated into the Eclipse Integrated Development Environment IDE as a plug-in. The choice of Eclipse was dictated by the fact that it is an open source IDE widely used to build open and scalable development platforms consisting of tools and runtimes for building, deploying and managing of software. In this section, we present our framework called “XML 2 Fuzzy OWL” detailing the different modules that make up its architecture. The first basic module is the table that displays fuzzy concepts from the XML file. Then, the second module concerns the attribution of

the membership functions and the values of the linguistic variable to each concept.

Subsequently, the third module deals with the search for concept synonyms through the SNOMED-CT browser and then moves on to the fourth module that adds the validated blurred concepts to the ontological source. Finally, the OWL file of the ontology will be modified. We will also present the tests of the functionalities of our framework realized with the disease “Paralysis cerebral”.

#### 4.1 XML 2 Fuzzy OWL framework architecture

To test the feasibility of our proposal, we proceeded to the realization of a framework called “XML 2 Fuzzy OWL”. It is a tool that allows the processing of the XML file, generated through an analyzer, after extracting and organizing fuzzy concepts from a fuzzy semantic data corpus. So, we have as input a corpus of semantic data.

In our framework “XML 2 Fuzzy OWL” we have four modules:

- The first module concerns the analysis of the XML file already generated by the lexical analyzer NooJ and the display of the fuzzy concepts in a table.
- The second module addresses the attribution of the membership function types and the attributes of the

linguistic variable of each fuzzy concept.

- The third module is devoted to the search for synonyms to explain the meaning of each fuzzy concept.
- The last module ensures the addition of fuzzy concepts that are already valid in the ontological source.

The first, second and third modules correspond to the second stage of our approach. As for the last module, it corresponds to the third stage of our approach.

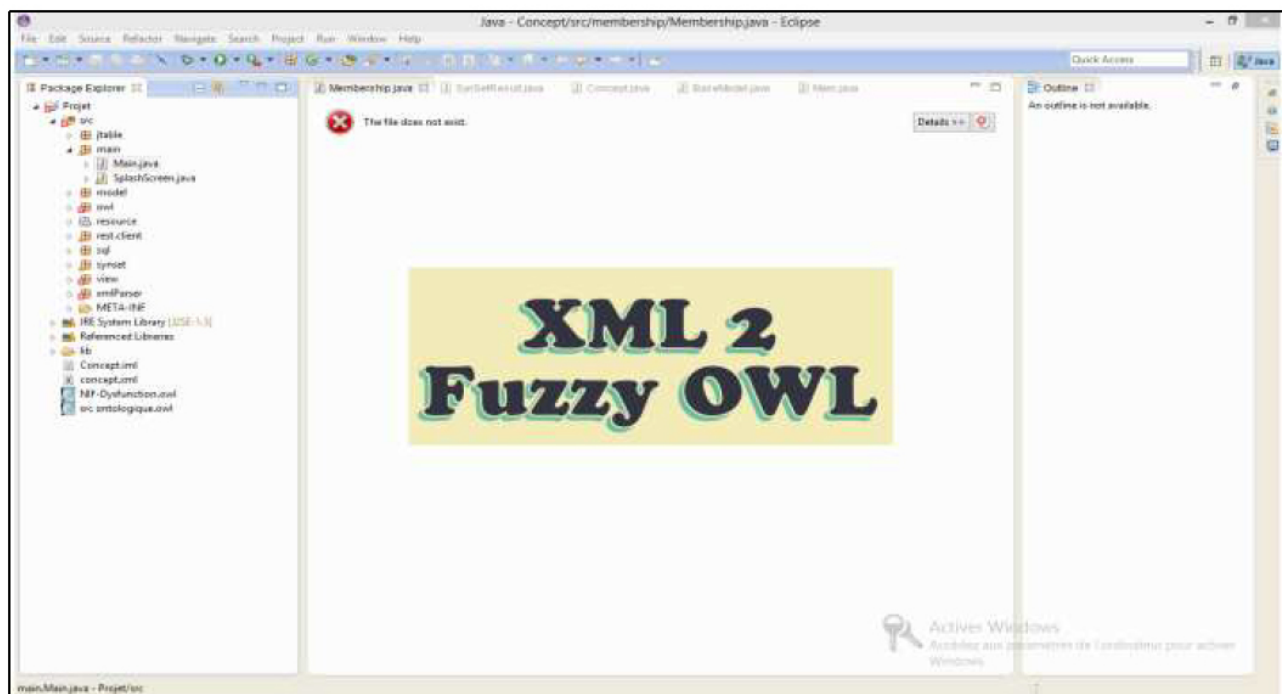
To realize our framework, we chose to use the Java language and the Eclipse integrated development environment. For the design of graphical interfaces, we used the Java Swing package. To save the coordinates associated with each fuzzy concept, we used the Easyphp environment, and for editing ontologies, we used Protégé.

Figure 3 shows the homepage of our framework.

Our framework implements the following features:

##### 4.1.1 Importing concepts

To import the fuzzy concepts for an application (i.e., fuzzy ontology) from an XML file already created in the NooJ part, the ontologist must select, in the File menu of XML Fuzzy OWL, the submenu “open” to choose the XML file to import. We then obtain a table of the concepts of the selected XML file.



**Figure 3.** The homepage of XML 2 Fuzzy OWL framework

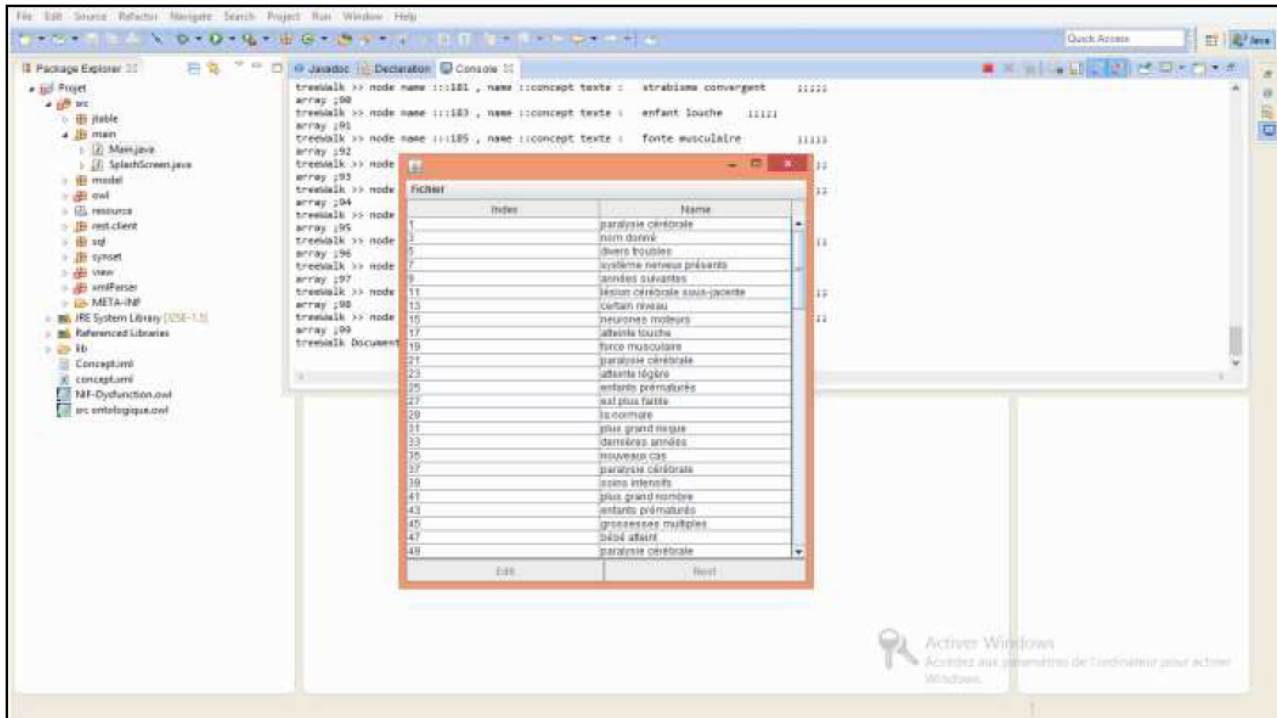


Figure 4. Table of fuzzy concepts

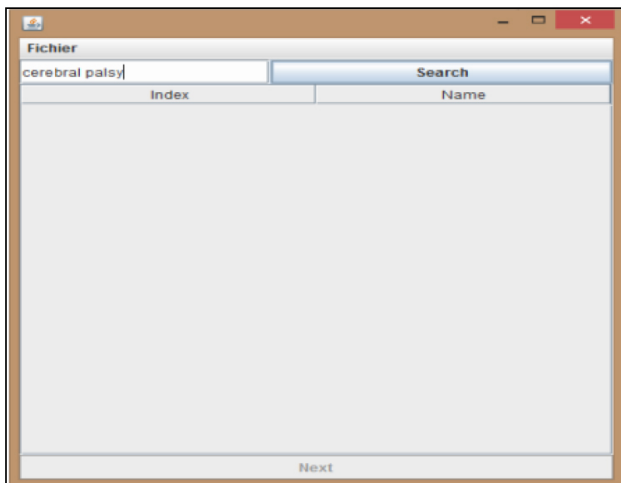


Figure 5. Search for synonyms of the concept entered

**4.1.2 Verification of concepts by the ontologist**

The ontologist can modify or delete concepts after the import step. To do this, he can use the “Edit” and “Delete” buttons, respectively, at the bottom of the table of fuzzy concepts (see Figure 4).

**4.1.3 Definition of linguistic variables**

We propose to the ontologist to specify, for each valid fuzzy concept, the type of the membership function and indicate the list of values of its linguistic variable. To do this, he can use the “Next” button at the bottom of the table of fuzzy concepts to go to the relevant interface.

**4.1.4 Finding synonyms for a fuzzy concept**

We propose to the ontologist to introduce the name of the concept, that he wanted to search for its synonyms (to explain more the meaning of this concept).

By clicking on the “Search” button, the results of the search for synonyms of the concept entered are displayed as shown in Figure 5. This functionality is achieved using the SNOMED-CT ontology.

**4.1.5 Enrichment of the ontological source**

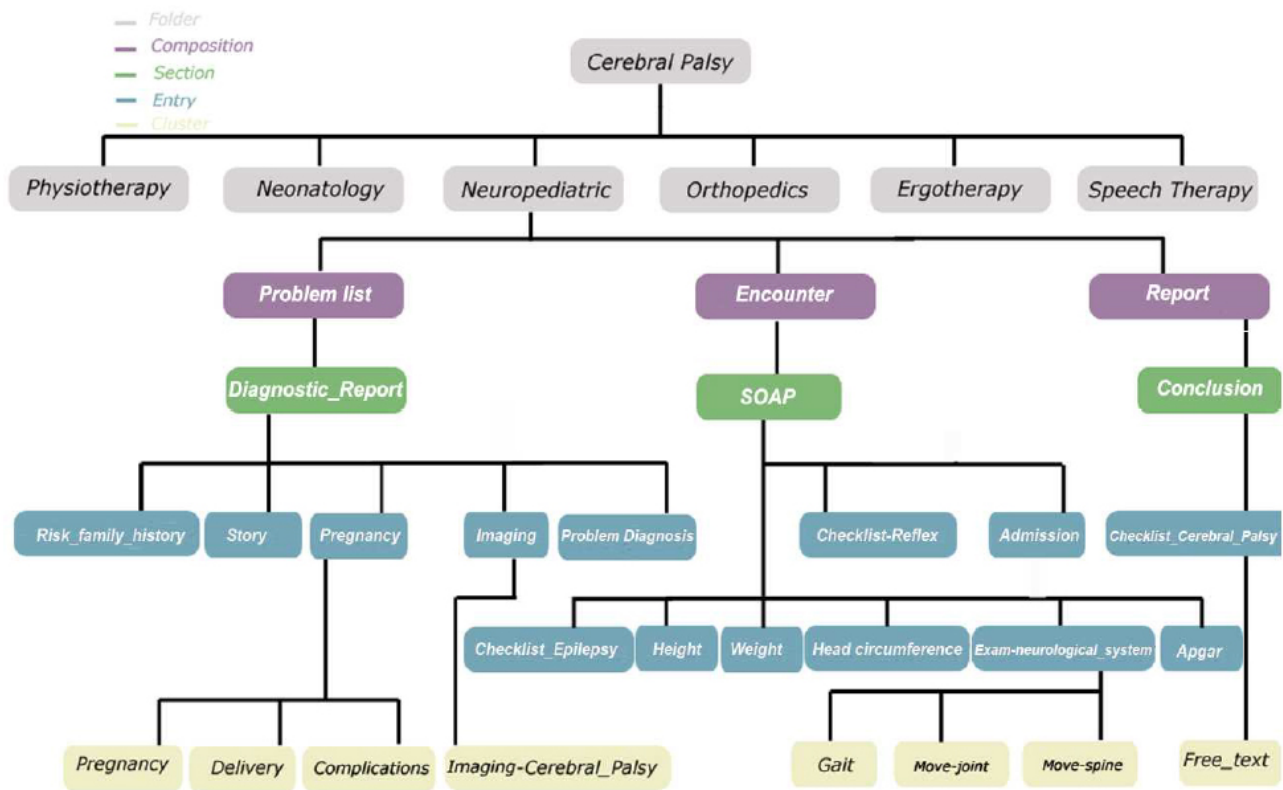
The fourth interface in our framework is that of adding the already valid concepts by the ontologist through the “Add” button located next to the listbox of the ontological source concepts. Just select the desired concept and click on the “Add” button. At this point, the fuzzy concept (son concept) will be added under the selected concept of the ontological source (parent concept). Then, the ontologist must click on the “Save” button to save the addition of this concept in the OWL file of the ontological source. In this way, the OWL file will be modified by adding lines of code through OWL 2 fuzzy language.

**4.2 Experimental study of XML 2 Fuzzy OWL framework**

In this step, we model a “Cerebral Palsy” type application. We define its precise and fuzzy concepts and semantic relations, while defining for the fuzzy elements the formulas for calculating degrees of belonging (the membership func-

tions). Then we evaluate the construction of precise and fuzzy ontological components by the XML 2 Fuzzy OWL framework in relation to this modeling. This evaluation aims

to determine the percentage of automation provided by this framework. This application represents an extension of the work presented.<sup>[4]</sup> Figure 6 shows our starting point.



**Figure 6.** Hierarchical structure of archetypes related to the neurological folder class

In this step, we will describe why we chose to work with the “Cerebral Palsy” disease related to pediatric neurology department of university hospitals. This illness requires a six-specialty intervention in its treatment as the physiotherapist, the occupational therapist, the speech therapist, the orthopedist, the neonatologist, the pediatric neuropathist. Hence came the need to standardize the concepts used. For example, the specialty of physiotherapy can reinvigorate, occupational therapy helps children develop practical skills, and speech-language pathologists can help solve language and diet problems. Sometimes orthopedic surgery and medication are part of the treatment as well. It is to standardize the vocabulary between the speakers that we chose cerebral palsy to work-on. We wanted to escape from falling into the problem of uncertainties and inaccuracies.

The semantic data corpus (the input of our framework), is a text, unspent, which describes the “Paralysis cerebral”. This disease affects children with Cerebral Palsy. It results from early cerebral lesions (two-year conception according to G. Tardieu), non-hereditary, stabilized (non-progressive), responsible for exclusive or predominant motor deficiencies.

The secular motor handicap associates varying degrees of posture and movement disorders. Specific disorders of the higher functions can be associated with it. This disease is related to a brain injury that occurred in the antenatal or perinatal period. It is a non-progressive motor disorder secondary to a defect or injury on a maturing brain. Specific disorders of the higher functions can be associated with it (perceptual disorders, praxic disorders - in relation to the representation and internalization of a succession of acts, sensory disturbances). However, brain damage has sufficiently preserved the intellectual faculties to allow schooling. This definition excludes children with mental retardation associated with motor disorder of cerebral origin, which is grouped under the term IMOC (Impairment Motor of Origin Cerebral) and children with multiple disabilities. Children with a motor disorder of brain origin represent 2.14/1,000 newborns, BMI in the strict sense 0.6/1,000. The incidence of this pathology has changed little over the past 10 years, despite the stricter surveillance of pregnancies and perinatal care (since 1970) and the reduction in prematurity (6.8% of births in 1975, 5.6% in 1981).

**4.2.1 Building Fuzzy Ontological Components in XML 2 Fuzzy OWL**

In this section, we will test our framework by calculating the percentage of automating the construction of an ontology using our XML 2 Fuzzy OWL framework for the construction of its components, which gives the results below:

After the import of the corpus, we proceeded to the automatic extraction of the fuzzy concepts using NooJ. The results of this extraction are saved in an XML file. Subsequently, their automatic import from this file allowed to display them in a table. To evaluate these results, we present below the statistical reports of the fuzzy concept development step in the table. We use three indicators:<sup>[21]</sup>

(1) Percentage of valid fuzzy concepts, which represents the percentage of fuzzy concepts developed correctly in relation to our ontological scheme.

Percentage of Fuzzy Concepts (FC) Valid = Number of Valid FC/Total FC

(2) Percentage of unnecessary fuzzy concepts, which represents the percentage of superfluous fuzzy concepts (i.e., identified in the XML file and not included in the ontology scheme) in relation to our ontological schema.

Percentage of Unclear Fuzzy Concepts (FC) = Unnecessary FC/Total Number of FC

(3) Percentage of missing fuzzy concepts: It represents the percentage of fuzzy concepts that have not been identified in the XML file compared to our ontology schema.

Percentage of Fuzzy Concepts (FC) Missing = Number of FC missing/Total FC

We present these percentages in Table 1.

**Table 1.** Experimental study of XML 2 Fuzzy OWL framework Total Instance type Number

Total	Instance type	Number	Percentage
94	Number of valid fuzzy concepts	62	66.95%
	Number of unnecessary fuzzy concepts	21	22.34%
	Number of missing fuzzy concepts	0	0%

Indeed, according to these results the percentages of useless fuzzy concepts are high. This comes down to the naivety of the extraction rules. Therefore, we plan in our future work to build filters to use after the extraction step to improve these results.

The average percentage of automatic fuzzy concepts extraction using XML 2 Fuzzy OWL is therefore about 66%.

We see that through the new fuzzy ontological source that we build using our XML 2 Fuzzy OWL framework, we can create fuzzy knowledge bases from which intelligent systems in the cited domain. These systems will ensure semantic interoperability and better support the uncertainties and imprecisions in the field.

**4.2.2 Results and discussions**

The author presents a study made for the extraction of a domain ontology using Text2Onto. The latter provides a satisfaction percentage of 11%.<sup>[22]</sup>

The average percentage of obtaining accurate and fuzzy ontological components correct by their proposed Text2FuzzyOnto framework is about 54%.<sup>[21]</sup>

After this comparison, we can say that our XML 2 Fuzzy OWL framework provides a satisfaction percentage of about 66%. We find that this percentage is satisfactory for the

ontologists, although it is a percentage of automation of the fuzzification of an important ontological source.

**5. DISCUSSION AND CONCLUSIONS**

In this paper, we presented an approach for the enrichment of an ontological source by new fuzzy concepts extracted through a lexical analyzer (NooJ). This approach is called Fuzzy Semantic OpenEHR. It presents an extension of Semantic OpenEHR’s methodology for building precise ontologies: a methodology for modeling EMRs in accordance with the openEHR standard, a methodology for integrating the semantic dimension into these EMRS.<sup>[4]</sup> The efficiency lever consists of the semantic interoperability and the fuzzy dimension; this is the lever that we have ensured via our approach.

Our approach is composed of three steps that are:

- Extraction of fuzzy concepts from a corpus: we used a NooJ lexical analyzer to analyze a corpus of semantic data and then extract fuzzy concepts according to a set of rules that we have already established. The result of this analysis is an XML file containing an enumerated list of fuzzy concepts.
- Processing the XML file through an algorithm called “FuzzyConcept INTO OntoSrc”: at the beginning, the first sub-step of this algorithm is used to analyze this

XML file generated by the parser, then, the validation step of these concepts by attributing each concept validates its type of membership function as well as the values of its linguistic variable. Subsequently, we recorded this information related to each concept in a database. The next sub-step of the algorithm consists in inserting the fuzzy concepts already valid in the ontological source, this is using the consultation of a “SNOMED-CT” terminology system which has already been chosen thanks to its character promoter of a content of 350,000 medical concepts.

- At the end, the last sub-step of the algorithm is to enrich the OWL file of the ontological source through the Fuzzy OWL 2 language by adding the tags to declare the new name of the concept as well as the relationship between the added concept and the parent concept.

We, then, implemented this algorithm to test the feasibility of our proposal. We proceeded to the realization of a framework called “XML 2 Fuzzy OWL”. The input of our framework is an XML file generated by the lexical analyzer during the step of extraction of the fuzzy concepts. The output is the OWL file of the ontological source that has been modified to follow an addition of a fuzzy concept. As for the architecture of our framework, it is composed of four modules which are:

- The first basic module is the table displaying the fuzzy concepts extracted from the XML file.
- The second module concerns the attribution of the membership functions and attributes of the linguistic variable to each concept.
- The third module deals with the search for synonyms of concept through the consultation of the SNOMED-CT navigator if the ontologist does not understand the meaning of the selected fuzzy concept. In the other case, the ontologist is not obliged to consult this terminology.
- The fourth module makes it possible to add the fuzzy concepts validated in the ontological source. In this way, the OWL file of the ontology will be modified through Fuzzy OWL 2 language.

After we proceed with a test for our framework through the percentage calculation of the automation of the construction of an ontology using “XML 2 Fuzzy OWL” for the construction of its components. Then we tried to evaluate this percentage and then to judge the value of our framework compared to another framework. We concluded that the average

percentage of the automatic census of fuzzy concepts by our framework is about 43%, while “Text2Onto”,<sup>[22]</sup> provides a satisfaction percentage of 11%, and “Text2FuzzyOnto”<sup>[21]</sup> the authors get a satisfaction percentage of about 54%. We can say then that this percentage is satisfactory for ontologists, although it is a percentage of automation of the fuzzification of an important ontological source.

We envision several avenues of work to improve the quality of our approach to enrich an ontological source by new fuzzy concepts extracted through an analyzer. Our first short-term perspective is to check the percentage of automation of our “XML 2 Fuzzy OWL” framework with another disease other than “Cerebral Palsy”. Also, we can improve our proposed solution by automating the task of integrating fuzzy concepts into the ontological source, and this is to lighten the work of the ontologist since in our approach, this task is semi-manual. Concerning the OWL file of the ontological source that has been modified after the additions of the fuzzy concepts, we plan to create a fuzzy ontology validation process which allows to validate syntactically, semantically and conceptually the fuzzy ontological model described in the form of the Fuzzy OWL 2 code, which must have many validation rules. Another perspective to improve the approach is to add fuzzy concepts to the ontological source by using formulas for calculating similarity between the two concepts that exist in the literature, otherwise we can reformulate one of these formulas to be functional in our case.

Finally, we envisage another long-term perspective on improving the performance of our extraction process, from the fuzzy semantic data corpus, inheritance relations (between precise concepts, fuzzy concepts, a concept blur and a precise concept), which are manifested by the keywords “is a”. This is through a parser and the extraction of semantic relations (i.e., synonymy, antonymy, equivalence, homonymy and taxonomy). Since we have noticed a deficiency at the level of ontological components fuzzy; it lacks, among other things, these relationships.

#### ACKNOWLEDGEMENTS

We would like to thank Professor Chahinez Triki, specialists in the Neurology Department in Hedi-Chaker-Sfax-Tunisia University Hospital for her valuable intervention in the process modeling of cerebral palsy EMR.

#### CONFLICTS OF INTEREST DISCLOSURE

The authors declare they have no conflicts of interest.



## REFERENCES

- [1] Dusserre L, Ducrot H. L'informatique médicale. Presses universitaires de France. 1985.
- [2] European Community. Semantic interoperability for better health and safer healthcare. Deployment and Research Roadmap for Europe. 2009.
- [3] International Standards ISO 13606-1. Health informatics — Electronic health record communication — Part 1: Reference model. First edition. 2008.
- [4] Ellouze AS, Bouaziz R, Ghorbel H. Integrating semantic dimension into openEHR archetypes for the management of cerebral palsy electronic medical records. *Journal of Biomedical Informatics*. 2016; 63: 307-324. PMID: 27568295. <https://doi.org/10.1016/j.jbi.2016.08.018>
- [5] Romeyer C. Système d'information fondé sur une traçabilité des activités: Intérêt et difficultés de mise en œuvre dans les hôpitaux. Thèse de doctorat: Université de la Méditerranée: Aix-Marseille II; 2002.
- [6] Agence Nationale d'Accréditation et d'Evaluation en Santé. Evaluation des pratiques professionnelles dans les établissements de santé; Dossier du patient: amélioration de la qualité de la tenue et du contenu réglementations et recommandations. Paris. 2003.
- [7] Blobel BG. Advanced EHR architectures—promises or reality. *Methods of Information in Medicine*. 2006; 45(1): 95-101. PMID: 16482378. <https://doi.org/10.1055/s-0038-1634044>
- [8] Eichelberg M, Aden T, Riesmeier J, et al. Electronic health record standards—a brief overview. 4th international conference on information and communications technology, Cairo, Egypt. 2006.
- [9] Resnik P. Semantic similarity in a taxonomy: an information based measure and its application to problems of ambiguity in natural language. *Journal of Artificial Intelligence Research*. 1999; 11: 95-130.
- [10] Vandebussche P. Définition d'un cadre formel de représentation des Systèmes d'Organisation de la Connaissance. Thèse de doctorat. 2011.
- [11] Charlet J, Declerck G, Dhombres F, et al. Construire une ontologie médicale pour la recherche d'information: problématiques terminologiques et de modélisation. 23es journées francophones d'Ingénierie des connaissances (IC' 2012). Paris, France. 2012; 33-48.
- [12] Cardoso S, Aimé X, Mora L, et al. Les ontologies aider à comprendre les parcours de santé dans le cadre des maladies neurodégénératives. Conference Paper. Conference: IA & Santé 2016 - Deuxième Atelier sur l'Intelligence Artificielle et la Santé, à Montpellier. 2016.
- [13] Zadeh L. The Concept of a Linguistic Variable and its Application to Approximate Reasoning. *International Journal of Information Science*. 1975; 4(4): 301-357. [https://doi.org/10.1016/0020-0255\(75\)90046-8](https://doi.org/10.1016/0020-0255(75)90046-8)
- [14] Ghorbel H, Bahri A, Bouaziz R. Fuzzy Ontologies Building Platform for Semantic Web: FOB Platform. IGI Global, Next Generation Search Engines: Advanced Models for Information Retrieval (Chapter 5). 2012; 92-113.
- [15] Sanchez E, Yamanoi T. Fuzzy Ontologies for the Semantic Web. Proc. of the 7th International Conference on Flexible Query Answering Systems (FQAS 2006). 2006; 691-699. [https://doi.org/10.1007/11766254\\_59](https://doi.org/10.1007/11766254_59)
- [16] Parry D. Evaluation of a Fuzzy Ontology-Based Medical Information System. Auckland University of Technology, New Zealand. 2009. <https://doi.org/10.4018/978-1-60566-050-9.ch080>
- [17] Zekri F, Ghorbel H, Bouaziz R. A decision support system based on fuzzy specialized rules for the Alzheimer disease. Proc. of 11th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD'2014). Xiamen, China. 2014; 490-496.
- [18] Silberztein M, Tutin A. Nooj, un outil TAL pour l'enseignement des langues. Application pour l'étude de la morphologie lexicale en FLE. *International Journal of Apprentissage des langues et systèmes d'information et de communication*. 2005; 8(3): 123-134.
- [19] Venot A, Burgun A, Quantin C, et al. Informatique médicale, e-santé – fondements et applications. ISBN: 978-2-8178-0337-1. 2013.
- [20] Spackman K. SNOMED RT and SNOMED CT: promise of an international clinical terminology. *MD Computing*. 2000. PMID: 11189756.
- [21] Ghorbel H, Maalej S, Bouaziz R. Un framework pour la génération semi-automatique d'ontologies floues: Text2FuzzyOnto. *Technique et Science Informatiques (TSI)*. 2013; 32(6/2013) : 671–699.
- [22] Hatala M, Gasevic D, Siadaty M, et al. Can Educators Develop Ontologies Using Ontology Extraction Tools: An End-User Study. Proc. of the 4th European Conference on Technology Enhanced Learning: Learning in the Synergy of Multiple Disciplines. 2009; 140-153. [https://doi.org/10.1007/978-3-642-04636-0\\_15](https://doi.org/10.1007/978-3-642-04636-0_15)