A Knowledge-based Clinical Decision Support System for the diagnosis of Alzheimer Disease

Eider Sanchez, Carlos Toro and Eduardo Carrasco
eHealth and Biomedical Applications Department
Vicomtech-IK4
San Sebastian, Spain
esanchez@vicomtech.org

Patricia Bonachela and Carlos Parra
UCAi Group
University Hospital Virgen del Rocío
Seville, Spain

Gloria Bueno
VISILAB Group, ETSII
University of Castilla-La Mancha
Ciudad Real, Spain

Frank Guijarro
R&D Department
Bilbomatica
Bilbao, Spain

Abstract— **Alzheimer Disease (AD) has become a major issue in developed countries due to medical advances that have extended the population longevity. Recent advances in early detection date the initial stages of AD several years before the first recognizable symptoms appear visible.**

While at present time, there has not been recognized a single cause for AD, the common approach to support the diagnosis is based on diagnostic image processing, psychological tests, neurological tests, etc. This method produces a large amount of data that has to be taken into account by the physicians when they perform their diagnosis.

In this paper we present a Knowledge Engineering diagnosis-support tool for the detection of AD where ontologies and semantic reasoning play a fundamental role. Our work is intended to aid physicians in the early detection of AD by using multidisciplinary knowledge gathered, and inference and reasoning over the underlying Knowledge Bases.

A test example of our tool is also shown and discussed.

Keywords- Decision support systems; Computer aided diagnosis; Knowledge Based systems

I. INTRODUCTION

Alzheimer Disease (AD) has become a major issue in developed countries, where medical advances have extended the population longevity [1]. AD affects 10% of the senior population and also more than 25 million individuals around the world [2]. Thus, the socio-economic impact of AD is huge and attempts to decrease such impact are being searched. For this reason, early detection has grown to be a major research topic, as it can contribute to a better understanding of the disease, and the search for a more reliable diagnostic techniques and efficacious therapies [1].

Recent advances in early detection date the initial stages of AD as early as 15 years before the first recognizable clinical symptoms appear visible [1]. While at present time, there has not been recognized a single cause for AD, the common approach to support the diagnosis is based on diagnostic image processing, psychological tests, neurological tests, etc.

This approach generates a large amount of parameters. All of these parameters (multidisciplinary in essence) have to be taken into account by the physician in order to support possible diagnosis.

Arguably, the amount of data is so large and the advances on the state of the art are so recent that making a proper diagnosis becomes a problem of knowledge handling. Knowledge Engineering (KE) is then an interesting approach to be used.

Among the possible KE techniques, semantic technologies and web inspired paradigms have become one of the most promising fields where possible solutions for the knowledge-handling problem described above, could be encountered [3], [4]. Semantic technologies are quite promising due to their high adaptability, robustness and reasoning capabilities which could provide feasible tools for medical diagnosis as shown by Segev *et al.* [5] and by Gnanambal *et al.* [6].

In this paper we present a KE diagnosis support tool for the detection of AD where ontologies and semantic reasoning play a fundamental role.

Our work is intended to aid physicians in the early stages and possible early detection of AD by using (i) multidisciplinary knowledge gathered, and (ii) inference and reasoning over the underlying Knowledge Bases. We differ from previous approaches as we aim to handle knowledge and not only data or information. To do so, we use ontologies as our supporting knowledge structures and a semantic reasoning system for the decision making process.

Our support system is based on production rules, which are provided by domain experts. Through a validation and reasoning process, our approach ultimately generates sets of suggestions that support the diagnosis.

This research was partially funded by the Centre for the Development of Industrial Technology (CDTI) of the Ministry of Industry, Tourism and Trade of Spain under the grant CENIT-20081013, which is part of the INGENIO2010 program.

This paper is arranged as follows: Section II presents an overview of the relevant state of the art; Section III introduces our approach as well as the reasoning system which supports the decision; Section IV presents a test example of our diagnosis system, showing some real world gathered data, and lastly, future work and conclusions are presented in Section V.

II. PREVIOUS WORK

AD is a neurodegenerative disease that presents an eminently multidisciplinary approach for diagnosis. This fact imposes some special requirements to possible computerized diagnostic support systems and the techniques to be applied. In this chapter, some previous work relevant to the scope of this paper is presented.

A. Clinical Decision Support Systems

Clinical Decision Support Systems (CDSS) are active knowledge resources that use patient clinical data to generate case specific advice [7]. CDSSs are aids to decision making in clinical processes and can be used with different purposes such as diagnosis or treatment. They are generally designed to integrate a medical Knowledge Base, patient data and an inference engine to generate case-specific advice.

CDSSs can support the process of the diagnosis of AD as shown by Lindgren [8]. Therefore, the medical community has a great interest on the design and development of CDSS and there are international organizations working on the standardization of CDSSs. For example, the Clinical Decision Support Work Group of Health Level Seven (HL7) is working on the Virtual Medical Record (vMR) [9] that is used for the representation of clinical information exchanged across multiple health information systems. Related to the vMR, a work worth mentioning is GELLO [10], which aims to provide a common format for data encoding and manipulation for any clinical application.

B. Semantics in the medical domain

Semantic technologies are used among others, (i) to support the integration of heterogeneous knowledge, (ii) the expression of rich and well-defined models for knowledge aggregation, and (iii) the application of logic for the generation of new knowledge [11]. In the medical domain, semantically structured clinical data are used in distributed environments, as shown in [6]. In this work, different approaches using semantic technologies are presented for several research directions in the medical domain.

In particular, ontologies are very promising from the point of view of the medical domain. For the computer science domain, Gruber defined an ontology as the explicit specification of a conceptualization [12]. It can fulfill 2 important issues: the need for organized and standardized terminologies and the need for reusable structures [13]. Ontologies have been applied to several problems in health care domain, such as the management of interoperability issues [14], the Patient-Centered Healthcare [15] and more recently, the mapping of terminologies [16].

Ontologies to be reused or aligned with third party systems can be found in repositories. Among the most recent repository services, Bioportal provides access to the Knowledge Bases via web services as well as resources for the community-based evaluation and evolution of ontology content [17].

Ontologies deliver interesting benefits as their nature allows the reasoning and inferring of new knowledge [4]. Among the most widely used we can mention the case of SNOMED CT [16] and SWAN [18]. The first one is a common standardized terminology for the medical domain and the second represents an effort to provide an integrated scientific knowledge for researchers to share their results. Being the most representative ontologies in this domain, interesting problems were found during our research (completeness of the Knowledge Bases, usability for the case of AD, etc). In section III we will discuss in further detail the reasons that made us develop an entirely new ontology (the MIND ontology) and align it to both SWAN and SNOMED CT.

1) SWAN - Semantic Web Application in Neuromedicine

SWAN is the result of a project intended for developing an integrated scientific knowledge infrastructure applied to AD, using Semantic Web technologies. It is published as part of the Alzheimer Research Forum website (www.alzforum.org) and it is a framework for integrating the scientific advances made within different projects and locations in the domain of AD.

The integration with SWAN endorses contents with hypotheses and publications extracted from the Alzheimer Research Forum, as shown by Lam *et al.* [18]. SWAN, however, is not designed for the diagnosis of AD and hence it does not consider the terms needed. In our approach we use this ontology to support the production rules as it will be explained in section III.

2) SNOMED CT - Systematized Nomenclature of Medicine Clinical Terms

SNOMED CT is a comprehensive clinical terminology that provides clinical content and expressivity for clinical documentation and reporting. SNOMED CT provides the core general clinical terminology for the Electronic Health Record (EHR). It describes different clinical concepts such as diseases and procedures, including those needed for the diagnosis of AD.

Thus, we believe that the need of a congruent ontology like the one we present in this paper to be aligned with SNOMED CT when possible provides an extra standardization to our work. When integrating a newly developed ontology (local ontology) with a standard, the reutilization of the local ontology by different organizations is possible [19]. Houshiaryan *et al.* dealt with this issue in their work [13]. SNOMED CT is very suitable for this integration, as the mapping schemas to other clinical terminologies have already been carried out, e.g. the mapping to ICD-10 [16].

III. CLINICAL DECISION SUPPORT SYSTEM FOR THE EARLY DIAGNOSIS OF AD

In this section, we present our CDSS for the early diagnosis of AD. The proposed solution consists of a support Knowledge Base and a reasoning system. Our approach consists of three different ontologies: SWAN, SNOMED CT and the MIND ontology.

As presented in Section II, SWAN links and endorses the criteria for the diagnosis of AD with the hypotheses and publications that are being held by the medical and scientific community. Hence, the contents of our system, such as tests carried out and rules applied for the diagnosis, can be validated and verified to be current and updated. Moreover at any given point, the physician will be able to check the scientific work that provided the production rule to be rationalized by the system.

The second ontology in our approach is SNOMED CT. This ontology is used for standardization purposes. In our case, the domain experts have concluded that SNOMED CT does not contain the terms needed for the complete diagnosis process and since the addition of new terminologies is arguably a slow process, we decided to develop a new ontology for the tests that serve as basis for the diagnostic and align it with SNOMED CT. This approach gives us both standardization and enough flexibility to develop a more specific domain based knowledge container.

The third and last ontology in our schema is called the MIND ontology and it describes the tests carried out to patients. It will be explained in detail in the next sections.

Fig. 1 depicts the three ontologies and their usability and context of application of the presented ontologies.

A. *The MIND ontology*

The MIND ontology, describes the neuropsychological, neurological, radiological, metabolomical and genetic tests carried out to patients. In order to develop this ontology, we consulted domain experts. They described the battery of tests that are usually performed on a patient under the suspicion of AD.

As we found that the tests and their descriptive parameters are in constant change, we decided on basing the whole ontology creation in a simpler paradigm and the result was a set of Graphical User Interface (GUI) which provided the contestant and structure of the MIND ontology.

The aforesaid approach gave us even more flexibility during the development of the ontology as well as in the future, as a new test (possibly one that has not exists yet) can easily be added to the ontology. The initial set of GUIs used was based on a system used by physicians to store the results of the tests.

Ontology	Description	Focus
SNOMED CT	Patient	Clinical
SWAN	Domain	Symptomatic
MIND Ontology	Tests	Diagnosis

Figure 1. Type of description and focus of each of the ontologies.

As the system is accessible via the web, there is no need for an ontology engineer to check or re-build the ontology. In the same way, there is also no need for any prior knowledge about the OWL standard specification.

Our system collects the tags in the tests GUIS and divides this information in two: (i) properties and classes for the ontology creation (domain ontology structural information) and (ii) values stored which become instances on the ontology and at the same time are stored/gathered to/from the clinical databases, allowing legacy systems stored knowledge integration (Fig. 2 depicts a screenshot of a GUI).

This part of our system is automatic and the gathered information is constructed following the Knowledge Base logical model that will be explained next.

B. *Main classes of the MIND ontology*

There are seven classes that are automatically mapped during creation time: **Doctor**, **Patient**, **Diagnosis**, **Enrollment**, **FollowUp**, **Test** and **TestValue**.

The **Test** class is the superclass of the different tests applied, and in general a new test will be considered as a sub class of this class.

The **Test** class is related with the class **Patient** with the *correspondingPatient* property, and with the class **Doctor** with the *orderingDoctor* property and at the same time with the class **FollowUp** with the *correspondingFollowUp* property.

Diagnosis and **Enrollment** classes are related to the class **Patient** with *hasDiagnosis* and *hasEnrollment* properties, respectively.

The instances of the **TestValue** class are the data gathered in the GUIs. Hence, **Test** and **TestValue** are related to the properties that refer to those parameters in the aforesaid GUIs. In other words, these are results of the different tests carried out. Fig. 3 depicts the described ontology.

C. *Inference system and rules engine*

With solely ontology and the mappings to SNOMED CT and SWAN, the physician is able to make queries to the

Figure 2. Screenshot of a source GUI (partial view)

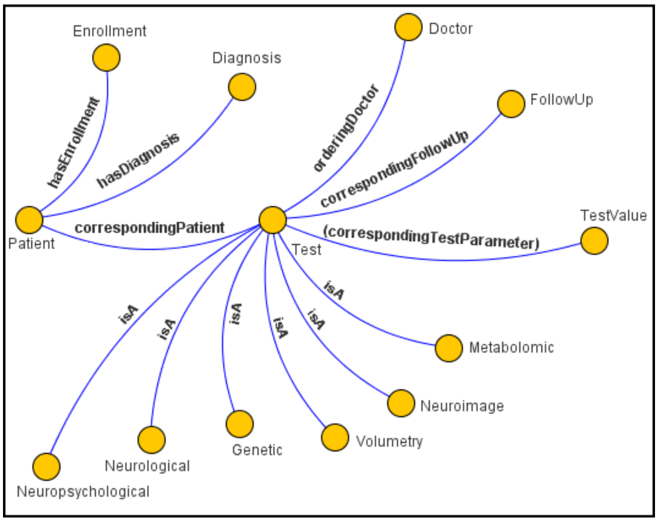

Figure 3. Overview of the MIND ontology

knowledge structure at any given time. At this point arguably no difference is distinguished when compared to a very good relational information storage mechanism or even some modern (state of the art) clinical information managing systems. However the intrinsic semantics that is embedded in our Knowledge Base shows its full potential when interrogated and inferred using production rules and DL reasoners.

In order to provide such mechanisms, we have developed a series of methods programmed using the Protégé OWL API.

Domain experts have produced a set of rules, which are used for the reasoning process.

In an attempt to provide a better result, we asked the domain experts to rank each rule (producing a sort of weighted rule). During reasoning time an initial importance hierarchy is provided. Last but not least, every rule is endorsed by the corresponding publication or bibliographic source when possible (via a link given by the mapping of the MIND ontology and the SWAN ontology). Fig. 4 presents one of our production rules which follow a classical if/then/else structure. Our approach for the rules syntax is inspired in RULE ML recommendation with minor changes given basically for usability reasons.

Lastly, our system presents the result of the query process and the initial weights used only for ranking purposes. At this time this process is at an experimental phase and not standardization on the rank is advised to the domain experts as

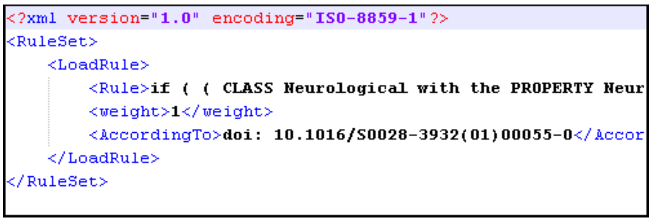

Figure 4. Production rule example

guideline. Fig. 5, depicts an example output from the console debug. In section IV, a test example will be presented.

IV. TEST EXAMPLE

Our approach was developed under the framework of the Spanish project MIND (www.portalmind.es), which aims to follow the multidisciplinary approach of AD. The MIND project carries out a clinical trial over 350 patients in 3 hospitals of Valencia, in Spain.

Patients taking part in the study fulfill the inclusion requirements set by our experts and can belong to one of the following 3 groups: Alzheimer, Mild Cognitive Impairment (MCI) or control. One of the most important objectives of the MIND project is to early detect which MCI patients are going to evolve to AD.

Patients are followed up every 6 months during 3 years and each time the set of tests is applied in its entirety to them. To be precise, the tests carried out are divided into the following:

- Neuropsychological Tests
- Neurological Tests
- Blood analysis (for the genetic and metabolomical results)
- Magnetic Resonance Imaging (MRI)
- Functional Magnetic Resonance Imaging (fMRI)

Parallel to the clinical trial, the clinical decision support system presented in the previous sections has been developed. It aids physicians during the whole diagnosis process. First of all, the above mentioned tests are carried out to patients and the data and results generated are gathered in a web based system called ODEI. This information is editable and can be reviewed at any time in the GUIs. Based on the knowledge in those GUIs, an OWL-DL ontology is created automatically, the MIND ontology. Hence, the MIND ontology contains the knowledge from the tests. The data corresponding to the results are stored in a data base which offers web-service access to it. The ontology instantiates the query-calls to the data. Fig. 6 and Fig. 7 show the data gathering process and the ontology generation for the corresponding part. Fig. 8 depicts an example of a query-call instance.

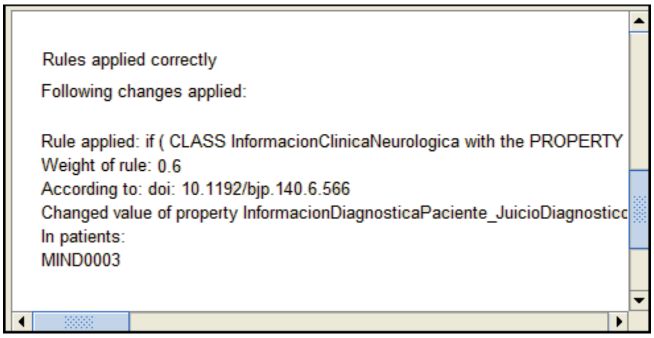

Figure 5. Output from the system (partial)

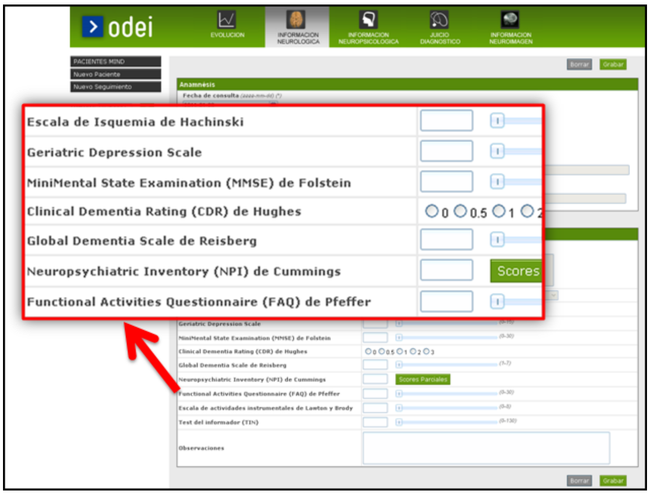

Figure 6. Data gathering process

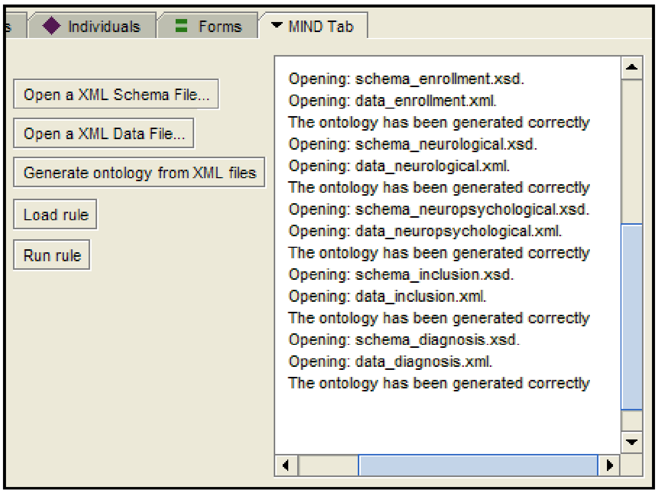

Figure 9. Appearance of the Protégé tab plug-in

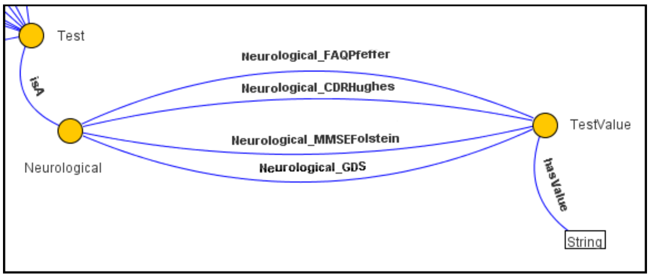

Figure 7. Ontology generation for the part corresponding to the data gathering in Fig. 4.

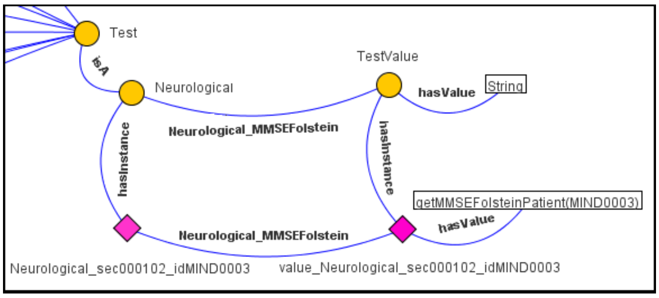

Figure 8. Instantiation to the query-call

This ontology creation is made programmatically and it is transparent for both, the final user and the ontology expert. For testing purposes a Protégé tab plug-in has been implemented, to show manually the ontology creation step by step. Every GUI generates a couple xml documents containing the structure (schema) on the one side and the data on the other. These xml documents are loaded manually in the tab plug-in and the corresponding ontology is created. Fig. 9 shows the appearance of this Protégé tab plug-in.

Due to intellectual property issues, we are not allowed to show the host application. Therefore, only the console output is shown in this paper, which has been also integrated in the aforementioned Protégé tab plug-in.

The xml documents from different GUIs (enrollment, neurological, neuropsychological, inclusion and diagnosis) are loaded and the ontology is generated correctly. Fig. 9 above shows the ontology loading and generation process.

Once the Knowledge Base has been created, a set of rules given by domain experts is loaded. As mentioned in the previous section, each rule is weighted and also endorsed by SWAN. Fig. 10 depicts the loading rules to the reasoner process.

The reasoner applies the rules loaded to the Knowledge Bases and infers the corresponding diagnoses.

For a given patient different rules may apply, each one with a corresponding diagnosis. The weights of the matching rules rank the diagnoses for the final suggestion. Fig. 11 shows the inference of the diagnosis. A matching rule is depicted.

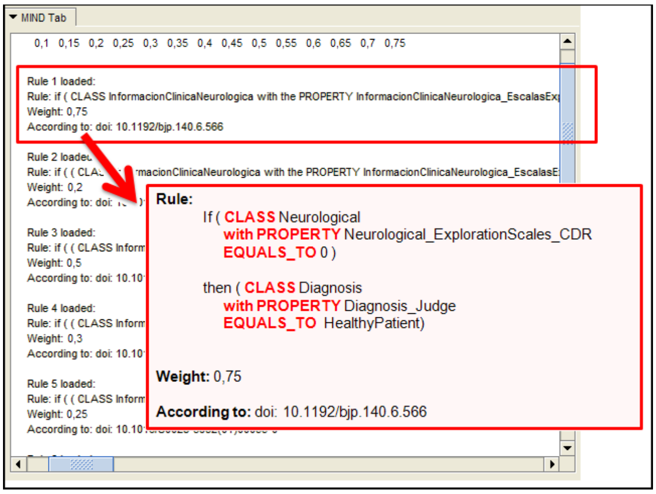

Figure 10. Loading rules to the reasoner

V. CONCLUSIONS AND FUTURE WORK

In this paper we have presented a Knowledge Engineering diagnosis-support tool for the detection of AD where ontologies and semantic reasoning play a fundamental role. Our work is intended to aid physicians in the early detection of AD by using multidisciplinary knowledge gathered and inference and reasoning over the underlying Knowledge Bases. A test example of our tool has also be shown and discussed.

Three aligned ontologies - the MIND ontology, SWAN and SNOMED CT – form the Knowledge Base. Diagnoses are inferred by a rule-based reasoner applied over the aforementioned Knowledge Base. Those rules are provided by domain experts. They are endorsed with the corresponding publications (via mappings to the SWAN ontology) as well as weighted (ranked according to an importance hierarchy given by the domain experts).

It would be important to extend the standardized ontologies to represent all the domain knowledge of Alzheimer's disease to improve the accuracy of reasoning.

The Clinical Decision Support System presented in this paper is currently being validated by 3 hospitals. Results of such validation will be presented in a future work.

Through the work in this project we have encountered 3 areas for future work. Firstly, during the mapping process a need for a future graphical ontology mapping tool has arisen, as well as a graphical rule editor for production rules.

Additionally, with regard to the reasoner, we would like to explore the Set of Experience Knowledge Structure (SOEKS) [20] technology and experience-based reasoning.

Lastly, the technologies and the system presented in this paper could also possibly be applied to other domains such as cardiologic diseases or autism, as well as extended to other purposes such as the treatment and monitoring of patients.

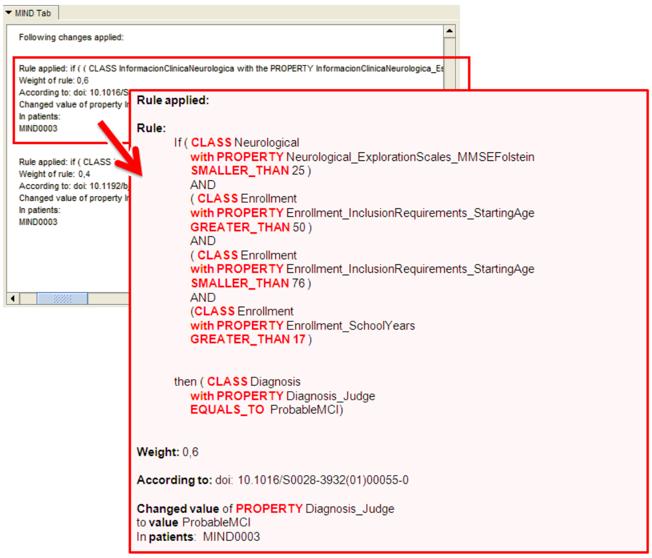

Figure 11. Inference of the diagnosis

ACKNOWLEDGMENT

We would like to thank the members of the MIND consortium and in specific we would like to express our gratitude to ERESA for leading this effort.

REFERENCES

[1] B. Monien, L. Apostolova, and G. Bitan, “Early diagnostics and therapeutics for Alzheimer’s disease – how early can we get there?” Expert review of Neurotherapeutics, vol. 6, 2009, pp. 1293-1306.

[2] M. Pleckaityte, “Alzheimer’s disease: a molecular mechanism, new hypotheses, and therapeutic strategies,” Medicina, vol. 46(1), 2010, pp. 70-6.

[3] E.-J. Ko, H.-J. Lee, and J.-W. Lee, “Ontology and CDSS based Intelligent Health Data Management in HealthCare Server,” World Academy of Science, Engineering and Technology, vol. 29, 2007, pp. 119-123.

[4] W.D. Yu and S.R. Jonnalagadda, “Semantic Web and Mining in Healthcare,” e-Health Networking, Applications and Services, 2006, pp. 198-201.

[5] A. Segev, M. Leshno, and M. Zviran, “Internet as a Knowledge Base for medical diagnostic assistance,” Expert Systems with Applications, vol. 33(1), July 2007, pp. 251 - 255.

[6] S. Gnanambal and M. Thangaraj, “Research Directions in Semantic Web on Healthcare,” Journal of Computer Science, vol. 1, 2010, pp. 449-453.

[7] L. Joseph, J. C. Wyatt, and D.G. Altman, “Decision tools in health care: focus on the problem, not the solution”, BMC Medical Informatics and Decision Making, 2006, vol. 6:4.

[8] H. Lindgren, “Decision support system supporting clinical reasoning process - an evaluation study in dementia care”, Studies in Health Technology and Informatics, 2008, vol. 136, pp. 315-20.

[9] P.D. Johnson, S.W. Tu, M.A. Musen, and I. Purves, “A Virtual Medical Record for Guideline-Based Decision Support" Medinfo, 2001, pp. 294-298.

[10] Mor Peleg and Daniel L. Rubin. "Querying Radiology Appropriateness Criteria from a virtual Medical Record using GELLO". Proc. Workshop on Knowledge Representation for Health-Care: Patient Data, Processes and Guidelines, in conjunction with Artificial Intelligence in Medicine Europe, Verona Italy, July 19, 2009

[11] M. Dumontier, B. Andersson, C. Batchelor, C. Denney, C. Domarew, A. Jentzsch, J. Luciano, E. Pichler, E. Prud’hommeaux, P. Whetzel, O. Bodenreider, T. Clark, L. Harland, V. Kashyap, P. Kos, J. Kozlovsky, J. McGurk, C. Ogbuji, M. Samwald, L. Schriml, P. J. Tonellato, J. Zhao and S. Stephens. “The Translational Medicine Ontology: Driving personalized medicine by bridging the gap from bedside to bench” Bio-Ontologies 2010, Proceedings of the 13th Annual Bio-Ontologies Meeting, Boston, USA

[12] T.R. Gruber, "Toward principles for the design of ontologies used for knowledge sharing", International Journal of Human-Computer Studies, 1995, vol. 43(5–6), pp. 907–928.

[13] Kambiz Houshiaryan, Hwa-Sun Kim, Hyung-Hoi Kim, Tran Tung, Ii-Kon Kim, Yun-Sik Kwak, and Hune Cho, “Customized Ontology-Based Intelligent Medical Terminology Tool,” Proceedings of 7th International Workshop on Enterprise networking and Computing in Healthcare Industry, 2005, pp. 320-324.

[14] P. Ganguly, S. Chattopadhyay, and P. N, Ray and Pradeep, “An Ontology-based Framework for Managing Semantic Interoperability Issues in e-Health,” 2008, pp. 73-78.

[15] J. Puustjiirvi and L. Puustjiirvi, “Providing Relevant Health Information to Patient-Centered Healthcare,” 12th IEEE International Conference on e-Health Networking, Applications and Services, 2010.

[16] M. Nyström, A. Vikström, G.H. Nilsson, H. Ahlfeldt, and H. Orman, "Enriching a primary health care version of ICD-10 using SNOMED CT mapping", Journal of biomedical semantics, 2010, vol. 1(7).

[17] N.F. Noy, N.H. Shah, P.L. Whetzel, B. Dai, M. Dorf, N. Griffith, C. Jonquet, D.L. Rubin, M.-anne Storey, C.G. Chute, and M.A. Musen,

"BioPortal : ontologies and integrated data resources at the click of a mouse," Nucleic Acids Research, vol. 1, 2009, pp. 1-4.

[18] H.Y.K. Lam, L. Marenco, T. Clark, Y. Gao, J. Kinoshita, G. Shepherd, P. Miller, E. Wu, G. Wong, N. Liu, C. Crasto, T. Morse, S. Stephens, and K.H. Cheung, "Semantic Web Meets e-Neuroscience: An RDF Use Case", In Proceedings of the ASWC International Workshop on Semantic e-Science, 2006, pp. 158-170

[19] C. Toro, M. Graña, J. Posada, J. Vaquero, and Edward Szczerbicki, "Domain Modeling Based on Engineering Standards", Knowledge-Based and Intelligent Information and Engineering Systems, 2009, vol. 5711, pp. 95-102

[20] C. Sanín and E. Szczerbicki, "Experience-based knowledge representation: SOEKS", Cybernetics and Systems, 2009, vol. 40: 2, pp. 99 — 122