Ontologies for Early Detection of the Alzheimer Disease and Other Neurodegenerative Diseases

Alba Gomez-Valadés, Rafael Martínez-Tomás, and Mariano Rincón-Zamorano

Dpto. Inteligencia Artificial, Universidad Nacional de Educación a Distancia UNED, Madrid, Spain
albagvb@dia.uned.es

Abstract. Nowadays technologies allow an exponential generation of biomedical data, which must be indexed according to some standard criteria to be useful to the scientific and medical community, being neurology one of the areas in which the standardization is more necessary. Ontologies have been highlighted as one of the best options, with their capability of homogenise information, allowing their integration with other kind of information, and the inference of new information based on the data that is stored. We analyse and compare the approaches taken by different research groups inside the area of the Alzheimer's disease, and the ontologies they developed with the objective of providing a common framework to standardize information, data recovery or as a part of an expert system. However, to make this approach work the ontologies must be maintained over the time, a critical point which is not been followed by any of the ontologies reviewed.

1 Introduction

During the last decade has been a revolution in the volume and complexity of data created in the life sciences, and with them, in the possibility of studying such data [12] . However, their utility depends fundamentally on the ability to know how to handle and interpret large heterogeneous datasets [8] scattered [4] across different databases and under different formats, so the integration and standardization of the data it's necessary in order to make information useful [6] while allow the data interoperability, to facilitate the extraction and retrieval of information [3], with diverse scientific and clinical objectives. Thus, the generation of adequate infrastructures to allow standardization, exchange and sharing information have become a key objective for the success of the current and future research and clinic [4].

In this framework, the development of ontologies has been established as one of the most adequate solutions to confront these problems, as in the domains of biomedical research, where specific ontologies have been developed across the different fields [3]. Within biomedicine, neurology is one of the fields where integration is more necessary to make the data useful, since mental processes extend across very heterogeneous levels [11]. Mild cognitive impairment (MCI) has attracted significant attention in neurological research, as it is a transition phase between normal aging and dementia, especially that related to Alzheimer's disease (AD), making MCI one of the main indicators of developed AD, that's why the development of methods that allows their early detection of AD is essential to improve the quality of life of patients [22].

For this purpose, various types of tests have been developed, which evaluate in different ways the state of the patients: biological test (such as the concentration measurement of beta-amyloid); imaging test (as Nuclear Magnetic Resonance or MRI), or neuropsychological tests. The underlying theory of those last tests is that these neurodegenerative diseases cause damage in certain areas of the brain, which affect different mental functions and cognitive processes. In the neuropsychological tests, these dysfunctions created by the MCI are reflected as different types of errors or deficiencies committed at the time of performing of each test. Those tests are the only ones capable of measuring the cognitive abilities in patients, such as short-term memory versus long-term memory, or the executive functioning [11]. In addition, neuropsychological tests have the advantage of being non-invasive, versatile and low cost.

The purpose of this article is to review the different approaches and objectives carried out by different research groups in this area, not only to solve the problems of interoperability and standardization in the AD domain, and other neurodegenerative diseases related to the MCI, but also to allow storage, recovery and making inferences from the information, as well as to help the MCI diagnosis.

2 Ontologies and Neurodegenerative Diseases

An ontology is a formal definition of classes, properties, and relationships between them, which is integrated inside some knowledge area, allowing the homogenization and consensus in the representation and reutilization of a domain [19]. It facilitates the exchange of information in biomedicine [17], the integration and recovery of heterogeneous data from a diverse variety of sources, with the aim to improve the diagnosis or the treatment of the disease. Depending on the approach and the final goal sought during the development of the ontology, it can be distinguished 3 main groups [21]:

- Ontologies for the standardization in the terminology: They seek to allow both the direct reutilization of the terminology by third parties, and the compatibility of the ontology with most of reasoning engines. In design they are characterized by a strong hierarchy of classes and a large amount of different types of metadata for a correct definition of the terms.
- Ontologies for the storage and recovery of the information: They seek to allow the recovery and inference of new knowledge from the stored data, as well as achieve a standardization of the domain. In design, they tend to display less

nesting in the class hierarchy than in the previous case, and the metadata usually are scarce or reduced to the minimum, although their axiomatic tends to be more specific.
- Ontologies as a diagnosis support: they are normally designed from scratch for a practical purpose, to be integrated into a larger system as a module or subsystem, normally working as a knowledge database. They usually are linked to the target application, being difficult their reutilization in other contexts. In design, the class hierarchy usually has only relevant term to the system, few levels of nesting and scarce metadata, although their axiomatic are heavy.

However, there are also some problems associated to the usage of the ontologies. The main and more immediate problem are the low reutilization of the existing terminology, generating redundancy problems and conflict in the term names, unstable references, redundancy in the class hierarchy, and inconsistencies between them [15]. To avoid this situation, a priority during the development of new ontologies must be the reutilization of already existent ontologies as much as possible, only adding new classes and instances when they are not covered by any of the consulted ontologies [10].

In the following sections, the main approaches are presented. They are organized according to the final objective of the ontologies described above.

2.1 Ontologies for the Standardization in the Terminology

The project of Gao et al. [9] of the Semantic Web Application in neuromedicine (SWAN) was one of the first biomedicine ontologies focus in the storage and contextualization of the existent information about the AD. The project was developed as an infrastructure that integrate in an effective way the scientific knowledge of the AD allowing the construction of a semantic web of hypothesis, publications and digital repositories [5]. SWAN was considered as the reference repository about the data regarding to the AD that were available in the web. However, this ontology and the associated application has been discontinued from all the online repositories where it had been stored, being no possible to retrieve it.

On the other hand, Jensen et al. [14] propose Neurological Disease (ND) Ontology, which seeks to provide a framework for the representation the key aspects of neurodegenerative diseases for study and treatment, providing a set of controlled classes connected in a logical way to describe the range of the neurodegenerative diseases, as well as their signs and symptoms, evaluations, diagnosis and interventions that have been found in the course of clinical practice. It can also serve to link and extend other ontologies of the same domain. In the Fig. 1, it can be observed the extension in those domain areas that ND performs to the Basic Formal Ontology (BFO) [1]base ontology.

Later, in the paper published by Cox et al. [7] is described NeuroPsychological Testing Ontology (NPT), which seeks to extend and to complete ND, specifically in the part of the domain relating to the neuropsychological tests. NPT provides a set of classes for the realistic representation and the annotation of a wide variety of neuropsychological tests which evaluate similar or overlapping domains of the cognitive function, such as the MMSE, as well as other associated data, allowing the integration of the results. This provides a realistic and detailed representation of the functioning of the cognitive process and the functions they involve. However, NPT also has the problem of excess of complexity, which make it little manageable due to that. This makes difficult to locate and focus in the relevant classes, since it has classes from very heterogeneous domains that have little to do with the domain in which this ontology is focused.

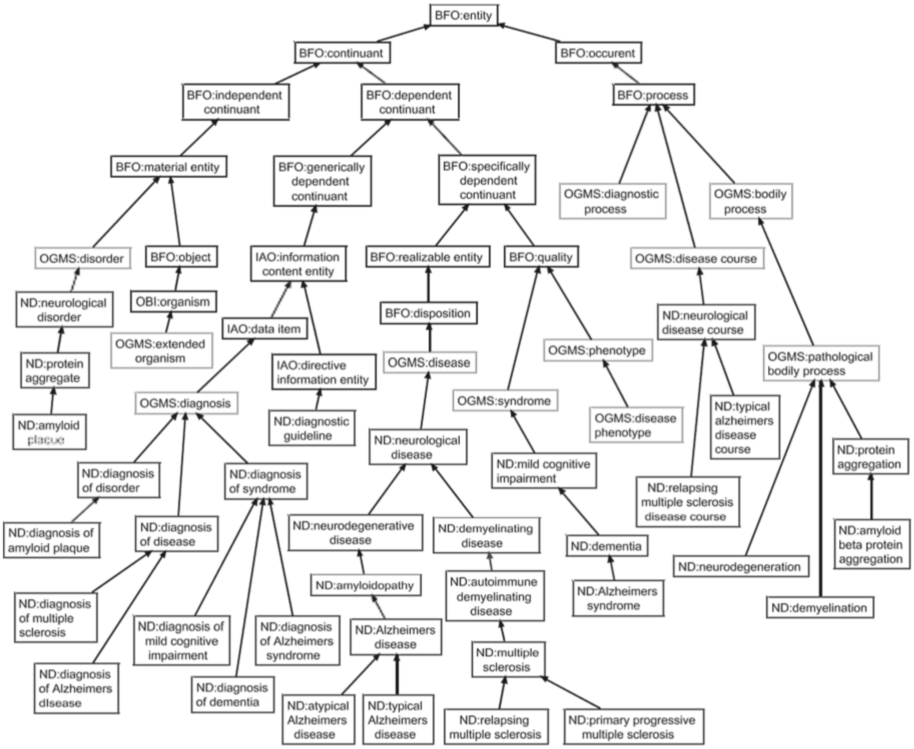

Fig. 1. Graphic depiction of the high-level terms in ND. A subset of classes of ND that shows the is_a relations between BFP [14]

2.2 Ontologies for the Storage and Recovery of the Information

According to the authors of SWAN, this ontology, was designed both as a standardization system and as a storage and information retrieval system [18] stored in the Alzforum digital library, in order to make information inference and generation, among other reasons. The information retrieval would be done through two methods: SWAN Workbench, available only for a reduce group of users, and the SWAN Browser, designed to be used more generally [5].

Malhotra et al. [16] propose Alzheimer Disease Ontology (ADO), following the guide of the BFO. This ontology seeks the widest possible coverage of the different aspects of the AD domain in a structured way, from diagnosis to treatments. Like SWAN, ADO was designed with the idea of allowing retrieval and inference of the information, that will be done through queries about the stored data. However, ADO has the problem of having a low reutilization of already existed ontologies, as well as an obtuse axiomatic of interpreting. For example, the classes generically_dependent_continuant and specifically_dependent_continuant are disjointed classes, but they are also marked as equivalent to dependent_continuant. Since they are subclases of dependent_continuant, this makes those classes both equivalent and disjointed between them. This kind of situations appears more than once across the ontology.

In other hand, the OntoNeuroLOG ontology from Batrancourt et al. [2] is a multilayer ontology of the instruments used to evaluate the brain and cognitive functions. OntoNeuroLOG is a multilayer ontology organized in sub-ontologies or modules located in three different levels of abstraction. It has been built using DOLCE as its main basis, which has been complemented in the different modules with other ontologies. Also, new terms were defined specifically for the ontology when necessary. Although it has the benefit of being one of the most complete ones, it also has the problem of being one of the biggest ontologies, making it impossible to integrate the 3 modules in which it's splitted up as a whole in Prot´eg´e.

2.3 Ontologies as a Diagnosis Support

Sanchez et al. [18] propose the MIND ontology as a decision support system that aid physicians in the early diagnosis of AD. This project merges an ontology and a semantic reasoner able to infer logic consequences starting from a series of facts or given axioms to help physicians in the early detection of AD using the multidisciplinary knowledge stored in the ontology. For that case the ontology describes the neuropsychological, neurological, radiological, metabolomical and genetic tests carried out to patients. Despite everything, this ontology does not relate with the different cognitive process, or the mistakes committed in the tests during their performance; instead it employs only in the final score.

The Ontology driven decision support for the diagnosis of MCI was proposed by Zhang et al. [22] as a method of decision support supported by ontologies designed to avoid subjectivity in the diagnosis of MCI through magnetic resonance imaging, for the detection of the cortical cortex thickness, since it is reduced in patients with MCI [20]. However, it has the disadvantage that the C4.5 decision tree used to developed the model is sensitive to the training data, with the consequent problem of overfitting. Also, this ontology is focused only on an imaging approach of MRI, modelling the ontology according to this criterion, and ignoring other systems such as neuropsychological tests.

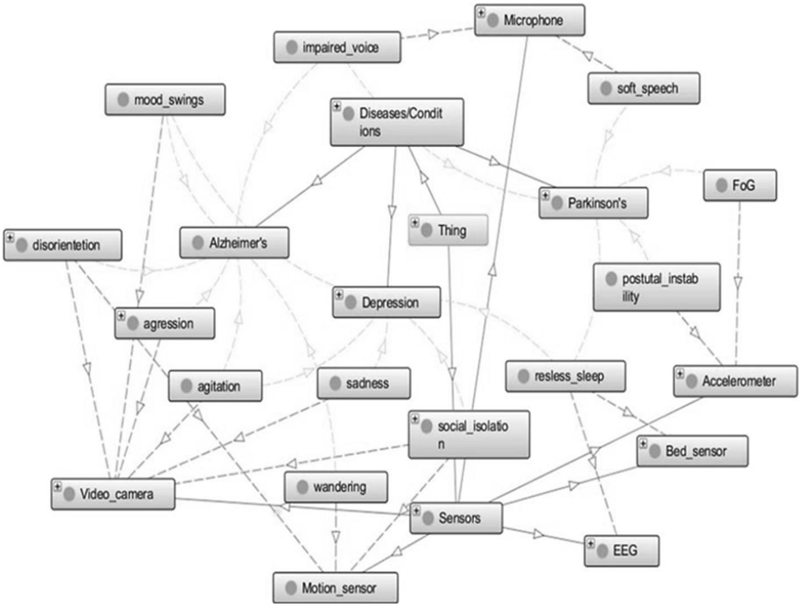

Fig. 2. DSS (Disease-Symptom-Sensor) ontology model [13].

Ivascu et al. [13] propose a multiagent ontology, designed to facilitate remote monitoring of patients who are susceptible to develop cognitive impairment disease. To achieve that goal, this ontology is built using a combination between an ontology developed as a Disease-Symptom-Sensor (DSS) system, which serves as a diagnosis component, which is shown in Fig. 2, and a multiagent system, to which the ontology complements, and which is a group of programs specialized in a task and capable of working together. This way, the system can predict the disease based on the registered symptoms. However, there are still some elements that must be solved before this system can be release, such as data privacy improvements, automated devices/sensors discovery, context sensitive information aggregation and activities correlations for users that match the same profile.

Finally, [21] propose an ontology based in the fuzzy logic, AlzFuzzyOnto, which are developed using MIND as the basis to model an expert system to aid to physicians in the early diagnosis of AD. The motivation is that there is a significant number of terms and concepts that constitute a source of imprecision and uncertainty. To solve this, fuzzy classes were added, and the concepts were connected using a fuzzy relationship of belonging, in which a crisp concept are related to a fuzzy one. Those relations have different weights of degrees of membership with each class with a value in the interval [0, 1], allowing the developed of a fuzzy ontology for the AD. However, the experimentation, validation and instantiation of this ontology has not carried out, but are left as a future work, as well as the construction of a fuzzy inference engine.

Table 1. Summary table of the ontologies. The Reutilization column is according to the paper if the ontology was not found. Axiomatics: High (high number of relationships between low or concrete clases) Medium (relationships between high or more abstract clases, which are inherited by the concrete ones) Low (scarce number of relationships in general). Reutilization: High (they use of different kind of ontologies to model the ontology domain, usually high level) Medium (they reuse no more of 1 or 2 ontologies) Low (they only use the squeme of one ontology for the more upper classes, or no reutilization at all). Hierarchy: High (great depth in nesting, usually a high number of classes -more than 1000-) Medium (presence of more classes in the root, great depth in nesting are more uncommon) Low (the nesting usually no overpass a 5 classes depth, and the number of classes are usually less than 100). Metadata: High (label, description, comments and other metadata tend to appear in most classes), Medium (label and a short description, they can also count with synonyms, comments, etc, rarely surpassing one line), Low (usually just the label; occacionally short descriptions, comments, synonyms, etc, usually of only a few words long)

Name	Publicly available	Last update	Focus	Axiomatics/ relations	Reutilization	Class hierarchy	Metadata
SWAN	No	2009	Terms standardiza-tion/Knowledge inference	–	–	–	–
ND	Yes	2012	Terms standardization	Medium	High	High	High
NPT	Yes	2013	Terms standardization	Medium	High	High	High
ADO	Yes	2013	Knowledge inference	Low	Low	Medium	Medium
OntoNeuroLOG	Yes	2013	Knowledge inference	High	High	High	Low
MIND	No	2011 (paper)	Part of an expert system	–	Low	–	–
Ontology Driven Decision	No	2013 (paper)	Part of an expert system	–	No	–	–
Multiagent	No	2015 (paper)	Part of an expert system	High	No	Low	Low
AlzFuzzyOnto	No	2015 (paper)	Part of an expert system	–	Low	–	–

3 Conclusions

The new technologies are allowing an exponential generation of biomedical data, which must be stored and indexed correctly to allow easy access and management, so they can be useful to the scientific and medical community. Inside the area of biomedicine, neurology, and more specifically, MCI related to the early detection of neurodegenerative diseases, has become one of the fields where such integration has become more necessary. Ontologies are a powerful tool in achieving the goal of integration and easy access to the data, since they can provide a standardized vocabulary for the representation, sharing and reuse of the knowledge, as well as storage, retrieval and inference of information. Moreover, they can be integrated into expert systems, working as a knowledge database during the early detection of MCI.

Throughout this article the main ontologies developed in this domain has been shown, either with the shared objective of standardizing and providing a common framework of existing information, data recovery or as a part of an expert system. The summary of the ontologies described is shown in Table 1. But for this to work, ontologies must be maintained, something that are not happening with any of the revised ones. Although this can be understandable in those ontologies designed to work as a part of an expert system to fulfil a specific function, in which case an operative version them would be enough, this situation would apply to standardization-oriented ontologies, though the absence of maintenance is common to all ontologies. Some of them only receive a single update, corresponding to the years of publication of the article. Those that received subsequent updates, they were made in the close years to the first publication, being ND and OntoNeuroLog the ones which received more support during more time, even though both have not received any updates in the recent years. Finally, there are several ontologies described in papers that were not possible to locate in any internet repository. Those are the ones oriented with a more practical goal, with the exception of SWAN, which was deleted from internet for unknown reasons.

Lastly, it has been found that term reuse between different ontologies is, on average, scarce or non-existent, the main exception being the ontologies of ND and the derivative ontology NPT.

Funding. We thanks to the Ministry of Education, Youth and Sports of the Community of Madrid, and the European Social Fund for a contract to A.G.-V. B. (PEJD-2017-PRE/TIC-4406) in the program of Youth Employment and the Youth Employment Initiative (YEI) 2018.

References

1. Arp, R., Smith, B.: Realizable entities in basic formal ontology. In: Proceedings of Bio-Ontologies Workshop, Intelligent Systemas for Molecular Biology (ISMB), p. 10 (2011)
2. Batrancourt, B., Dojat, M., Gibaud, B., Kassel, G.: A multilayer ontology of instruments for neurological behavioral and cognitive assessments. Neuroinformatics **13**(1), 93–110 (2015)
3. Blake, J.A., Bult, C.J.: Beyond the data deluge: data integration and bio-ontologies. J. Biomed. Inf. **39**(3), 314–320 (2006)
4. Burgun, A., Bodenreider, O.: Accessing and integrating data and knowledge for biomedical research. Yearb. Med. Inf. **17**(01), 91–101 (2008)
5. Ciccarese, P., et al.: The SWAN biomedical discourse ontology. J. Biomed. Inf. **41**(5), 739–751 (2008)
6. Costa, F.F.: Big data in biomedicine. Drug Disc. Today **19**(4), 433–440 (2014)
7. Cox, A.P., Jensen, M., Ruttenberg, A., Szigeti, K., Diehl, A.D.: Measuring cognitive functions: hurdles in the development of the neuropsychological testing ontology. In: Proceedings of the 4th International Conference on Biomedical Ontology 2013, Montreal, Canada, p. 6, July 2013

8. Decety, J., Cacioppo, J.: Frontiers in human neuroscience: the golden triangle and beyond. Perspect. Psychol. Sci. **5**(6), 767–771 (2010)
9. Gao, Y., et al.: SWAN: a distributed knowledge infrastructure for Alzheimer disease research. Web Seman.: Sci. Serv. Agents World Wide Web **4**(3), 222–228 (2006)
10. Gomez-Perez, A., Fernandez-Lopez, M., Corcho, O.: Ontological Engineering: with Examples from the Areas of Knowledge Management, E-Commerce and the Semantic Web. Advanced Information and Knowledge Processing. Springer, New York (2004). https://doi.org/10.1007/b97353
11. Hastings, J., et al.: Interdisciplinary perspectives on the development, integration, and application of cognitive ontologies. Front. Neuroinf. **8**, 62 (2014)
12. Hoehndorf, R., Schofield, P.N., Gkoutos, G.V.: The role of ontologies in biological and biomedical research: a functional perspective. Briefings Bioinf. **16**(6), 1069–1080 (2015)
13. Ivascu, T., Manate, B., Negru, V.: A multi-agent architecture for ontology-based diagnosis of mental disorders. In: 2015 17th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), pp. 423–430. IEEE. September 2015
14. Jensen, M., et al.: The neurological disease ontology. J. Biomed. Semant. **4**(1), 42 (2013)
15. Klein, M.: Combining and Relating Ontologies: an analysis of problems and solutions. In: Ontologies and Information Sharing, vol. 47, May 2001
16. Malhotra, A., Younesi, E., Gndel, M., Mller, B., Heneka, M.T., Hofmann-Apitius, M.: ADO: a disease ontology representing the domain knowledge specific to Alzheimer's disease. Alzheimer's Dement. **10**(2), 238–246 (2014)
17. Mead, C.N.: Data interchange standards in healthcare IT - computable semantic interoperability: now possible but still difficult, do we really need a better mousetrap? J. healthc. Inf. Manage. (JHIM) **20**, 71–78 (2006)
18. Sanchez, E., et al.: A knowledge-based clinical decision support system for the diagnosis of Alzheimer disease. In: 2011 IEEE 13th International Conference on e-Health Networking, Applications and Services, pp. 351–357. IEEE. June 2011
19. Trokanas, N., Cecelja, F.: Ontology evaluation for reuse in the domain of process systems engineering. Comput. Chem. Eng. **85**, 177–187 (2016)
20. Whitwell, J.L., et al.: MRI patterns of atrophy associated with progression to AD in amnestic mild cognitive impairment. Neurology **70**(7), 512–520 (2008)
21. Zekri, F., Bouaziz, R., Turki, E.: A fuzzy-based ontology for Alzheimer's disease decision support. In 2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1–6. IEEE. August 2015
22. Zhang, X., Bin, H., Ma, X., Moore, P., Chen, J.: Ontology driven decision support for the diagnosis of mild cognitive impairment. Comput. Methods Programs Biomed. **113**(3), 781–791 (2014)