24th International Conference of the European Federation for Medical Informatics
Quality of Life through Quality of Information – J. Mantas et al. (Eds.)
MIE2012 / CD / Short Communications (Oral)

Harnessing Linked Data for the Consumption and Publishing of Unified Diagnosis-Based Knowledge

Alejandro RODRÍGUEZ-GONZÁLEZ[1a], Miguel Ángel MAYER[b], Juan Miguel GOMEZ-BERBIS[a] and Ángel GARCÍA-CRESPO[a]
[a] *Computer Science Department, University Carlos III of Madrid, Spain*
[b] *Research Programme on Biomedical Informatics, IMIM-Universitat Pompeu Fabra, Barcelona, Spain*

Abstract. Linked Data has become one of the main technologies in knowledge representation, publishing and consumption. The breakthroughs of its use in the bioinformatics domain have been demonstrated in several fields such as pharmaceutical research and drug discovery among others. However, the generation of a particular dataset with medical information related to the diagnosis of diseases has not been addressed yet. In this paper, the authors propose an approach to solve this caveat based on the consumption of biomedical data from existing biomedical sources such as Bio2RDF, OBO-Foundry and SNOMED-CT among others to create a dataset with information about the main entities present in the differential diagnosis process and its relations.

Keywords. Linked data, medical informatics, semantic technologies, diagnosis

Introduction

The term Linked Data has gained momentum since its definition, which was coined by Sir Tim Berners-Lee in 2006. The concept has been introduced as part of a new breed of cutting-edge technologies for the Web, and more concretely, as a new part of Semantic Technologies. The concept has become very important in several areas such as Economics and Finances [1], IT Professionals [2], Enterprises [3], and scientific publications [4] among others.

In Bioinformatics we can also see the importance of Linked Data in this area [5]. We have some concrete examples of fields such as pharmaceutical research [6] or genes, proteins, drugs and clinical trials [7-8] among others. One of the most relevant efforts in the use of Semantic Technologies in "bio" area is Bio2RDF [7]. As it is described in [7] "*There are numerous bioinformatics databases available on different websites. Although Resource Description Framework (RDF) was proposed as a standard format for the web, these databases are still available in various formats. With the increasing popularity of semantic web technologies and the ever growing number of databases in bioinformatics, there is a pressing need to develop mashup systems to help the process of bioinformatics knowledge integration*." The use of such

[1] Corresponding Author. Alejandro Rodríguez-González, Computer Science Department, University Carlos III of Madrid, Spain; Email: alejandro.rodriguez@uc3m.es

systems to automatically extract and categorize biomedical information represents a great advance to provide access to biomedical information and establish relations between the concepts stored in these repositories.

From the diagnosis point of view, most of the medical literature provides information about the main entities involved in the diagnosis process of a concrete disease (also known as diagnostic criteria). However, these relations are not always prone to be isolated or even to find in order to create computational models which allow developing diagnosis systems which can assist physicians in some activities such as the diagnosis process, research and/or training, to mention a few.

The purpose of this paper is to present the adoption of Linked Data from a publishing and consumption point of view with the aim of developing medical knowledge about differential diagnosis field, allowing a future use of this knowledge.

1. Applying Linked Data in Medical Diagnosis

The representation of medical knowledge in diagnostic environments typically follows a concrete model, based on the definition of a disease as a diagnostic criterion [9-11]. This information is very valuable to physicians, researchers or students during their training. The use of biomedical knowledge sources such as Bio2RDF, OBO-Foundry and SNOMED-CT among others allows researchers to perform a consumption of the data stored in these repositories with the aim of building models like the ones described before. The solution proposed is based on the architecture depicted in Figure 1.

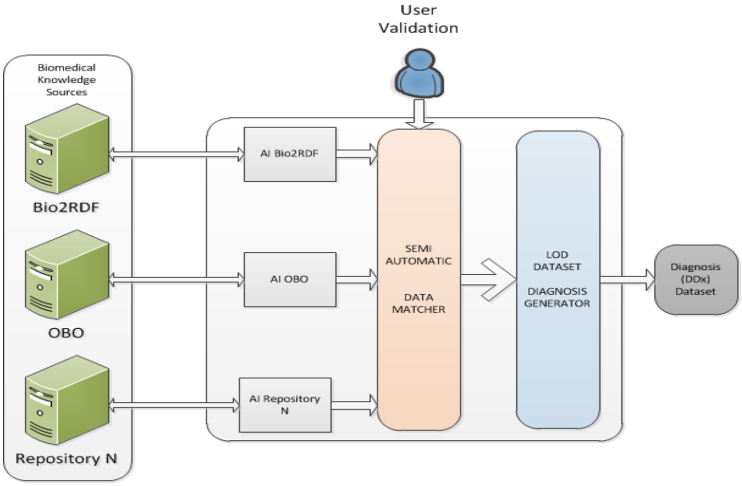

Figure 1. Architecture proposal.

The idea presented in this paper stems from, and builds on, using a software platform to consume medical data from several input platforms to get information about a disease. Once the information about the disease is obtained (from Bio2RDF for example we can talk about consumption from a Linked Data perspective), a model which represents the disease and the information obtained is generated. After that, a semi-automatic data matching process will be executed to find coincidences (e.g. symptoms or signs which are related to the disease in several of the input sources used), in order to generate the best model for a disease. This process can be performed in an automatic way (with the use of heuristics) or in semi-automatic way (with the intervention of an expert in the field). Once the model, which only contains information about the diagnosis process of the disease, is generated, it will be converted to RDF triples and stored in the diagnosis dataset. These RDF triples will be also linked to other datasets of the Linking Open Data (LOD) Project Cloud and with the original sources used (Bio2RDF, OBO, etc...), allowing a future consumption of the Diagnosis Dataset through the LOD Cloud.

2. Conclusions

Despite Linked Data has recently gained wide acceptance in a number of different efforts populating the Linked Open Data (LOD) Cloud, we strongly believe it could also significantly leverage current approaches for publishing and consuming data in the biomedical domain. Fundamentally, in this particular domain, it is critical to rely on the maximum of information available and ready to be queried and explored. This availability is the very competitive advantage for future clinical decision support systems. In the area of diagnosis, physicians will normally need information about entities mentioned by the literature that can occur on a concrete disease. Nowadays, achieving access to this information is a difficult task without knowing where the source of information actually is. The fact of having a dataset with the relevant information regarding this domain is crucial and can suppose a difference in the quality of healthcare application in several areas (clinical practice, knowledge reusing, training, etc.).

References

[1] O'Riain S, Harth A, Curry E. Linked Data Driven Information Systems as an Enabler for Integrating Financial Data. In A.Y. Yap, editor. Information Systems for Global Financial Markets: Emerging Developments and Effects. Palo Alto, USA: IGI-Global. 2012.

[2] Colomo-Palacios R, Sánchez-Cervantes JL, Alor-Hernández G, Rodríguez-González A. Linked Data: perspectives for IT professionals. Int J Hum Cap Inf Technol Prof. 2012; 3(3):1-13.

[3] Wood D. Linking Enterprise Data. New-York, USA: Springer-Verlag, 2010.

[4] Van de Sompel H, Lagoze C, Nelson M, Warner S, Sanderson R, Johnston P. Adding eScience Assets to the Data Web. Proceedings of the 2nd Workshop on Linked Data on the Web (LDOW2009); 2009

[5] Zhao J, Miles A, Klyne G, Shotton D. Linked data and provenance in biological data webs. Briefn Bioinform. 2009; 10(2):139-152.

[6] Samwald M, Jentzsch A, Bouton C, Kallesøe CS, Willighagen E, Hajagos J, Marshall MS, Prud'hommeaux E, Hassanzadeh O, Pichler E, Stephens S. Linked open drug data for pharmaceutical research and development. J Cheminform. 2011; 3(1):19.

[7] Belleau F, Nolin M, Tourigny N, Rigault P, Morissette J. Bio2RDF: Towards a mashup to build bioinformatics knowledge systems. J Biomed Inform. 2008; 41(5):706-16.

[8] Jentzsch A, Hassanzadeh O, Bizer C, Andersson B, Stephens S. Enabling Tailored Therapeutics with Linked Data. Proceedings of the 2nd Workshop on Linked Data on the Web (LDOW2009); 2009

[9] Bertaud-Gounot V, Duvauferrier R, Burgun A. Ontology and medical diagnosis. Inform Health Soc Care. 2011; 0(0):1-11

[10] Burgun A, Bodenreider O, Jacquelinet C. Issues in the classification of disease instances with ontologies. Stud Health Technol Inform. 2005; 116:695-700

[11] Peelen L, Klein MCA, Schlobach S, De-Keizer NF, Peek N. Analyzing Differences in Operational Disease Definitions Using Ontological Modeling. Proceedings of the 11th conference on Artificial Intelligence in Medicine; 2007.