Semantic Annotation of Data on Neurodegenerative Diseases in Patients using Ontologies

Panče Panov[1,2], Ana Kostovska[1,2*], Ilin Tolovski[1,2*] and Sašo Džeroski[1,2,3]
[1]Jožef Stefan Institute, Ljubljana, Slovenia
[2]Jožef Stefan International Postgraduate School, Ljubljana, Slovenia
[3]Centre of Excellence for Integrated Approaches in Chemistry and Biology of Proteins, Ljubljana, Slovenia
{Pance.Panov, Ana.Kostovska, Ilin.Tolovski, Saso.Dzeroski}@ijs.si

In this work, we propose a mid-level ontology for representing various types of data on patients with neurodegenerative diseases. The proposed ontology can be used for semantic annotation of datasets that contain different diagnostic data (clinical, imaging, biomarker, etc) about neurodegenerative diseases and its progression, collected on patients by the hospitals. Having an ontology for describing data on patients with neurodegenerative diseases is important from two different perspectives: (1) from a viewpoint of ontology-based data access (ODBA) [1] it would allow federation queries on data produced and stored at different hospitals; (2) from viewpoint of data analytics it would allow (semi) automatic creation of data analysis workflows based on the datatypes that occur in the datasets, annotated with ontology terms.

The proposed ontology was constructed following best practices from ontology engineering. This involved the use of a top-level ontology (Basic Formal Ontology [2]) as a template, and a set of standard formally defined relations. We heavily reused classes and identified mappings to domain terms that are defined in previously developed biomedical ontologies and vocabularies available at BioPortal (http://bioportal.bioontology.org/). This included domain terms from ontologies and vocabularies covering general medicine (such as SNOMED, NCIT, MESH, LOINC, ICD10), neuroscience (such as NIF, BRCT, NeuroMorpho.org) and neurodegenerative diseases (such as ADO, PDON).

The ontology was constructed in a hybrid fashion (see Figure 1). For this purpose, we used two instances of datasets on patients with neurodegenerative diseases [3,4], originating from two well-known studies concerning neurodegenerative diseases: Alzheimer's Disease Neuroimaging Initiative (ADNI) [5] and Parkinson's Progression Markers Initiative (PPMI) [6]. We also used the domain terms that appear documentation of ADNI and PPMI studies (study objectives, study protocols, study procedures, schedule of activities and others) [7-11], as we believe that the data produced by the hospitals in the project will most probably be subsets of types of data that occur in ADNI and PPMI studies. To address the data analytics perspective, we also reused and extended our previously developed ontology of data types (OntoDT) [12] and ontology of core data mining entities (OntoDM-core) [13] to represent specific domain datatypes that occur in the datasets from the domain of neurodegenerative diseases. The ontology construction and the semantic annotation of the two instances of neurodegenerative diseases datasets was performed using semantic web technologies (RDF, OWL, RDFS), which are currently a popular solution to data and knowledge sharing and integration.

* Student author

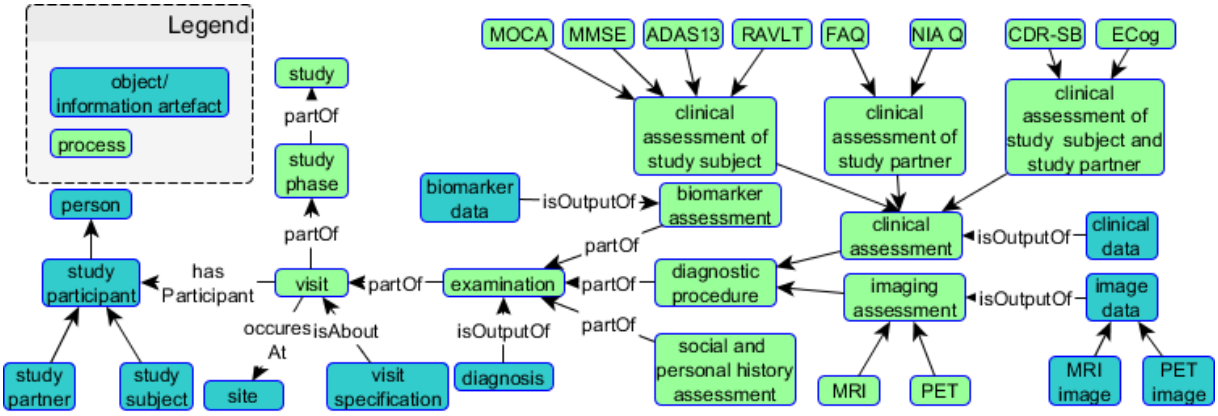

Figure 1. Part of the structure of the constructed ontology for representing data on patients with neurodegenerative diseases. The arrows that are not labeled represent IS-A relations.

Acknowledgements:
Data used in this work were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at https://goo.gl/43TsyJ.. Data used in the preparation of this work was also obtained from the Parkinson's Progression Markers Initiative (PPMI) database (www.ppmi-info.org/data). For up-to-date information on the study, visit www.ppmi-info.org. "PPMI – a public-private partnership – is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including list the full names of all of the PPMI funding partners found at www.ppmi-info.org/fundingpartners.We also acknowledge the European Commission's support through the Human Brain Project (Grant No. 604102).

References:

[1] A. Poggi, D. Lembo, D. Calvanese, G. De Giacomo, M. Lenzerini, and R. Rosati. Linking data to ontologies. J. on Data Semantics, X:133–173, 2008

[2] B. Smith, R. Arp, A. D. Spear. Building Ontologies with Basic Formal Ontology, MIT Press, 2015.

[3] D. Gamberger, B. Ženko, A. Mitelpunkt, N. Shachar, and N. Lavrač. Clusters of male and female Alzheimer's disease patients in the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. *Brain Informatics* 3(3):169-179, 2016.

[4] V. Mileski, D. Kocev, B. Draganski, and S. Džeroski. Multi-dimensional analysis of PPMI data. In Proceedings of the *8th Jožef Stefan International Postgraduate School Students' Conference 2016, Ljubljana, Slovenia,* Part 1:175–178, 2016.

[5] Alzheimer's Disease Neuroimaging Initiative (ADNI) database. URL: http://adni.loni.usc.edu/ (accessed 20.10.2017)

[6] Parkinson's Progression Markers Initiative (PPMI) database. URL: http://www.ppmi-info.org/ (accessed 20.10.2017)

[7] ADNI Procedures Manual. URL: https://goo.gl/UbR2Ad (accessed 20.10.2017)

[8] ADNI GO Procedures Manual. URL: https://goo.gl/od8yUC (accessed 20.10.2017)

[9] ADNI 2 Procedures Manual. URL: https://goo.gl/VQ4pL7 (accessed 20.10.2017)

[10] PPMI Study Protocol. URL: https://goo.gl/QrD7p8 (accessed 20.10.2017)

[11] PPMI Case Report Forms. URL: https://goo.gl/JCpZRw (accessed 20.10.2017)

[12] P. Panov, L. N. Soldatova, and S. Džeroski. Generic ontology of datatypes. *Information Sciences* 329:900-920, 2016.

[13] P. Panov, L. Soldatova, S. Džeroski. Ontology of core data mining entities. *Data Mining and Knowledge Discovery* 28 (5–6):1222-1265, 2014.