Medical Informatics in a United and Healthy Europe
K.-P. Adlassnig et al. (Eds.)
IOS Press, 2009

doi:10.3233/978-1-60750-044-5-317

Entrez Neuron RDFa: A Pragmatic Semantic Web Application for Data Integration in Neuroscience Research

Matthias SAMWALD [a,c,j,1], Ernest LIM [b], Peter MASIAR [b], Luis MARENCO [b,e], Huajun CHEN [d], Thomas MORSE [g], Pradeep MUTALIK [b], Gordon SHEPHERD [g], Perry MILLER [b,e,f], Kei-Hoi CHEUNG [b,e,h,i]

[a] *Digital Enterprise Research Institute, National University of Ireland Galway, IDA Business Park, Lower Dangan, Galway, Ireland*
[b] *Yale Center for Medical Informatics, Yale University School of Medicine, New Haven, USA*
[c] *Konrad Lorenz Institute for Evolution and Cognition Research, Altenberg, Austria*
[d] *School of Computer Science, Zhejiang University, Hangzhou, 310027, China*
[e] *Department of Anesthesiology, Yale University School of Medicine, New Haven, USA*
[f] *Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, USA*
[g] *Department of Neurobiology, Yale University School of Medicine, New Haven, USA*
[h] *Department of Genetics, Yale University School of Medicine, New Haven, USA*
[i] *Department of Computer Science, Yale University, New Haven, USA*
[j] *Section on Medical Expert and Knowledge-Based Systems, Medical University of Vienna, Austria*

Abstract. The amount of biomedical data available in Semantic Web formats has been rapidly growing in recent years. While these formats are machine-friendly, user-friendly web interfaces allowing easy querying of these data are typically lacking. We present "Entrez Neuron", a pilot neuron-centric interface that allows for keyword-based queries against a coherent repository of OWL ontologies. These ontologies describe neuronal structures, physiology, mathematical models and microscopy images. The returned query results are organized hierarchically according to brain architecture. Where possible, the application makes use of entities from the Open Biomedical Ontologies (OBO) and the 'HCLS knowledgebase' developed by the W3C Interest Group for Health Care and Life Science. It makes use of the emerging RDFa standard to embed ontology fragments and semantic annotations within its HTML-based user interface. The application and underlying ontologies demonstrate how Semantic Web technologies can be used for information integration within a curated information repository and between curated information repositories. It also demonstrates how information integration can be accomplished on the client side, through simple copying and pasting of portions of documents that contain RDFa markup.

Keywords. semantic web, neuroscience, OWL, web user interface, reasoning, RDFa, information integration

[1] Corresponding Author: Matthias Samwald, Ferrogasse 45/14, 1180 Vienna, Austria; E-mail: samwald@gmx.at.

1. Introduction

The Semantic Web (SW) has the potential to expand the frontier of neuroinformatics, as it has the capability of providing a standard technological platform for integrating diverse types of data. Recently, a collection of data sets (provided by different sources in different formats) has been converted by the Semantic Web for Health Care and Life Sciences (HCLS) Interest Group [1] into the Web Ontology Language (OWL) format, which forms one basic core of neuroscience knowledge. Some basic queries (written in SPARQL) spanning over several of the integrated data sources have been demonstrated [2]. While the HCLS community continues to convert more data sets to add to its knowledge base, other neuroscience data providers have begun making their data available in OWL format. As the SW neuroscience knowledge base continues to grow, it becomes important to develop SW applications that support neuroscientific queries in a user-friendly fashion.

In this paper, we describe a pilot application called "Entrez Neuron RDFa" that allows the neuroscientist to query multiple OWL ontologies including diverse types of neuroscience data. This application does not require the user to learn any RDF/OWL query language. Instead, it provides a keyword-based interface for users to search the knowledge base.

Entrez Neuron RDFa covers a broad range of neuroscientific information spanning from basic neuronal properties over the description of computational models up to microscopic images and brain anatomy.

2. Methods

The data incorporated into Entrez Neuron RDFa come from several neuroscientific ontologies: NeuronDB and ModelDB (parts of SenseLab [3]), Subcellular Anatomy Ontology (SAO) [4] and an OWL conversion of the Cell Centered Database (CCDB) [5]. These ontologies also serve as entry points for a large array of other ontologies that are mapped to these primary ontologies, but are not yet represented in the user interface of Entrez Neuron. These ontologies were selected because their combined information serves practical needs of research in cellular neurobiology.

The NeuronDB and ModelDB ontologies have been developed at the Yale Center for Medical Informatics and are derived from the SenseLab database collection (http://senselab.med.yale.edu). The NeuronDB ontology is focused on anatomic locations, cell architecture and physiologic parameters of neuronal cells. The ModelDB ontology is a large repository of computational neuroscience models and simulations that are linked to entities in the NeuronDB ontology. The SAO and the OWL conversion of CCDB have been developed at the US National Center for Microscopy and Imaging Research. They contain detailed information about cells, their parts and cytological characteristics.

Most ontologies that are accessed by Entrez Neuron RDFa are based on the Basic Formal Ontology (BFO) [6] and selected ontologies from the Open Biomedical Ontologies (OBO) collection [7]. This ensures that the ontologies adhere to sound ontological principles and are more easily integrated with other ontologies that adhere to similar design paradigms.

The RDF/OWL of the ontologies is stored in an Oracle 11g database [8] that has built-in support for RDF/OWL storage, querying and inferencing. This support for

OWL constructs is required by Entrez Neuron RDFa to handle mappings between ontologies that are formulated through OWL constructs (owl:sameAs, owl:equivalentClass), as well as querying over transitive properties (such as rdfs:subClassOf).

The HTML pages generated by Entrez Neuron RDFa contain embedded Semantic Web statements encoded in accordance to the RDFa standard [9] which has recently been published by the World Wide Web Consortium (W3C). These RDFa statements are dynamically generated from queries to the RDF/OWL store during HTML page generation, and represent relevant fragments of the underlying OWL ontologies. RDFa makes it possible to seamlessly embed RDF/OWL statements (ontology fragments) into HTML documents by using HTML attributes. This, in turn, makes it possible to tightly integrate the human-readable information (text) with the machine-readable information (RDF/OWL statements) in a single document.

3. Results

The application can be accessed on the web at http://ycmi.med.yale.edu/entrez_neuron.html. Links to additional material, demonstrations and the underlying ontologies are available on the web at [10]. The user

Figure 1. A screenshot of the general layout of the Entrez Neuron application. Ontology fragments (encoded in RDFa, not visible) are embedded throughout the entire page, enriching human-readable content with machine-readable information.

interface makes it possible for users to quickly start information exploration with very simple, broad keyword queries, without requiring them to be familiarized with the detailed structure of the ontologies. In other words, the interface guides the users through their information search and offers opportunities to both narrow and broaden the search process in a way that would not be possible without the deep knowledge of the ontology structure.

The user interface has three main panels (see Figure 1). The upper panel is used to initiate keyword searches against the selected data sources. The left panel consists of a hierarchy of brain regions and neurons found in these regions and can be used to constrain search results. This hierarchy is dynamically updated so only those brain regions with matched query results are shown. The hierarchy is composed based on the NeuronDB and SAO ontologies, querying over relations such as 'has part', a relation defined in OBO. The large panel on the right displays the detailed information retrieved from each data source about a class of neurons selected from the tree; each tab corresponds to one data source.

The RDFa-encoded statements from Entrez Neuron can easily be 'mashed up' (integrated) with data from other portals that makes use of RDFa, such as the Science Commons Text Annotations service [11].

4. Discussion and Conclusions

While significant progress has been made in creation and storage of RDF/OWL data and ontologies for neuroscience and biomedicine, the development of practical user interfaces has been lagging behind. Ironically, the greatest strengths of RDF and OWL, namely their flexibility and expressiveness, also pose challenges in the creation of good user interfaces. Developers of Semantic Web applications often face the choice between user interfaces that are either a) not specialized to a certain type of data, but too generic to be user-friendly or b) user-friendly, but very inflexible and specialized to a certain type of data.

We chose to prioritize user friendliness over the ability to display arbitrary data, but some flexibility was retained. Many components of the application, such as the brain anatomy hierarchy, use rather generic RDF queries and work quite independently from the ontologies and knowledge domain of our application. Other components, such as the query results views, use queries that are very specific to the ontologies we are currently using. We found that such restrictions were necessary to keep the user interface well-structured and intuitive. We chose brain anatomy as our structural organization, as it is an intuitive way for neuroscientists to view and access neuron-related information.

We also aimed to find a good balance between simple graphical visualizations (the hierarchy of brain locations) and simple text/table structures, thereby avoiding problems associated with more complex visualization techniques such as graphs.

So far, the evaluation of the practicability of user interfaces for Semantic Web applications was done by the developers of applications themselves, and a small group of scientists. Further studies are needed to reach a more objective and robust evaluation.

The RDFa standard has a great potential for unifying the user-oriented world with the machine-oriented world, a feature that has been somewhat lacking in previous Semantic Web standards. One major advantage of RDFa is the possibility of embedding RDF statements in appropriate parts of the HTML document. In this way,

the RDF can be set into relation with the human-readable document at a very fine level of granularity, e.g., at the level of sentences and images. This makes it possible for the user to selectively "copy & paste" parts of the document and to extract the associated RDF data in parallel. In cases where RDF is used to annotate text, this also makes it possible to trace the RDF back to the annotated text, giving the machine-readable data a human-readable context. The Entrez Neuron RDFa prototype enabled us to demonstrate that RDFa-based metadata can be derived from complex OWL ontologies to enable data integration with external RDFa resources, solely by copying and merging HTML fragments.

Acknowledgements. Thanks to Maryann Martone and Willy WaiHo Wong for providing the OWL ontologies of the CCDB database and assisting with their integration into Entrez Neuron. Thanks to Melliyal Annamalai and Alan Wu from Oracle Corporation for their technical assistance. This work was presented at the WWW 2008 Semantic Web for Health Care and Life Sciences Workshop, but it was not published. This work was supported in part by NIH grants P01 DC04732 and R01 DA021253 and by funding from the Fidelity Foundation. The work presented in this paper has been funded in part by Science Foundation Ireland under Grant No. SFI/08/CE/I1380 (Lion-2). The work was also funded in part by the Konrad Lorenz Institute for Evolution and Cognition Research.

References

[1] http://www.w3.org/2001/sw/hcls/.
[2] Ruttenberg, A. (2007) Harnessing the Semantic Web to Answer Scientific Questions. 16th International World Wide Web Conference, http://www.w3.org/2007/Talks/www2007-AnsweringScientificQuestions-Ruttenberg.pdf.
[3] Crasto, C.J., Marenco, L.N., Liu, N., Morse, T.M., Cheung, K.-H., Lai, P.C., Bahl, G., Masiar, P., Lam, H.Y., Lim, E., Chen, H., Nadkarni, P., Migliore, M., Miller, P.L., Shepherd, G.M. (2007) SenseLab: new developments in disseminating neuroscience information. *Briefings in Bioinformatics* 8(3):150–162.
[4] Fong, L.L., Larson, S.D., Gupta, A., Condit, C., Bug, W.J., Chen, L., West, R., Lamont, S., Terada, M., Martone, M.E. (2007) An ontology-driven knowledge environment for subcellular neuroanatomy. *OWL: Experiences and Directions*, Innsbruck, Austria, CEUR Workshop Proceedings, ISSN 1613-0073, http://CEUR-WS.org/Vol-258/, June 6–7, 2007.
[5] http://ccdb.ucsd.edu/.
[6] http://www.ifomis.uni-saarland.de/bfo.
[7] Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W. et al. (2007) The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology* 25(11):1251–1255.
[8] http://www.oracle.com/.
[9] http://www.w3.org/TR/xhtml-rdfa-primer/.
[10] http://neuroweb.med.yale.edu/senselab/.
[11] http://whatizit.neurocommons.org.