# Using semantic web techniques for validation of cognitive models against neuroscientific data

Balkenius, Christian; Gustafsson, Marie

*Published in:*
Proceedings of AILS 04 ( Report / Lund Institute of Technology, Lund University ; 151)

2004

# Using Semantic Web Techniques for Validation of Cognitive Models against Neuroscientific Data*

Marie Gustafsson
`mat99mgu@student.lu.se`

Christian Balkenius
Lund University Cognitive Science
Kunshuset, Lundagard
S-222 22 Lund, Sweden
`christian.balkenius@lucs.lu.se`

In 1966 Stephen Kuffler established the Department of Neurobiology at Harvard Medical School, and thus, by bringing together physiologists, biochemists and anatomists to focus their efforts on the nervous system, he helped found the discipline of neuroscience. Several decades later, more than 50,000 neuroscientists worldwide study everything from individual molecules to complex behaviors in species from nematode worms to humans [**?**]. Together they fill more than 300 journals, having created one of the largest, most unmanageable datasets in science.

"The sharing, distribution, organization, and managing of data, models, and tools are essential elements in the path from sheer information to knowledge and understanding in neuroscience" [**?**]. In catering to this need, the field of neuroinformatics has emerged. Neuroinformatics is the information science infrastructure of neuroscience. It relates to the tools, databases, models and mechanisms of information flow that serve all of the clinical and research efforts in this field. There have been several attempts to create neuroinformatic systems, but none seem to have reached wide spread adoption yet.

This is especially true in the in the area of cognitive modeling where the goal is to develop computational models of cognitive processes. When these models claim to parallel processes in the brain, it would be useful if this could be tested in an automatic way by invoking some form of neuroinformatic database. In principle, many structural and functional claims about a model could be automatically validated against neuroscientific data if both the model and the data were represented in form suitable for an inference engine.

In the Ikaros project at Lund University Cognitive Science (Balkenius & Morén, 2003), we aim to develop an open infrastructure for system-level modeling of the brain including databases of experimental data, models and structural and functional brain data. A core component is the Ikaros kernel that allows platform independent execution of models and communication with experimental databases as well as external devices such as sensors and actuators. Simulations are controlled by experiment files in XML format that contain structural information about the models in the form of individual components and interactions between them.

The Semantic Web is a vision in which Web resources are machine-processable, where the information can be shared and processed both by automated tools and by humans. Semantic mark-up is central to this sharing of information. For different agents to understand the terms used, ontologies are needed in which the terms are described. An ontology is basically a collection of definitions of concepts. Well-designed, well-defined, and Web-compatible ontology languages with supporting reasoning tools are needed. Recently, both the Resource Description Framework (RDF) and the Web Ontology Language (OWL) have become World Wide Web Consortium (W3C) recommendations. RDF is used to represent information and exchanging knowledge on the Web, while OWL is used for publishing and sharing ontologies.

The Semantic Web is the product of many different desires and influences, where the aim is to make better use of the Web. [**?**] outline three main influences. One is anxiety over the disorder

*The work described here is Marie Gustafsson's thesis at the Department of Compter Science, Lund University.

of digital documents. Another comes from Artificial Intelligence and its "maturing sense of the kinds of computations that can take place given formal representations" (Marshall and Shipman). There is also a "utopian desire to offload the burden of information overload and the complexity of everyday life onto the computer" (Marshall and Shipman). All these desires are justified, but which are realistic? What are appropriate expectations for the Semantic Web?

To what extent can ideas and techniques from the possibly developing semantic web be used for testing cognitive models against neuroscientific data? Data will consist of neuroanatomical connection data from published articles, to be encoded in manner appropriate to the semantic web techniques.

The first part of this work is to decide on how to encode the data. During and after that process, an suitable inference engine has to be found. Once the data is encoded and brought to cooperate with the inference engine, this technique will be tested with direct questions and with parts of cognitive models from the Ikaros project.

A combination of RDF and OWL seems to be an appropriate way of encoding the data. Protg, a knowledge-base program developed at Stanford medical, has an OWL-plugin that will be used for this. The experiment files from Ikaros can be transformed to these formats using XSLT. For reasoning, Jess or RACER will probably be used.

This is an interesting and important topic of investigation for several reasons. One part of it is the relevance of testing and applying semantic web ideas in order to find out what the concept has to offer, what possible limitations exist, as well as what might be good areas of application. Further, what relevance might the semantic web have for neuroinformatics? And how can the techniques to be investigated further our understanding by showing strengths and weaknesses in cognitive models?

Although we hope to develop a minimal system as a proof of concept, many questions remain unanswered. Who will enter data into the knowledge base for a full scale system? One possibility arising with Semantic Web techniques is that the researcher could publish meta data about their findings on their own, using some GUI adapted to the task, then either register it or have it discovered. Making this a reality would require a large scale agreement on the methods and formats used to represent data and there would, of course, be huge problems of verification and credit. Fortunately, these problems lies outside the scope of this work.

# References