Web services as tools within a scientific knowledge engineering environment

G.A.P.C. Burns, W.-C. Cheng, N. Zhang & S. Ghandeharizadeh
University of Southern California
gully@usc.edu, weichenc@usc.edu, ningz@cs.usc.edu, shahram@pollux.usc.edu

Abstract

Neuroscience poses a number of challenges to the concept of "software and data as a service" advocated by Web Services. A primary one is the heterogeneity of the data sources at a conceptual level. In this paper, we describe a knowledge engineering system (called 'NeuroScholar') that acts as a Web Service (WS) for knowledge models based on the published neuroscientific literature. It incorporates several existing web services into a framework that empowers a neuroscientist to represent the neural basis of a predefined physiological phenomenon (e.g.: stress). This paper provides an overview of NeuroScholar and outlines its future research directions.

1. Introduction

A Web Service (WS) is based on the concept of "software and data as a service". Its essence is a remote procedure call (RPC) that consumes and processes some input data in order to produce output data. It is a concept that renders web applications extensible: By identifying and combining WS components, an organization may rapidly develop a new web application. The new web application may consist of web services that span the boundaries of several (if not many) organizations.

Neuroscience, the academic study of how the brain works, is an ideal WS application. This application domain suffers massively from information overload and heterogeneity [7]. This is because the subject is extremely complex with a large amount of data becoming available due to the emergence of 'Neuroinformatics' as a discipline [1] and there is little standardization within neuroscientific nomenclature (see [11] for a historical review). The emerging technology of web services provides standard methods of communication and consolidation with immediate, direct applications within this subject.

We present a knowledge engineering environment for neuroscientists, based on the published literature, that seeks to permit users to synthesize new theories about how the brain works [2]. This environment, called 'NeuroScholar', uses web services in three separate ways: by (a) utilizing existing web services, (b) acting as a specialized web service for its own content and (c) making use of software engineering tools to automatically generate web service applications from a design specification. Within this paper, we will introduce the NeuroScholar application and describe how web services offer a powerful methodology for solving problems that would otherwise be intractable. We will also describe the state of implementation within the system and our future research directions.

2. Approach

The NeuroScholar system itself is a prototypical knowledge engineering environment to enhance neuroscientists' capability to make sense of an extremely complicated subject [2]. Our approach is to provide easy-to-use tools that permit non-computational scientists to access and process heterogeneous data sources from within our application, so that then, they can synthesize the product of these tasks into a single unified construct: the knowledge model.

Initially, they may manage their online library of downloaded documents (we only support the use of Portable Document Format files, 'PDF's, since they are universally used within the academic sphere). They may use a highlighting tool from within the application to any select text or figures of interest and then compile a library of text or image fragments in a database (the 'fragmenter', see [6]). They may use any combination of a set of specialized tools to build structured representations based on the human-readable contents of the fragments. These structured representations could be anything: editable maps of brain sections, flowcharts representing different experimental designs, large scale networks of interconnected brain regions, or neural modeling tools (see [6] for two examples). The large-scale NeuroScholar application can use smaller applications (written in Java) as plugins as long a suitable adaptor class is written to link them to the main application. The NeuroScholar system represents these data as nodes in a graph, permitting the relationships between them to be viewed as well (as edges linking the nodes). The visibility of types of nodes and edges may be turned on and off, allowing the user to scrutinize different combinations of data nodes and their interrelationships.

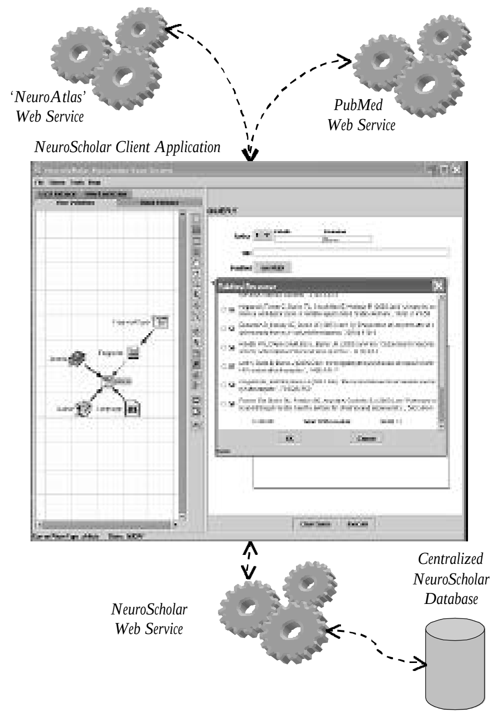

Figure 1. Application Organization

Figure 1 illustrates how the NeuroScholar system is organized. Within this application, we make use of available web-services within the plugin tools and specialized form controls. One such form control (the 'Article Robot') accesses the National Library of Medicine's PubMed Web Service to lookup citation information, retrieve data and even assist end users to navigate to full-text articles for subsequent download and fragmentation.

An example of a plugin tool that demonstrates the principle of NeuroScholar's use of web services is the Atlas Mapper. This is a tool that allows users to draw delineations of regions on specified sections of a brain atlas [2]. This permits neuroscientists to accurately state where they believe a volume of brain tissue is located in the standardized framework of a brain atlas. The images that these delineations are protected by copyright and there are a total of 73 separate atlas levels [10]. A web-service removes the need for any local storage of files and can be used to provide low-resolution images (with appropriate copyright statements) following the agreement of their copyright owners. We envisage web services being used at all levels of the NeuroScholar to provide access to standardized data sets or named services (such as simulation or analysis processes).

The data contained within the centralized NeuroScholar database (see Figure 1) will be served to the end-user as a web service providing information concerning the data of interest (see [3] for a complete description of the data model underlying the NeuroScholar system). This web service could interact with the NeuroScholar client application of any given user to examine the work of all users of the system, and could also provide a range of data services to anyone wishing to use it. This would permit users to examine the contents of the system without the need for the specialized plugins contained within the client application.

Finally, the NeuroScholar system is built on a software engineering framework called the View Primitive Data Model framework (or 'VPDMf', [5]) that operates on a data model design to generate a database, a user interface and simple non-specialized web services. Such a generated system could be used to as a small repositories of local data or could serve as the template for a more complex application.

One of the reasons why web services are ideal for the purposes of the NeuroScholar application is that different neuroinformatics developers or groups tends to use a different data model (with the exception of some standard data formats such as NeuroML, [8]). For example, there are several neuroinformatics databases concerned with neural connectivity and each one differs substantively in their internal representation ([4, 9]). If each of these systems presented themselves as a web service within a clearly defined namespace, then XSL-based transformation approaches could straightforwardly allow data to be mediated between sources.

3. Current Status

At present, NeuroScholar is still a prototype application. The system is functional, but only one plugin (the Fragmenter) has been incorporated into the system's implementation. The system's behaviour as a web service is still in the alpha phase of development. The most mature section of the system is the VPDMf system builder, which is regularly used to build the test application used in development.

4. Future Activities

Our specific goal is to address real problems within Neuroscience and to use that as a springboard into developing applications for other disciplines. To that end, we are collaborating with experimental neuroscientists to investigate neural systems involved in the physiological phenomenon of stress. We are incorporating statistical analysis engines to perform inference, consistency checking and data analysis. The ultimate vision of this work is that experimental neuroscientists should be able to construct models of their own knowledge within an online environment without having any specialized computational expertise. They should be able to share these models with each other so that their ideas can be understood, used, tested and maybe even adopted by other scientists in their models.

5. References

[1] Ascoli, G., E. De Schutter and D. Kennedy, "An Information Science Infrastructure for Neuroscience." *Neuroinformatics*, Humana Press, Inc, Ithaca NY, 2003, v1 (1) pp: 1-2.

[2] Burns, G., "Knowledge Mechanics and the Neuroscholar project: a new approach to neuroscientific theory." from M. Arbib and J. Grethe *Computing the Brain*, Academic Press, San Diego, 2001, pp: 319-336.

[3] Burns, G. A. "Knowledge management of the neuroscientific literature: the data model and underlying strategy of the

NeuroScholar system." *Philos Trans R Soc Lond B Biol Sci* (2001) Royal Society, London, v356(1412) pp: 1187-208.

[4] Burns, G. A. P. C., *Neural connectivity of the rat: theory, methods, and applications.* D.Phil dissertation, laboratory of Physiology. University of Oxford, Oxford 2003.

[5] Burns, G. A. P. C., F. Bian, W.-C. Cheng, S. Kapadia, C. Shahabi and S. Ghandeharizadeh. "Software engineering tools and approaches for Neuroinformatics: the design and implementation of the View-Primitive Data Model framework ('VPDMf')." *Neurocomputing*, Elsevier, Amsterdam, 2002, v44-46, pp:1049-1056.

[6] Burns, G. A. P. C., A. M. Khan, S. Ghandeharizadeh, M. A. O'Neill and Y.-S. Chen (2002). "Tools and Approaches for the Construction of Knowledge Models from the Neuroscientific Literature." *Neuroinformatics*, Humana Press, Inc, Ithaca NY, 2003, v1 (1), pp: 81-109.

[7] Chicurel, M., "Databasing the brain." *Nature*, 2000 v406(6798), pp: 822-5.

[8] Goddard, N. H., M. Hucka, F. Howell, H. Cornelis, K. Shankar and D. Beeman, "Towards NeuroML: model description methods for collaborative modelling in neuroscience." *Philos Trans R Soc Lond B Biol Sci*, Royal Society, London, 2001, v356(1412), pp: 1209-28.

[9] Stephan, K. E., L. Kamper, A. Bozkurt, G. A. Burns, M. P. Young and R. Kotter, "Advanced database methodology for the Collation of Connectivity data on the Macaque brain (CoCoMac)." Philos Trans R Soc Lond B Biol Sci 2001, Royal Society, London, v356(1412), pp: 1159-86.

[10] Swanson, L. W. (1998). *Brain Maps: Structure of the Rat Brain.* , Elsevier Science, Amsterdam, 1998.

[11] Swanson, L. W. (2000). "What is the brain?" *Trends Neurosci*, Elsevier Science, Amsterdam, 2000, v23(11): pp: 519-27