

# **Tools for neuroinformatic data exchange: An XML application for neuronal morphology data**

Weihong Qi<sup>1</sup>, and Sharon Crook<sup>2</sup>

<sup>1</sup> Department of Computer Science, University of Maine, Orono, ME 04469

<sup>2</sup> Department of Mathematics and Statistics, University of Maine, Orono, ME 04469

## **Abstract**

We describe an Extensible Markup Language (XML) application that defines a common data format for the storage and processing of neuronal morphology data. The application, MorphML, is consistent with current neuronal digitization methods and is also compatible with existing simulation software environments for creating compartmental models of neurons based on morphology data. We also describe a tool, nemo2xml, which generates MorphML documents corresponding to digitized neuronal morphology files for one particular data format, the Nevin or binary branch tree (BBT) syntax.

Keywords: Neuronal morphology data; XML application; Digitized neurons

## **Introduction**

Existing neuronal morphology data have been obtained using many different neuron tracing systems such as Eutectic's Neuron Tracing System, Microbrightfield's Neurolucida, the Nevin BBT syntax, and the Douglas syntax (e.g. [2]). Generally, these systems represent neuronal structure using a collection of points and connections in three dimensions.

The different tracing systems use a variety of representations and vocabularies to describe the connectivity between points based on different conceptual models for the morphology data. In each case, the generated files are flat text files so that cell structure is not available explicitly without reconstructing the cell. Consequently, each system requires a different reconstruction tool for use in spatial analysis or for creating compartmental models for simulation studies. There has been growing interest in defining a common data format for describing and exchanging neuronal cell and model information using Extensible Markup Language (XML) technology [1,3]. This work addresses the development of an XML application, MorphML, as a common data format for neuronal morphology data and also describes associated tools.

### **MorphML: An XML application for neuronal morphology data**

XML provides a portable format for computer documents [4,6,7] where data are surrounded by text that describes the structure and semantics of the data; that is, the important details about the content of the document are explicitly described by a set of tags known as a markup. XML allows developers to define tags as they are needed, and for the purpose of interoperability, researchers in the same domain often agree to use a particular vocabulary for the markup, resulting in a set of tags known as an XML application. The markup permitted in a particular XML application can be documented using a document type definition (DTD) or a schema language such as the W3C XML Schema Language which includes more extensive validation facilities [8].

The XML application that we have developed, MorphML, provides a markup for representing digitized neuronal morphology files. Our chosen vocabulary is an abstract representation language that is independent of any particular digitization syntax so that these MorphML documents have the same format regardless of the syntax of the original digitized morphology files. Because the documents are also designed to be stored in databases, viewed by visualization tools, and passed to a variety of analysis and simulation software, the data format has not been modeled on the internal structure of any specific application software. However, for all downstream applications, neuronal cell structure information is essential. Thus, in addition to storing original information about the digitization, MorphML documents provide reusable information about cell structure as shown in Figure 1.

<Insert Figure 1 near here.>

The application has been well documented since MorphML will be used to share neuronal cell morphology data among neuroscience research labs. The schema is available in W3C Schema Language (morphology.xsd) and as a DTD (morphology.dtd) as a guide for using the application and for validation. Every element in the schema is a named complex type, which facilitates data type reuse and application extension. To facilitate easy handling of white space by downstream application software, every element contains only complex content, and no element contains character

data. The schema was developed in consultation with several research groups that store and use morphology data and has been defined to accommodate many different types of cells, cell features, and other histological features as demonstrated in Figure 2. The complex types and elements shown in Figures 1 and 2 are only a subset of the available data types in MorphML. The schema also provides a comprehensive and controlled support for extending the markup using the `extension`, `include`, `redefine`, and `import` elements.

<Insert Figure 2 near here.>

Another XML application, NeuroML, has been developed to support discussion, development, and exchange of neuronal models [3,9]. MorphML can be used separately from NeuroML as a stand-alone application for representing morphological data. However, it is also suitable for inclusion in NeuroML using the simple inclusion method. An element of NeuroML can be defined as `Morphology` in order to create a container element that contains the definition of our schema. Because NeuroML is defined using object-oriented Java classes, our application is also defined by a Java class package. Thus, MorphML can also be incorporated into NeuroML by writing descriptive scripting code in which both packages are loaded. The attributes `elementFormDefault` and `attributeFormDefault` are set to `qualified` to force every element from MorphML to be explicitly qualified with the namespace and avoid potential tag-name collision. We have also developed a Java Application

Program Interface (API) for MorphML including routines to parse and generate MorphML documents to and from an instantiated Java object tree.

### **The tool nemo2xml**

The ultimate goal of our work is to coordinate the many components of a simulation environment that can be used to store, reconstruct, model, and study networks of neurons from the cricket cercal sensory system. Interface modules are required to make these components work together as a system. The cricket neurons were digitized using the Nevin syntax and saved as BBT files. Because these digitized morphology files generally contain a large number of data points, the corresponding MorphML documents must be created automatically by software tools. Therefore, a Java tool, nemo2xml, was developed to automate the generation of a MorphML document from one or more BBT files. Because the MorphML document is based on cell structures, a neuron must first be reconstructed in memory from the input BBT files. Then nemo2xml uses the dynamic tree structure to generate the correct tags for encoding in MorphML. Many tools are available that can be used to validate documents generated by nemo2xml against the defined MorphML DTD or schema such as XMLSPY [10], Topologi's Schematron Validator [11], and Xerces [12]. When programs that process MorphML documents are developed in the future, the parser's API should be designed to validate documents.

### **Acknowledgements**

The MorphML package and nemo2xml are both available at <http://www.umemat.maine.edu/faculty/crook/MorphML.html>. We would like to thank Robert Cannon, Hugo Cornelis, Laszlo Zaborszky, and Zoltan Nadasdy for helpful discussions of neuronal morphology and the MorphML schema design. We also gratefully acknowledge the support of the National Science Foundation through IBN 0091117. This material is also based on work supported by the National Science Foundation at the Mathematical Biosciences Institute under Agreement 0112050.

## References

- [1] D. Beeman, J.M. Bower, E. De Schutter, E.N. Efthimiadis, N. Goddard, and J. Leigh, The GENESIS simulator-based neuronal database, in: S.H. Koslow and M.F. Huerta eds., *Neuroinformatics: An Overview of the Human Brain Project*, (Lawrence Erlbaum Associates, Mahwah, NJ, 1997) 57-80.
- [2] J. Capowski, *Computer Techniques in Neuroanatomy* (Plenum Press, New York, 1989).
- [3] N.H. Goddard, M. Hucka, F. Howell, H. Cornelis, K. Shankar, and D. Beeman, Towards NeuroML: Model description methods for collaborative modelling in neuroscience, *Philosophical Transactions of the Royal Society B* 356 (2001) 1209-1228.
- [4] E.R. Harold and W.S. Means, *XML in a Nutshell* (O'Reilly & Associates, Inc., Sebastopol, CA, 2002).
- [5] M.L. Hines and N.T. Carnevale, The NEURON simulation environment, *Neural Comp.* 9 (1997) 1179-1209.

- [6] H. Maruyama, K. Tamura and N. Uramoto, XML and Java<sup>TM</sup>:  
Developing Web Applications (Addison Wesley Longman, Inc.,  
Reading, MA, 1999).
- [7] B. McLaughlin, Java and XML (O'Reilly & Associates, Inc.,  
Sebastopol, CA, 2001).

### **Web Sites**

- [8] <http://www.w3.org>
- [9] <http://www.neuroml.org>
- [10] <http://www.xmlspy.com>
- [11] <http://www.topologi.com>
- [12] <http://xml.apache.org>

### **Figure Captions**

Figure 1. The complex type, CellType, has a number of subelements that can be used to describe cell morphology at different levels of complexity.

Figure 2. Additional elements have been developed for describing other anatomical and histological features such as spines, synaptic locations, and fiducial structures.

### **Biographical Information**

Weihong Qi received a Bachelor's degree in Biology from Wuhan University in 1994. She was admitted to the graduate school of the Chinese Academy of Science and

graduated in 1997 with a Master's degree in Microbiology. In 2001 she received a Ph.D. in Biological Sciences from the University of Maine. She also received a Master's degree in Computer Science from the University of Maine in 2002. After graduating, she began a postdoctoral research position in Bioinformatics at Michigan State University.

Sharon Crook received a Ph.D. in Applied Mathematics from the University of Maryland in 1996 for her work on the role of delay in cortical dynamics. In 1997 she became a postdoctoral researcher in the Center for Computational Biology at Montana State University. While there, she worked with the Miller and Jacobs group modeling sensory processing in the cricket cercal system. Currently, she is an Assistant Professor at the University of Maine doing research in the area of mathematical neuroscience.



**Figure 1**

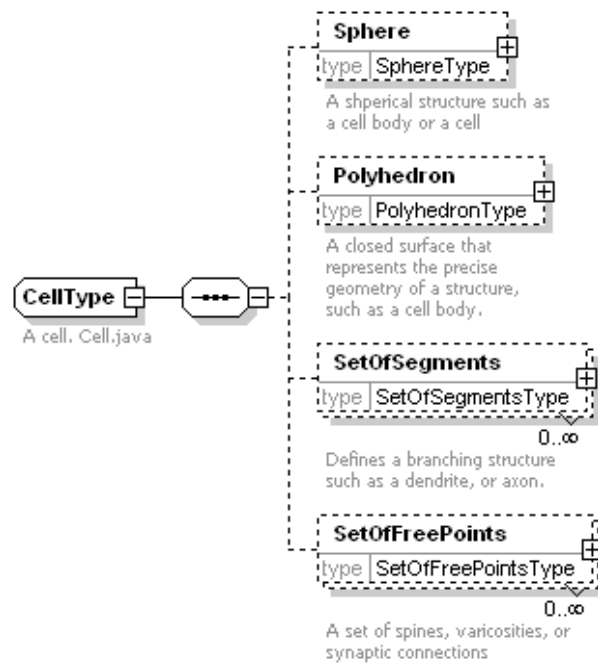


Figure 2

