

## News Item

---

### Semi-Automated Population of an Online Database of Neuronal Models (ModelDB) With Citation Information, Using PubMed for Validation

Andrew P. Davison,<sup>1</sup> Thomas M. Morse,<sup>\*2</sup> Michele Migliore,<sup>2,3</sup> Gordon M. Shepherd,<sup>2</sup> and Michael L. Hines<sup>2</sup>

<sup>1</sup>Centre National de la Recherche Scientifique, 1 Avenue de la Terrasse (Bat. 33) 91198 Gif-sur-Yvette, France; <sup>2</sup>Department of Neurobiology, Yale University School of Medicine, 333 Cedar Street, New Haven, CT; <sup>3</sup>Institute of Biophysics, National Research Council, Palermo, Italy

#### Abstract

Citations play an important role in medical and scientific databases by indicating the authoritative source of the data. Manual citation entry is tedious and prone to errors. We describe a method and make available computer scripts which automate the process of citation entry. We use an open citation project PERL

module (PARSER) for parsing citation data that is then used to retrieve PubMed records to supply the (validated) reference. Our PERL scripts are available via a link in the web references section of this article.

**Index Entries:** Populating databases; literature-search; neuroscience database; interoperability; data entry.

Citations play a valuable role when included in medical and scientific databases. They indicate the association of the data with authoritative research reports or broader review articles. Searches based on citations facilitate further inquiry into the literature. Tools and methodologies that populated our database, ModelDB (*see* ModelDB and SenseLab under Web Resources later), with citation information, are

available for investigators and developers who wish to include citation based tools in their databases.

The main purpose of ModelDB is to make the computer code that describes a neuronal model publicly available (Migliore et al., 2003; Davison et al., 2003; Mirsky et al., 1998; Miller et al., 2001). ModelDB has grown to contain over 100 models. The usefulness of

\*Author to whom all correspondence and reprint requests should be sent.  
E-mail: tom.morse@yale.edu.

any database depends on the database being well populated and on having search methods that help the user find the element(s) of interest. We provide several search methods to find models of interest on the ModelDB homepage.

Through these search methods links one can arrive at the web page for a model with just two mouse clicks. The web page for each model provides information including the source code, and tools for finding related models, among which is the "Citation Browser."

The Citation Browser link loads a page as seen in Fig. 1A. This page in turn allows the user to view an annotated list of the titles of (centered headers) and reference bibliographies (left column) from the one or more papers that are associated with the model, and a list of papers that cite the paper(s) (right column) associated with the model. Each paper in ModelDB that is cited by two or more papers in the database is highlighted in green and has a hyperlink to the Citation Browser page for that paper. Each paper that has a model stored in ModelDB is annotated by a bulleted, pink-highlighted hyperlink to that model. Related papers and models may be found just by following the hyperlinks.

The Citation Browser supplies a convenient method for finding papers and models of interest. In order to implement the Citation Browser, it is necessary for the citation lists of at least model-associated papers to be entered into the database. Such population would be tedious if done by hand. The only complete electronic public source of citation information, ISI Web of Knowledge (see "Web Resources"), requires a subscription/license and does not indicate if a paper's model is electronically available. There are research-oriented open citation projects (see "Citebase Search and ParaCite" under "Web Resources"), but they are still in preliminary stages. We therefore developed a web-based interface to automate the entry of citation

lists obtained from the electronic versions of the original articles, from PDF or HTML files, or from scanning followed by optical character recognition (OCR, Hewlett-Packard bundled software: OfficeJet G Series Scan version 2.0). By leveraging open-source citation-parsing tools (Web Resources: The Open Citation Project), we were able to produce web form interfaces that automate the process of entering citation data from these electronic versions of journal articles, and that also validate the data using PubMed.

The user interface for entering citations is an HTML form (Fig. 1B). In summary, the procedure is as follows (see "Online Supplement" for a simplified flow-chart and PERL code excerpts):

- The paper for which the reference list is to be entered is retrieved from ModelDB using either its PubMed ID, or its ModelDB Object ID (the latter is for references that are not in PubMed—mostly books and book chapters; see Marengo et al., 2003 for more detail on the architecture of ModelDB). If the paper is not in the ModelDB database but has a valid PubMed ID, the paper is entered into the database.
- The list of references is pasted into the text box. The principal formatting requirement is that each reference be on a single line.
- Each reference is parsed to extract the list of authors, the name and volume of the journal, and the starting page of the article.
- References that have been successfully parsed are sent as a batch to PubMed, which returns, for each reference, either a PubMed ID or a message that the reference was not found.
- The PubMed IDs are sent back to PubMed, which returns an XML file containing data for each reference. The XML is parsed, the data entered into ModelDB, and the reference linked to the original paper whose reference list is being entered.

**A**Address  <http://senselab.med.yale.edu/senselab/ModelDB/indexCitations.asp?mode=p&id=7177&wcolor=True>Go 

## Citation Relationships

Key: [Link to a Model](#) [Reference cited by multiple papers](#)

De Schutter E, Bower JM (1994) An active membrane model of the cerebellar Purkinje cell. I. Simulation of current clamps in slice. *J Neurophysiol* 71:375–400 [PubMed]

- [Cerebellar purkinje cell: De Schutter and Bower 1994](#)

References and models cited by this paper	References and models that cite this paper
<a href="#">Adams RB, Constanti A, Brown DA, Clark RB (1982)</a>	<a href="#">Aradi I, Holmes WR (1999)</a>
<a href="#">Adelman JP, Shen KZ, Kavanaugh MP, Warren RA, Wu YN, Lagrutta A, Bond CT, North RA (1992)</a>	<a href="#">Blackwell KT (2001)</a>
<a href="#">Aldrich RW, Stevens CF (1987)</a>	<a href="#">Coop AD, Reeke GN (2001)</a>
<a href="#">Alzheimer C, Schwindt PC, Grill WE (1993)</a>	<a href="#">De Schutter E (1997)</a>
<a href="#">Bernander O, Douglas RJ, Martin KA, Koch C (1991)</a>	<a href="#">De Schutter E (1998)</a>
<a href="#">Bhalla US, Bittich DH, Bower JM (1992)</a>	<a href="#">De Schutter E (1999)</a>
<a href="#">Bhalla US, Bower JM (1993)</a>	<a href="#">De Schutter E, Smolen P (1998)</a>
• <a href="#">Olfactory Mitral cell: Bhalla and Bower 1993</a>	<a href="#">Desjardins AE, Li YX, Reinker S, Miura RM, Neuman RS (2003)</a>
<a href="#">Bossu JL, Dupont JL, Feltz A (1989)</a>	<a href="#">Doliron B, Longtin A, Turner RW, Maier L (2001)</a>
<a href="#">Bower JM, Koch C (1992)</a>	<a href="#">Engel J, Schultens HA, Schild D (1999)</a>
<a href="#">Bush PC, Sejnowski TJ (1991)</a>	<a href="#">Genet S, Delord B (2002)</a>
<a href="#">Calabrese RL, De Schutter E (1992)</a>	<a href="#">Kotter R (1999)</a>
<a href="#">Chan CY, Hounsgaard J, Nicholson C (1988)</a>	<a href="#">Kotter R, Feizelmeier M (1998)</a>
<a href="#">Chen CF, Hess P (1990)</a>	<a href="#">Maex R, Schutter ED (1998)</a>
<a href="#">Connor JA, Stevens CF (1971)</a>	<a href="#">Mainen ZF, Joerges J, Huguenard JR, Sejnowski TJ (1995)</a>
<a href="#">Cooley JW, Dodge FA (1966)</a>	• <a href="#">Spike Initiation in Neocortical Pyramidal Neurons from Mainen et al 1995</a>
	<a href="#">Miyasho T, Takagi H, Suzuki H, Watanabe S, Inoue M, Kudo Y, Miyakawa H (2001)</a>

**B**

Address <http://senselab.med.yale.edu/senselab/ModelDB/CitationEntry.asp> Go

**SenseLab** **ModelDB**

### Reference List Entry

Here we enter papers into ModelDB, linking them to one paper as its references. If you wish to enter papers without linking them together [click here](#).

This tool uses the [PubMed database](#) from the [National Center for Biotechnology Information \(NCBI\)](#). Please read the [NCBI's Disclaimer and Copyright notice](#) before using this tool.

PubMed ID of paper  or Object ID of paper  (Object ID takes precedence)

Click "Submit" to check that the PubMedID you have entered above is valid, then paste the list of references (one per line) into the box below and click "Submit" again. If you do not know the PubMed id you can find the paper may already be in the database under an object id. To find that search [here](#) and use the number in the URL of the reference.

If there is no PubMed id and no object id then use the manual entry button.

Any references that do not result in a PubMed ID number will be left in the below box, so that you can edit them and try again. The information that is parsed from the raw reference will be appended to the reference in the following format:

```
<| journal | year | volume | firstpage |>
```

Edit the original reference, not the parsed information.

If you know the PubMed ID for a reference, you can replace the reference line with the ID.

Submit

[Add paper manually.](#)

Fig. 1. (A) Example of the Citation Browser page. (B) The "Reference List Entry" page uses existing PubMed IDs and reference lists from papers to automatically extract PubMed data and enter it into our (Citation Browser) database.

- References that have been successfully entered into ModelDB are removed from the text box. References that have not been successfully parsed or that are not in PubMed are left in the box.
- If there is an error in the reference, e.g., incorrect page number, this can be corrected in the text box and the reference resubmitted to PubMed.
- For references that can be found by manually searching PubMed, but cannot be successfully parsed, the PubMed ID can be entered instead of the reference string.
- References that resist parsing and/or are not in PubMed can be transferred to a form that allows manual reference entry. After each submit of the manual entry page, the papers are searched-for in ModelDB and, if found, a drop-down list box appears with all the papers from the first author. The user may choose to select the paper if it happens to be in ModelDB already.

ModelDB's semi-automated citation entry software has enabled us to populate ModelDB with 400 reference lists consisting of over 10,000 papers and 11,000 distinct author names (note that the bulk of computational paper references are to the experimental literature). Including the citations in ModelDB allows us to construct a citation database that is focused on the computational neuroscience domain. This modeling citation database allows investigators to use our web application tool, the Citation Browser, to explore how models were used by other researchers.

We hope that this article will help database developers create similar tools to assist the importation of relevant references to the developer's own context. Citations provide an invaluable resource for medical and scientific databases, and indeed for any database that maintains data for which there is a substantial associated literature. The research reports and review articles contain the authoritative

description of the item(s) stored in the database and provide the larger context in which the usefulness of the item can be understood. The references in the ModelDB papers also provide a starting place for searches based on citation indices, useful for finding and understanding the work that the paper built upon, and also for seeing which studies potentially built upon the work under consideration.

Citations, because of their varying degrees of relatedness to the article of interest (Web Resources Garfield, 1994a), are frequently the objects of search and retrieval engines. PubMed is premised on this concept, as are online versions of most journal articles, and indeed, most of modern Library Science. The Institute for Scientific Information (ISI) supplies notable citation search tools. They offer complete lists of citation relationships for multidisciplinary analysis. A common use of citations is the ranking of the importance of papers, individuals, institutions, countries, and journals (Web Resources: ISI Web of Knowledge and Garfield, 1994b) by the number of citations that point to each of these entities. Our citation database is, by comparison, narrower in focus, being limited to publications in the field of computational neuroscience and the references contained therein. We have added value to our Citation Browser's reference lists by annotating them with a color coding that indicates if a publication has a model available in ModelDB. This can help investigators who are interested in modeling to know that this paper may have special interest to them.

**Future Directions:** We are exploring tools that can identify model papers automatically from electronically available journals (Crasto et al., 2003). These same tools will enhance the use of citations by attempting to annotate them automatically with the keywords that we currently use to search for models. Researchers could then more easily find papers on neuronal and network modeling topics.

## Acknowledgments

We would like to acknowledge Luis Marengo and Chiquito Crasto for helping with technical issues, discussions on past and future directions, and proofreading.

We gratefully acknowledge the support of the National Institute for Deafness and Other Communicative Disorders, National Institute of Mental Health, the National Institute of Neurological Disorders and Stroke, National Science Foundation, and the National Cancer Institute from their combined Human Brain Project (grant number 5P01DC004732-04).

## Web Resources

Citebase Search: <http://citebase.eprints.org/cgi-bin/search>

Garfield E. (1994a) Expected Citation Rates, Half-Life, and Impact Ratios: Comparing Apples to Apples in Evaluation Research: <http://www.isinet.com/essays/citationanalysis/10.html>

Garfield E. (1994b) The Relationship Between Citing and Cited Publications: A Question of Relatedness: <http://www.isinet.com/essays/useofcitationdatabases/5.html>

ISI Web of Knowledge—Essential Science Indicators: <http://www.isinet.com/media/presentrep/tspdf/sem-esi-1-0-0702.pdf>

ModelDB: <http://senselab.med.yale.edu/senselab/modeldb>

ParaCite: <http://paracite.eprints.org>

PubMed (National Center for Biotechnology Information, National Library of Medicine, US): <http://www.pubmed.gov>

SenseLab Citation entry and browser code: <http://senselab.med.yale.edu/senselab/modeldb/neuroinf2004.zip>

The Open Citation Project: <http://opcit.eprints.org>

## References

Crasto C. J., Marengo L., Migliore M., Mao B., Nadkarni P. M., Miller P. L., and Shepherd G. M. (2003) Text Mining Neuroscience Journal Articles to Populate Neuroscience Databases. *Neuroinformatics* 3, 215–238.

Davison A. P., Morse T. M., Migliore M., Marengo L., Shepherd G. M., and Hines M. L. (2003) ModelDB: A Resource for Neuronal and Network Modeling, in *Neuroscience databases: a practical guide*, R. Kotter, ed., Kluwer Academic, Boston, pp. 99–109.

Marengo L., Tosches N., Crasto C., Shepherd G., Miller P. L., and Nadkarni P. M. (2003) Achieving evolvable Web-database bioscience applications using the EAV/CR framework: recent advances. *J. Am. Med. Inform. Assoc.* 10, 444–453.

Migliore M., Morse T. M., Davison A. P., Marengo L., Shepherd G. M., and Hines M. L. (2003) ModelDB Making models publicly accessible to support computational neuroscience. *Neuroinformatics* 1, 131–134.

Miller P. L., Nadkarni P., Singer M., Marengo L., Hines M., and Shepherd G. (2001) Integration of multidisciplinary sensory data: a pilot model of the human brain project approach. *J. Am. Med. Inform. Assoc.* 8, 34–48.

Mirsky J. S., Nadkarni P. M., Healy M. D., Miller P. L., and Shepherd G. M. (1998) Database tools for integrating and searching membrane property data correlated with neuronal morphology. *J. Neurosci. Methods* 82, 105–121.