Biomedical Knowledge Engineering tools based on Experimental Design: A case study based on neuroanatomical tract-tracing experiments

Gully APC Burns
USC Information Sciences Institute
4676 Admiralty Way
Marina del Rey, CA 90292
burns@isi.edu

Thomas A. Russ
USC Information Sciences Institute
4676 Admiralty Way
Marina del Rey, CA 90292
tar@isi.edu

ABSTRACT

Curating information from the literature for storage in databases is a crucial task in biological research, and many groups assign a particular person or team to that process. We are developing a *general-purpose* approach to representing the design of a biomedical experiment and provide a manageable template for that experiment's data: "Knowledge Engineering from Experimental Design" (KEfED). The KEfED model allows us to impose a formal and well-grounded structure on the data contained in scientific articles, based on relationships between the dependent and independent variables that make up a scientific experiment. We use this structure to add value to the data contained in the articles by performing directed information retrieval, adding basic forms of reasoning using additional information such as anatomical atlases and taxonomies from external ontologies. We use a graphical interface for constructing KEfED models and a first-order logic reasoning system that performs inference over such models.

Categories and Subject Descriptors

I.2.4 [**Artificial Intelligence**]: knowledge representation formalisms and methods; J.3 [**Life and Medical Sciences**]: medical information systems, biology and genetics

General Terms

Design, Experimentation, Human Factors, Management

Keywords

Knowledge Acquisition, KEfED, Scientific Reasoning, Ontologies

1. THE "KEFED" MODEL

The premise of the KEfED approach [3] is that scientific reasoning is based on assertions about our understanding of phenomena that are *supported by data*. This might involve taking measurements of an appropriate dependent variable that are different from one another in a statistically-significantly way when the value of a relevant independent variable is changed. Therefore, our approach centers on the relationship between **independent variables** (providing the *constraints* governing an experimental design) and **dependent variables** (providing the *measurements*). Dependent variables are treated as multi-dimensional spaces where each dimension is governed by a single independent variable.

2. TRACT-TRACING EXPERIMENTS

Tract-tracing experiments [1, 6] are used to investigate the connections between different regions of the brain. These experiments injecting tracer chemicals into carefully defined locations of the brains of laboratory animals. The tracer chemical is transported along neural pathways and deposited in other regions of the brain that are connected to the injection site. An examination of the location where the tracer chemical ends up reveals the presence of connections between spatially separate parts of the brain.

3. SYSTEM DESIGN

We use a graphical editor to build a description of the work and data flow of an experiment. This produces a graphical model, such as the model of the tract-tracing experiments shown in Fig. 1. The diagrammatic representation allows us to the protocol to identify the context for data values. The logical representation design is currently a manual step that we plan to automate in future development.

The data is entered in a tabular form automatically derived from the KEfED model of the experiment type and then transformed into the logical representation used by the reasoning engine.

3.1 KEfED structure editor

We used Kap-Lab's Diagrammer program (`http://labs.kapit.fr/display/diagrammer/Diagrammer`) as the basis for our structure editor tool. This is a Flex component that

K-CAP'09, September 1–4, 2009, Redondo Beach, California, USA.
ACM 978-1-60558-658-8/09/09

permits users to construct graphs from elements that defined as SVG-based primitives. Diagrammer links the graphical elements to ActionScript classes defined by external developers. As the basis for these internal data-structures, we adopted the graph-based representations from the Flare Prefuse ActionScript library (`http://flare.prefuse.org/`), in order to use their graph-traversal and shortest-path algorithms.

Figure 1 shows a complete generic workflow for tract-tracing experiments designed to (a) describe the main elements of the experimental design so that the relationships between the independent and dependent variables may be correctly tabulated and (b) accurately capture the process of obtaining data by a series of transformations applied to the object under study.

3.2 Reasoning over tract-tracing data

We perform our reasoning using the PowerLoom [4] first-order logic knowledge representation and reasoning system. PowerLoom provides us with a deductive reasoning engine that supports numerical calculations, n-ary relations and closed-world reasoning. It has a query language that allows us to access the information from our encoding of the experimental structures.

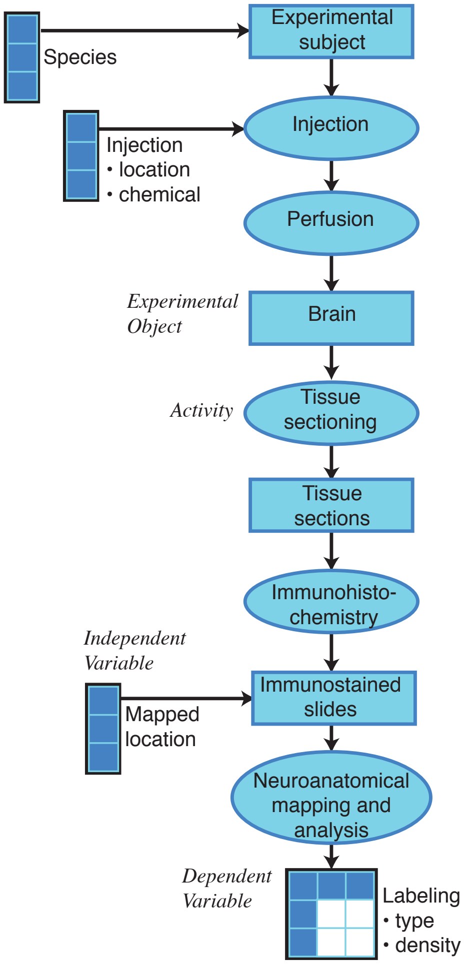

Figure 1: Generic experimental design schema for a tract-tracing experiment. This diagram shows steps involved and allows us to derive the data dependencies between the variables.

One example is the search for evidence concerning the existence and strength of connections between brain structures. We may look for *anterograde* (forward-tracing) experiments between a region of interest and other brain regions where signs of the tracer are found. We also use the reasoning system is to provide support for reasoning about geometric relationships between different brain regions. We have tools that import the basic geometric relationships from the brain atlases. We translate an XML-encoded version of the Swanson brain atlas containment information [2] provided by the Brain Architecture Management System into PowerLoom. We use the RCC5 variant of the Region Connection Calculus [5] to encode the topology of brain regions.

4. ACKNOWLEDGMENTS

This research is funded by the U.S. National Institutes of Health under grant MH-083871 for the 'BioScholar' project `http://bmkeg.isi.edu/`.

5. REFERENCES

[1] T. Blackstad, L. Heimer, and E. Mugaini. General approaches and laboratory procedures. In L. Heimer and M. Robads, editors, *Neuroanatomical tract tracing techniques*. Plenum Press, New York and London, 1981.

[2] M. Bota and L. Swanson. The brain architecture management system. Web site `http://brancusi.usc.edu/bkms/`, 2009.

[3] G. Burns, E. Hovy, and T. Ingulfsen. Biomedical knowledge engineering approaches driven by processing the primary experimental literature. In *Frontiers in Neuroinformatics. Conference Abstract: Neuroinformatics 2008*, Stockholm, Sweden, 2008.

[4] PowerLoom. Powerloom knowledge representation & reasoning system. Web site `http://www.isi.edu/isd/LOOM/PowerLoom/`, 2009.

[5] D. A. Randell, Z. Cui, and A. Cohn. A spatial logic based on regions and connection. In B. Nebel, C. Rich, and W. Swartout, editors, *KR'92. Principles of Knowledge Representation and Reasoning: Proceedings of the Third International Conference*, pages 165–176, San Mateo, California, 1992. Morgan Kaufmann.

[6] L. W. Swanson. *Brain Architecture, understanding the basic plan.* Oxford University Press, Oxford, 2003.