

TICK: A Content Management System Framework for Semantic Web Research and Instruction

Robert W. McGrail and S. Rebecca Thomas

Bard College Laboratory for Intelligent Data Systems
P.O. Box 5000, Bard College, Annandale-on-Hudson, NY 12504
mcgrail@bard.edu, thomas@bard.edu

Abstract

The goal of the TICK data management system is to make online publication of the fruits of research as easy as the publication of opinions on weblogs. In order to avoid the reinvention of many useful features, TICK leverages an existing content management system. This system already provides a framework for undergraduate research projects at Bard College. It is the authors' position that the TICK system also presents an appropriate platform for integrating Semantic Web instruction into undergraduate coursework. This article provides an overview of TICK along with a plan for integrating TICK into our computer science program.

Introduction

The TICK (**TICK Is Collaborative Knowledge**) data management system is a project of the Laboratory for Intelligent Data Systems at Bard College. The goal of TICK is to harness Semantic Web technology (Hendler, 2005) in order to ease the online publication of scientific research data and results, assisting scientific researchers in the same way that blogging software assists amateur political pundits.

The genesis of the TICK system lay in a senior research project that led to collaboration between the Ecology of Lyme Disease Laboratory at the Institute of Ecosystems Studies and Bard computer science faculty in 2001 (Shaw et al, 2003). Through this interdisciplinary work, the authors were exposed to a very substantial corpus of research data in need of major reorganization. The period since that time has witnessed the formation of the field of ecological informatics, which includes the application of Semantic Web technologies to the biological sciences (Finin & Sachs, 2004). Notable developments in this area include the SPIRE meta-project (Sachs et al, 2006).

Nonetheless, a wide gap persists between such advances in Semantic Web technology and the computational tools readily accessible to the average researcher. The authors believe that the time has come for a lightweight, user-friendly system that facilitates the integration of research

data into the Semantic Web. TICK is a step in this direction. The authors propose a semantically sophisticated data manager built upon an existing content management system (CMS). TICK will serve as a test bed for computer science courses as well as student and faculty research projects.

TICK Architecture

The OpenCms content management system (Butcher, 2006) forms the basis of TICK. It is based on Java (Gosling & McGilton, 1996) and employs Java Server Pages (JSP) as its template language. OpenCms supports code reuse through modules, which bundle templates, custom XML formats, and configuration files. An OpenCms microsite is a module defining a site of low to moderate complexity.

TICK Modules

TICK will consist of several OpenCms templates supporting a collection of new microsite modules. The Faculty and Research Laboratory microsites will house research data. The Seminar microsite is included in order to attract more research-minded academic users to TICK.

Seminar. An instance of this microsite is intended to serve as the online home for an academic seminar, colloquium, or other similar event. It employs XML for configuration details and basic seminar information. A seminar also utilizes a custom XML format for the announcement of public presentations. Presentations are syndicated via RSS (Winer, 2005). The seminar microsite will serve as a simple, complete model microsite for newcomers to TICK and OpenCms.

Faculty. This microsite represents the online professional home of a faculty member. It has a rich array of XML formats, most of which pertain to the various types of academic publications: journal articles, conference proceedings, books, technical reports, etc. Publications will have several RSS feeds corresponding to, say, lists of recent or selected publications. Other XML formats include news announcements and student worker position descriptions. A faculty member's website will be

automatically subscribed to the RSS feeds associated to their Seminar and Research Laboratory sites.

Research Laboratory. The Research Laboratory microsite is a collection of OpenCms elements that support the dissemination of academic research. XML formats include general news items and conference announcements, as well as the full range of academic publication types. Adjunct to this module is the research project microsite. A new research project is created from within an existing research laboratory site and is attached to the lab site's navigation tree.

TICK and Computer Science Courses

This section describes some points of intersection between the TICK project and the Bard College computer science program. The authors would like to stress that these activities do not collectively constitute any structural adjustment to our computer science curriculum. Indeed, each course and research project described below fulfills some mandate within our current program.

Introduction to Computing: Semantic Web. This course is a primer on XML technologies for a general audience. Students complete a series of laboratory and homework exercises that prepare them to build blog-style community websites, which will be implemented as instances of OpenCms microsite modules. Participants will first develop some simple static HTML web pages. Next they introduce new XML formats using XML Schema, and then build XSLT specifications to translate their XML profile data into HTML. JavaServer Pages are then introduced to add dynamic functionality. This strongly motivates the use of Java. The course then proceeds as an introduction to basic Java programming.

Databases. Recent trends in undergraduate database textbooks reveal dwindling interest in the object-oriented data model coupled with increased interest in semi-structured data (Ullman & Widom, 2002). In this spirit, the next iteration of the databases course at Bard College will place equal emphasis on the relational, entity-relationship, and semi-structured data models. New early laboratory exercises will focus on the design of XML formats for the TICK project.

Topics in Advanced Artificial Intelligence. The next version of this course will concern the use and development of ontologies. Since the audience for the course is advanced students, it is perfectly appropriate to dedicate a significant portion of the course to research problems arising from TICK.

Undergraduate Research. Two of our current seniors are conducting senior research projects under the TICK umbrella. One student is developing a semi-automatic system for generating custom XML formats from direct user interaction. The other is applying clustering algorithms to the categorization of research documents.

Conclusions

The status of the TICK system as of February 2007 is as follows. There is a working prototype of the Seminar module. Several templates and XML formats have been implemented for the Faculty and Laboratory modules. The introductory Semantic Web course will provide impetus for further development of these modules.

It is imperative that the project team balance the utility of TICK with its correctness as well as its user-friendliness relative to its target user base. We must be careful about the early choice of research domains to serve. We will focus on knowledge management in finite universal algebras (Burris & Sankappanavar, 1981). Many researchers within this area are quite computationally savvy, owing to their liberal use of computer algebra software. Also, the complexity of research data within this field (operation tables) is very manageable compared to those of the natural sciences.

References

- Burris, S. & H.P. Sankappanavar, 1981. *A Course in Universal Algebra*. New York: Springer. 1981.
- Butcher, M. 2006. *Managing and Customizing OpenCms* 6. Birmingham, UK: Packt Publishing Ltd. June 2006.
- Finin, T. & Sachs, J., 2004. *Will the Semantic Web Change Science?* University of Maryland, Baltimore County Technical Report. Sept. 2004.
- Gosling, J. & McGilton, H. 1996. *The Java Language Environment*. Retrieved Nov. 12, 2006 from <http://java.sun.com/docs/white/langenv/index.html>
- Hendler, J. 2005. Knowledge is power: A view from the Semantic Web. *AI Magazine*. 26(4): 76-84. Dec. 2005.
- Jackson, P., 1999. *Introduction to Expert Systems*. Essex, UK: Addison Wesley Longman. 1999.
- Sachs, J., Parr, C., Parafitnyk, A., Pan, R., Han, L., Ding, L., and Finin, T., 2006. Using the Semantic Web to Support Ecoinformatics. *Proceedings of the AAAI 2006 Fall Symposium on Semantic Web for Collaborative Knowledge Acquisition*. Arlington, VA. Oct. 2006.
- Shaw, M.T., Keesing, F., McGrail, R., and Ostfeld, R.S., 2003. Factors Influencing the Distribution of Larval Blacklegged Tick on Rodent Hosts. *American Journal of Tropical Medicine and Hygiene*. 68(4): 447-452. April 2003.
- Ullman, J. & Widom, J., 2002. *A First Course in Database Systems, 2nd Edition*. Upper Saddle River, NJ: Prentice Hall. Oct. 2002.
- Winer, D., 2005. *The RSS 2.0 Specification*. Retrieved on Nov. 19, 2006 from <http://blogs.law.harvard.edu/tech/rss>