Chapter XIII

Unified Citation Management and Visualization Using Open Standards: The Open Citation System

Mark Ginsburg, University of Arizona, USA

ABSTRACT

Scientific research is hindered when there are artificial barriers preventing the efficient and straightforward sharing of bibliographic information. In today’s computing world, the barriers take the form of incompatible bibliographic formats and constraining operating-system and vendor dependencies. These incompatible platforms isolate the respective camps. In this chapter, we demonstrate and discuss a new approach to unify citation management: the Open Citation System (OCS). OCS uses open XML standards and Java-component technologies. By providing converter tools to migrate citations to a centralized hub in BiblioML format (an XML tag set based on the UniMARC standard), we then make use of XML topic maps to provide an extensible framework for visualization. We take as an example the ACM classification code and show how the OCS system displays citations in a convenient focus and context hyperbolic tree interface. We conclude by discussing future directions planned to extend the OCS system and how open citation management can supply an important piece in our inexorable march toward a worldwide digital library.
INTRODUCTION

The Internet infrastructure enables many large- and small-scale citation databases that researchers enjoy today. Some of the citations are directly available from online digital libraries (DLs). Others are parsed from free-standing Web documents and placed into an online DL as the CiteSeer project accomplishes with its autonomous citation indexing (Lawrence & Giles, 1999a). The DLs, coupled with traditional journals, conference proceedings, and stand-alone citation collections, place voluminous citations online.

It would seem we are well on the way to the goal, proposed by Cameron (1995), of having a “universal, bibliographic and citation database linking every scholarly work ever written.” Yet, researchers struggle with several incompatible “spoke” formats based partially on historical reasons separating the physical and social sciences. Unified citation management is the goal, but at present, it is a fractured endeavor with ad hoc bridging attempts littering the landscape but no cohesive themes or design principles underlying the efforts.

What should a unified citation-management system accomplish? At the very least, it should provide a convenient interface to convert citations to and from various spokes: a set of spoke-to-spoke converters. It should also be able to group citations into a database to support fielded queries in order to retrieve specific subsets for the task at hand. Later, we will review some of the software systems that accomplish these basic goals. Thus, researchers can move back and forth between formats using simple Web forms and construct simple fielded queries on databases. The systems generally input and output serial lists of citations in the requested format. It is difficult to know which topics a given set of citations covers, though. Researchers must do fielded queries, for example, by author, title, abstract, or keyword to retrieve citation subsets. Ad hoc visualization efforts local to various spoke formats, while possible, have not been widely adopted in practice and do not address the fundamental problem of the citation information islands.

This chapter seeks to approach the problem of citation-topic exploration with more design cohesion and stability by using an integrated set of open XML (extensible markup language) standards and Java components. As Krechmer (2002) points out, open standards are “a codification that a society (collection of users and implementers) wishes to maintain unchanged for a span of time,” which directly yields the desired stability. Our citation-management system prototype, the Open Citation System (OCS), is introduced later in the chapter. The OCS implements converter classes, written in Java and implemented as a Web-based servlet, to transform spoke formats into a special XML tag set: the bibliographic markup language (BiblioML; Cover, 2001) “hub.” After discussing the advantages of the choice of XML as the hub format, we present additional XML tools, such as XML topic maps (XTMs; Auillans, 2002; ISO/IEC 13250 topic maps, 2000), that work in conjunction with a Java servlet and a hyperbolic...