

PubsOnline: Open Source Bibliography Database

Scott A. Myron

Indiana University
2711 E 10th St

Bloomington, IN 47408

+01 (812) 855-2627

samyron@indiana.edu

Richard Knepper

Indiana University
2711 E 10th St

Bloomington, IN 47408

+01 (812) 855-9574

rkneppe@indiana.edu

Matt Link

Indiana University
2711 E 10th St

Bloomington, IN 47408

+01 (812) 855-6339

mrlink@indiana.edu

Craig Stewart

Indiana University
Franklin Hall

Bloomington, IN 47408

+01 (812) 855-4240

stewart@iu.edu

ABSTRACT

Universities and colleges, departments within universities and colleges, and individual researchers often desire the ability to provide online listings, via the Web, of citations to publications and other forms of information dissemination. Cataloging citations to publications or other forms of information dissemination by a particular organization facilitates access to the information, its use, and citation in subsequent publications. Listing, searching, and indexing of citations is further improved when citations can be searched on by additional key information, such as by grant, university resource, or research lab.

This paper describes PubsOnline, an open source tool for management and presentation of databases of citations via the Web. Citations with bibliographic information are kept in the database and associated with attributes that are grouped by category and usable as search keys. Citations may optionally be linked to files containing an entire article. PubsOnline was developed with PHP and MySQL, and may be downloaded from <http://pubsonline.indiana.edu/>.

Categories and Subject Descriptors

H.3.5 [Information Systems]: Information Storage and Retrieval – Web-based services

General Terms

Management, Documentation, Standardization.

Keywords

Publications, citation management, record-keeping, grant writing, bibliographies, BibTeX, open source

1. INTRODUCTION

Tracking of intellectual contributions to scientific progress is often of value to universities and colleges. An entire university or college, a school or department, or an individual researcher may want to provide information about their publications and presentations. Years ago, universities or departments within universities often provided such information by binding a year's worth of reprints together in hardback book form; the weightiness of such books was taken to be an indication of the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SIGUCCS '05, November 6-9, 2005, Monterey, California, USA.

Copyright 2005 ACM 1-59593-173-3/05/0011...\$5.00.

heft of intellectual accomplishment. Today, the Web is a critical means of delivering information to potential users of research results. The ability to provide information via the Web may increase the number of people who read or cite a research paper. In addition, the tracking of publications, talks, press releases, and other forms of dissemination of scientific research is now commonly a legal requirement of grants. The ability to deliver this sort of information via the Web can make compliance with such grant requirements easier for grantees, and may make access to such information simpler for granting organizations.

Indiana University began tracking publications and talks using a Web-accessible online database in 2004, so that publications supported by grants involving University Information Technology Services could be easily searched, sorted, and accessed via the Web [7]. This has proven to be a valuable aid in providing information about Indiana University's intellectual accomplishments and giving indications of the impact of particular grants. The software that has made this possible—called PubsOnline—has now been released as open source. The purpose of this paper is to describe PubsOnline and its features, so that any institution may put it to use.

2. MOTIVATION

The primary motivation for the creation of PubsOnline was to provide for straightforward presentation of information about publications and presentations stemming from several grants to Indiana University. Such information had earlier been managed by hand using standard HTML pages with no search capability. Presenting citation information in this way became unmanageable once hundreds of such citations had been accumulated, associated with one or more of several grants received by University Information Technology Services of Indiana University.

Combining citations into a central management system allows for easier management of information. Search and retrieval functions make it easier for users of this information to find information pertinent to a particular topic. PubsOnline provides these capabilities, and can also keep a copy of publications in the database for download.

FEATURES

3.1 Basic Features

PubsOnline provides several important features for managing citation and bibliographic information. First, it provides the administrator with the ability to design the ways in which citations are organized and specify attributes to associate with each citation, to facilitate searches of the bibliographic information for particular pre-defined purposes. It also provides

a straightforward Web-based form to enable authorized individuals to add new entries to the database. Most importantly, PubsOnline facilitates the retrieval of information via the Web, with no special software requirements (other than access to a standard Web browser) placed upon the individuals accessing information. People who retrieve information see by default a standard HTML page with citation information (and links to copies of files, if provided within PubsOnline by the information managers), and may optionally obtain a text file of the citations in BibTeX format [2].

The ability to provide customized attributes is vital. Citation attributes may be any additional information that is useful in organizing citation information. For example, citations might be associated with their supporting laboratory, department, or school in order to facilitate organization and sorting. It may be desirable to identify all publications associated with a particular grant, or all publications that made use of a particular computer system. With such customized attributes for sorting publications, it is easy to generate lists of citations organized by characteristics important to presenters and users of the information held in a PubsOnline database. PubsOnline may be used to present citation information alone, or citation information and links to copies of documents stored on the server that runs PubsOnline.

There are three categories of individuals who may access information within a PubsOnline database: the information manager, who specifies attributes and has all rights to the underlying software on which PubsOnline depends; individuals who may add entries to the database but not change the database organization itself; and individuals who access information stored in a PubsOnline database, who may only read – but not add or edit – information in the database. PubsOnline provides for authentication for individuals authorized to manage the database or add entries to it. PubsOnline does not, itself, provide for capabilities to limit access by potential consumers (readers) of information, but authenticated or limited access to read information from a PubsOnline database may be provided by any means usable to control access to any HTML page.

PubsOnline provides great flexibility in the management and presentation of citation information. Each citation is associated with one or more attributes specified by the manager(s) of the information in the database. Each attribute is a member of a category, which is used to group attributes together. For example, in the case of Indiana University, attribute categories are used to organize publications by the research computing systems used in a particular research effort and the grant(s) which supported that effort. Figure 1 shows the PubsOnline database implemented by the Research and Academic Computing Division of University Information Technology Services at Indiana University. A paper published by an Indiana University researcher might have depended upon use of Indiana University's IBM SP. The citation for this paper would have the attribute "IBM SP" in the "Systems" category. Similarly, a paper may have been supported in part by the Indiana Genomics Initiative grant to Indiana University. In this case, the paper would have the attribute "Indiana Genomics Initiative" in the "Grants" category. When added to the database, these attributes are associated with the citation information in the database. PubsOnline, then, provides a great deal of flexibility and sophisticated capabilities to the presenters of information as to how information is presented and organized.

PubsOnline provides good tools for input and management of citation information. Entries may be added to a PubsOnline database via a Web-based form. The person entering information initiates the process by logging in to the input form. Adding a particular field is initiated by selecting the BibTeX type for the citation and filling in required and optional metadata for the publication. The individual adding entries may also indicate a file to be uploaded and linked from the citation. These files may be stored in any format. PubsOnline stores the uploaded papers in a directory and verifies that each has a unique name. PubsOnline will indicate a URL field in the citation to the indicated file to enable download of the file (if a person tried to add a file name that is already in use within an individual PubsOnline instance, an integer is added to name of the new file being added to eliminate the conflict). The person entering information then assigns attributes to each citation, per the attributes established within a particular PubsOnline instance.

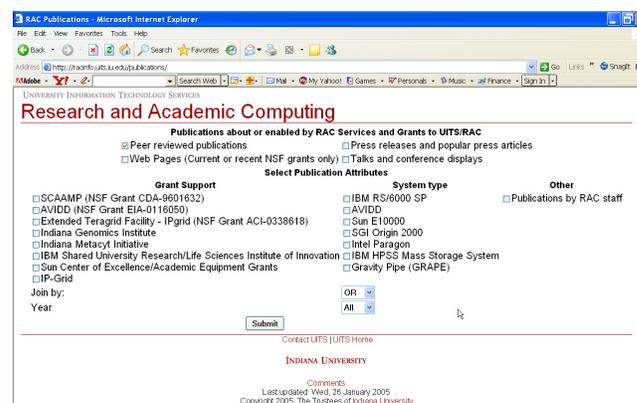


Figure 1. Online database search screen implemented by the Research and Academic Computing Division of University Information Technology Services to organize citations for publications, presentations, Web pages, and press releases [7]. The checkboxes indicate attributes that may be selected within three categories specified within PubsOnline (systems, grants, and whether or not a staff member of the Research and Academic Computing Division is among the authors).

Most importantly, PubsOnline provides significant flexibility and capabilities for the users of information contained in a PubsOnline database. PubsOnline presents, by default, a list of checkboxes that can be selected based on the attributes specified by the information administrators (Figure 1). A person retrieving information from a PubsOnline database can select all citations from the database, or a subset of citations based on criteria associated with particular attributes. A user may specify multiple attributes in selection of citations, and may use AND or OR logical joins to utilize multiple attributes at once. A person retrieving information from a PubsOnline database may also specify a keyword to be matched in the titles of the publications in the database, providing a great deal of flexibility to select information in ways not already made automatic through the information attributes. Results are presented to the user sorted by year and by BibTeX type. PubsOnline supports all of the BibTeX types, including articles, books, Master's Theses, Ph.D. Dissertations, and fields that may be extended to include other categories such as seminar presentations and press releases. BibTeX was chosen because it is a standard format in computer

science for providing bibliographic references, and because other bibliographic software commonly supports import and export of this format.

3.2 Installation and Requirements

PubsOnline may be downloaded from <http://pubsonline.indiana.edu> [6]. PubsOnline requires a Web server with PHP [4], such as Apache [1] with mod_php installed. The template function of PubsOnline requires that the Smarty template engine [5] for PHP needs to be installed. PubsOnline also requires a MySQL[3] database to store publication records. Currently, the PubsOnline database must be created manually; later versions of PubsOnline will allow users to create the database with a PHP form. In order to install the software, an administrator unpacks the PHP files into a directory that is made available to the Web server. Once the PHP files are unpacked and the MySQL databases have been created, the PubsOnline software is ready for use. PubsOnline is available for download as package of PHP scripts and documentation.

3. DEVELOPMENT

4.1 Basic Development and Template Engine

PubsOnline was developed using PHP, MySQL and Smarty. PHP was chosen based on its simplicity and large standard library; MySQL was used due to its proven interaction with PHP; and Smarty was used as a template system to allow for customization of the presentation layer without disturbing the logic that runs PubsOnline. This will enable any organization to use PubsOnline and integrate it seamlessly into their website.

4.2 BibTeX Database Schema

The database schema used by PubsOnline is designed around and mimics BibTeX. Each type has a set of required and optional fields. Like BibTeX, PubsOnline requires that certain fields are specified for each type. For example, the ‘article’ type has 4 required fields: authors, title, journal, and year. There are times, however, when PubsOnline is stricter than BibTeX. For example, in a ‘book’ type, PubsOnline requires that an author is specified even though BibTeX allows for an author or an editor to be listed. Furthermore, PubsOnline allows the administrator to extend these types as necessary to fit their application. For example, PubsOnline has extended the ‘misc’ BibTeX type to include a ‘Seminar/Talk’ publication type. This type is exactly like the ‘misc’ type; however the ‘note’ field is already set to ‘Seminar/Talk’.

In the current version of PubsOnline, the user has the option to export the search results to a BibTeX document. It should be noted, however, that this feature is for now highly experimental. More testing is necessary in order to determine if the output produced by the exporter is valid BibTeX in all cases. It is worth mentioning that special types will not be preserved during the export. For example, an entry that has the type ‘Seminar/Talk’ is actually exported as a ‘misc’ entry. Importing from a BibTeX document is not currently implemented and is in the design process.

4.3 Future Enhancements

PubsOnline has a number of milestones for future development. The import from BibTeX feature needs to be finished. This will allow organizations to quickly insert data into the PubsOnline database if they already have BibTeX indexes of their

publications. Second, the exporter needs to be thoroughly tested to verify that it produces valid BibTeX documents. Another feature that will be enhanced is the publication formatter. This would fix some of the minor citation errors that are possible (for example, a double period at the end of the title) by removing errors or prompting the user to fix them.

4. CONCLUSION

PubsOnline is a powerful open source tool that allows an organization to track and present information about its intellectual accomplishments. Tracking of data important to an organization is facilitated when information beyond basic bibliographic information is included in the publications database. At Indiana University, PubsOnline has facilitated grant writing by permitting grant writers to easily search through publications organized by relevant information. PubsOnline has also allowed administrators to see which grants and research systems have been associated with the largest amount of research. PubsOnline not only allows organizations to efficiently and effectively manage publications and track the use of grant resources, but it also allows users to search the database to see what the organization has accomplished. Perhaps most importantly, by making it easier to provide citation information (and optional links to copies of publications), PubsOnline may facilitate increased use of the information and innovation so painstakingly developed and presented in publications by researchers – thus increasing the impact and value of such research.

5. ACKNOWLEDGMENTS

The Unix Systems Support Group at Indiana University provided hardware for development and distribution for this project. D. Hart, M. Lingwall, B. Minton, R. Repasky, J. Rinkovsky, C. Shields, and E. Wernert provided feedback on at least one of the following: software, abstract, and manuscript. This publication was supported in part by the following: National Science Foundation under Grant No. 0116050; grant number 1U24AA014818-01 from NIAAA/NIH (the contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the NIAAA/NIH); the Indiana Genomics Initiative; and the Indiana METACyt Initiative. The Indiana Genomics Initiative and the Indiana METACyt Initiative are supported in part by the Lilly Endowment, Inc.

REFERENCES

- [1] The Apache Software Foundation. Apache HTTP Server Project. <http://httpd.apache.org/> (visited July 5, 2005).
- [2] D. Jacobsen. The BibTeX Format. <http://www.ecst.csuchico.edu/~jacobsd/bib/formats/bibtex.html> (visited July 7, 2005).
- [3] MySQL A.B. MySQL Developer Zone. <http://www.mysql.org/> (visited July 6, 2005).
- [4] PHP Group. PHP Hypertext Processor. <http://www.php.net/> (visited July 5, 2005).
- [5] PHP Group. Smarty Template Engine. <http://smarty.php.net/> (visited July 5, 2005).
- [6] Trustees of Indiana University. PubsOnline. <http://pubsonline.indiana.edu/> (visited July 9, 2005).
- [7] Trustees of Indiana University. PubsOnline. <http://racinfo.indiana.edu/publications/> (visited July 5, 2005).