

THE CREATION OF A CUMULATIVE INDEX ON CAD: "CUMINCAD"

Bob Martens[a], Ziga Turk[b]

Abstract

To researchers in many disciplines, Internet is quickly becoming the dominating environment to search for publications. Commercial bibliographic databases tend to be too general, are not up-to-date and require special skills and efforts to be searched. On the other hand research is also published on the Web that also enables collaborative creation of references by the specialists in the field. CUMINCAD is such a bibliographic database and compiles papers related to Computer Aided (Architectural) Design. The database is available on the Web and allows searching and browsing in the ways usual on the Web. It provides a "historical evolution" to learn from previous efforts and draws attention to older original works that could have been ignored because they could not be found on the Web. We believe that CUMINCAD will help focus future CAAD research and improve the education.

Keywords: CAAD-related Publications; Web-based Bibliographic Database; Conference Proceedings; Searchable Index; Grey Literature

Introduction

Since 1993, when architects and engineers started to use the Web [1], Internet has changed the ways in which people search for information. We use it to get the latest news, stock quotes, travel information, software documentation, reviews of books and films, etc. We are used to relying on search engines like Yahoo or AltaVista to find any information and get it for free. Web browsers have established themselves as the main tool to access the information and the paradigm of Web browsing and hypertext technology is what most people are getting used to. Using the Web is by far the most convenient. Quite often, if something cannot be found on the Web and for free, people are less and less inclined to bother getting it. They look for a similar publication available for free on the Web.

Exposure to the Web has also changed the ways in which scientists are doing literature studies and retrieve copies of papers. In the past several information services have been established to assist them, however, most of them are commercial, not available on Internet, or both (i.e. ICONDA, RWSB, URBADISC). Each uses a proprietary search technique, requires a special software and may only be available in the local library where one is supposed to physically walk to. They are not updated as promptly as new contents is being authored.

Getting full texts is even more awkward - paper copies are ordered and may take weeks to arrive. The Web also changed the ways in which information is being published [2]. It is considered the most democratic medium where the ratio between the number of authors and readers is much higher than with any other medium. Scientists and researchers have started to use it immediately to collaborate and share their ideas and research results with their colleges. Several con-

ferences and journals, including Automation in Construction, make their content available on the Web. Some journals are even purely electronic and available for free on the Web. This has been the case for the past few years but many of the older work is not available on the Web and therefore not immediately available to the new generation of researchers who use Web search engines to find information - any information.

For example, a number of conferences on CAD have taken place ever since the early eighties. The question of effectiveness of accessibility to this knowledge, particularly of the individual contributions has been raised. Recently, more and more conference proceedings have been labelled with an ISBN-number, thus facilitating ordering. Individual proceedings have doubtlessly been collected in libraries, however, a complete archive recording of complete series of all proceedings is rather rare.

The project goals have been identified as follows:

- Literature indexes should be available on the Web.
 - General search engines of the Web are not good enough to find publications within a specialized field such as CAAD.
 - Commercial databases are expensive and are not as easy to use as the Web, they are not updated as quickly as possible and may not include relevant papers and reports published internally or on highly specialized conferences and workshops.
 - Papers published in the pre-Web age have the disadvantage of not being equally easy to find as the latest.
- To tackle these issues the following solution is proposed:
- Offer a free index of major publications related to CAAD on the Web.
 - Use it to establish a critical mass that would attract researchers to register their work themselves.
 - Distribute the effort of maintenance and data collection while maintaining a minimal quality control.
 - Do it on a low budget.

An interface designed in this way has been created. Technical features will be presented in Section 2 and some issues related to data entry and context in Section 3.

CUMINCAD: The Design of the Interface

CUMINCAD stands for "Cumulative Index on CAD". At the time of writing it contains already over 2000 bibliographic entries. It is a collaborative effort of two persons, the authors, who have never physically met so far, but collaborated on this project using the Internet.

Data Structure

Each entry includes (1) typical bibliographic data, (2) extended bibliographic data, (3) Internet links and (4) administrative data. The bibliographic section includes information on authors, title, where and when it was published. This information is not structured into very small chunks as in most bibliographic libraries but broken into the four fields only. This facilitates the input that can be done quickly by simply using cut and paste of fairly large chunks of data. Rather smart full text searching techniques compensate for

the lack of structure. Extended bibliographic data includes summary of the publication and the keywords (if available). The keyword set is not controlled, which is appropriate for a fast moving field such as CAAD. Internet related fields include possible links to full text of the publication and to author's coordinates. In case these are not defined explicitly, search into the Web can be launched and there are quite good chances of finding the full text or the authors. Administrative information is about when an entry was inserted or modified, by whom, etc.

User's Interface

CUMINCAD-homepage [http://www.fagg.uni-lj.si/cumincad/] that offers access to most popular functions. Several search and browse options are available. When searching, a user enters a few words he is interested in, for example "Smith visualization". This would return papers written by a Smith and those dealing with visualization, the ones about visualization and Smith listed first. Each entry can be opened and full details about a publication are displayed. As demonstrated above, a user can search quite efficiently without knowing any special syntax to define the searches. For more complex searches, the user can use search syntax of the perhaps most popular Internet search engine - AltaVista. For example "Smith Jones +visualization -series: ACADIA" would return only records containing word visualization that were not part of the ACADIA series. Those by Smith or Jones would be listed on top.

Searching is important and useful, however, users are not always sure what they are looking for. CUMINCAD offers three browsing options - by author, by words in titles and by words in keywords. Indexes are automatically generated and therefore include some garbage, but they are nevertheless useful. Clicking on a word in index starts a search for that word and eventually ends in a listing.

We expect users to come back from time to time to see what is new. The interface uses some Web tricks to maintain what a user has seen last and he can therefore quickly find the changes in the database since his last visit. Another advanced feature (currently disabled for the general public), is to order a built in agent to keep one up-to-date with new papers about a certain topic. The interface is quite simple: at the end of a listing, a user is prompted if he wished an agent (a blonde called Angie) to do such searches for him. He enters his e-mail address and when (for example every Monday) he would like to receive the updates, and that is all. Search results will be delivered by e-mail, every Monday.

Editor's Interface

CUMINCAD uses a Web-oriented database. This means that not only the end users' literature research, but the editorial work can be done using the Web. Costly connection time is avoided and the principle of mobility comes in useful, i.e. input may be carried out from any computer equipped with Web browser, a procedure which surely was inconceivable until recently. The input is not as friendly as it could have been if programmed e.g. on a PC under Windows. The response times are a second or two, instead of a fraction of it. Still, the advantage of mobility and distribution outweighs the disadvantages. Many people can contribute the data via

Web. Import from structured PC formats (i.e. Word tables, Excel) is also possible. The database can hold tens of thousands of records, however, the authors do not intend to sacrifice the quality and focus of the entries for sheer volume.

Currently only the authors are granted sufficient privileges to add records into CUMINCAD. Quality control has produced so far hardly no problems as the main focus was laid on the compilation of contributions from respective Conference Proceedings on CAD. Later, different CAD-Associations resp. the general public will be allowed to add as well, but an editor will be asked to confirm that the entry is suitable for CUMINCAD. Furthermore, it will then be necessary to define more properly criteria in order to secure quality control.

Acknowledgments

CUMINCAD has been created by the authors. The understanding regarding their effort, shown by their Universities, is appreciated. The hardware for the service has been made available by the Institute of Structural, Earthquake Engineering and Construction IT of the University in Ljubljana. Great interest in cooperation to this end by the different CAD-associations can be taken for granted.

References

- [1] Z. Turk, *A Network Based Collaboration Framework, eCAADe Conference Proceedings, Eindhoven, The Netherlands, 1993.*
- [2] B-C. Bjoerk, *The Electronic Journal Revolution to the Scientific Publishing Process, Swedish Building Research, pp. 8-10, 3/1996.*

[a] Faculty of Regional Planning and Architecture, Vienna University of Technology, A-1040 Vienna (bmartens@email.tuwien.ac.at)

[b] Faculty of Civil and Geodetic Engineering, University of Ljubljana, SI-1000 Ljubljana (ziga.turk@ikpir.fagg.uni-lj.si)

