Go to contents 01

Digital Proceedings: Experiences regarding Creating and Using

MARTENS, Bob; TURK, Ziga; CEROVSEK, Tomo
Vienna University of Technology (Austria) and University of Ljubljana (Slovenia)
http://itc.fgg.uni-lj.si/cumincad/

This paper describes the developments of the CUMINCAD database since 1999 when it was first presented and some statistical information, how the service is being used. CUMINCAD started as a bibliographic database storing meta information about CAAD-related publications. Recently, full texts are being added. The process of creation of electronic copies of papers in pdf-format is described as well as decisions which were taken in this context. Over the last two years 20.000 users visited CUMINCAD. We present a brief analysis of their behavior and interaction patterns. This and the forthcoming possibility of a full-text-search will open up a new perspective for CAAD-research.
Keywords: *CAAD-related Publications; Web-based Bibliographic Database; Searchable Index; Retrospective CAAD Research; Purpose Analysis*

Introduction

In the eighties, first considerations regarding the implementation of CAAD occurred at many architecture schools, soon resulting in a need for exchange of ideas and experiences. Platforms such as ACADIA, CAADRIA, eCAADe and CAAD Futures were established. Up to the second half of the nineties paper-based proceedings were published, generally with a small (50-500 copies) circulation only. What happens to smaller editions is that they are safely stored away in the studies of participants of the conferences and university members, but hardly become available to the public. Library networks have so far not been systematically supplied with copies, as far as could be traced back.

CUMINCAD, an acronym for “Cumulative Index on CAD”, is a bibliographic index that compiles papers related to computer aided architectural design (Martens and Turk, 1999 resp. 2000; http://itc.fgg.uni-lj.si/cumincad/). This work started in 1998 and helps focussing on future CAAD education and research activities. Implemented with a database, it 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 otherwise. CUMINCAD thus supports the search and the dissemination of CAAD-related publications.

As of today, all available CAAD-conference proceedings are recorded in CUMINCAD and apart from general bibliographical data also a summary is presented. However, if a record is of interest, it might be rather difficult to retrieve the full-paper. I.e.: the already mentioned problems regarding “gray literature” results in the fact that procuring e.g. via the international inter-library loan is principally possible, but very time-consuming and rather uncertain as to its outcome. Studies are showing (Bjoerk and Turk, 2000) that particularly the younger researchers do not wish to bother with using any literature but the one that is conformably and for free available to them through the Internet.

Perhaps one succeeds in contacting the specific author directly via the registered e-mail address, the possibility of offering the digital or digitalized full papers, however, surely is to be regarded the required completion of CUMINICAD. A further aspect in favor of digitalizing is that a complete collection of Proceedings would become "transportable" by means of the CD-ROM medium granting a clear overview of the scientific conference performance of an association. At the time of writing more than 3.000 records are available and approx. a third of them were imported from the CADLine-Database (Kalay, 1991).

Digital eCAADe Proceedings (1983-2000)

Until the 1996 Annual Conference, paper-based eCAADe proceedings were in a relatively small number published. Starting with the 1997 conference, every year also a CD-Rom was produced, which made this kind of information indeed widely available. There are hardly any individual teachers, researchers, librarians, etc., which avail of a complete set of eCAADe proceedings. The small number of printed proceedings as well as the policy of changing conference hosts lead to a situation that an ordering afterwards is quite uncertain in terms of having success. However, eCAADe holds the copyright on those publications and in fact this set of publications is worth an important "capital" of the association. Making nearly two decades of teaching and research work in this specialized field of CAAD available would therefore be of great importance (around 3.600 pages). For this reason the eCAADe-council decided to finance this project by own funding.

This decision results from the offer of the Viennese company Mediatecture (http://www.mediatecture.at). The practical experience of this enterprise in this area, - the 1991- Proceedings serving as a test set in this framework with approx. 300 pages – was particularly convincing.

In order to keep the documents efficiently small for the internet resp. providing a full-text search, not only just a scanning job - page by page -, was performed. Furthermore the intention to present the original layout, tracing back to the original proceedings, was focussed on. The main item of performance was the conversion to pdf-files. To the general work belonged OCR-scanning and conversion to MSWord; furthermore editing and eliminating scanning mistakes resp. finally creating Adobe-pdf-files. The abbreviation "pdf" stands for "portable document format" which can be implemented on practically all current platforms (http://www.adobe.com). Pdf-files have become popular and aim mainly at producing a printing format in its original form on any available computer by means of a reader (free of charge) or a plug-in, which means that complex software-installations become obsolete and using is available without any "ifs and buts".

Text recording and conversion in MS Word-documents was achieved by means of Omnipage (http://www.caere.com). Using the original fonts and type sizes so possible was attempted. Working with the produced Word-document soon showed that the handling of the given file size might become tricky due to the number of pictures. Thus dividing into several (part-) documents seemed meaningful. Moreover, a footer was inserted in order to indicate digital processing as such and subsequently entering page numbering wherever missing. The main work, however, was to eliminate mistakes resulting in the course of text recording. The submitted printing quality of the individual Proceedings played a decisive role in this context: i.e.: unsatisfactory printing performance leads to time-consuming corrections.(figure 1).

The Word-documents produced herewith still remain very memory-consuming in their corrected form. By means of pdf-conversion effective compriming was achieved. And again experimenting was necessary due to unintended occurring of pixels leading to illegibility.

On- and off-line access

The possibility of producing html-documents has so far not been documented. Wishing to be able to create the best-possible match of the original this variation

Go to contents 01

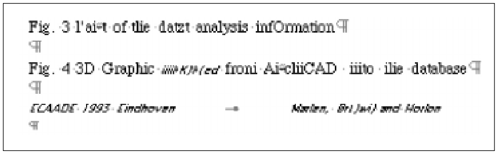


Fig. 3 1'ai=t of tlie datzt analysis infOrmation ¶
¶
Fig. 4 3D Graphic *aw.K.h/ed* froni Ai=cliiCAD iiito ilie database ¶
¶
ECAADE 1993 Eindhoven → Marlen, Bri.Jwi) and Horlow
¶

Figure 1. Example of an Extremely "Distorted" OCR-Document.

was not pursued, as html-presentations are not necessarily made for printing-out purposes (problems with page break and page references) and this is what the users are actually interested in. On the other hand the eCAADe 1997-Proceedings available exclusively in html-format were converted to pdf, not e.g. loosing any links specified. As soon as the complete "rough material" was available in the pdf-format the next step was to create two different final versions:

- a complete indexed file for offline use (CD-ROM)
- a collection of stand-alone files for online use (connected with CUMINCAD)

What has not been considered so far is that pdf-based publication provides the possibility of options and restrictions. If so required document printout can be prohibited, read-only rights being granted, whereas the addition of annotations or even text editing may be available provided "Adobe Distiller" (involving charges) is installed. Any restriction may be canceled by entering a password.

For the production of the CD-ROM-version the individual documents were compiled and indexed according to year of appearance. Definitive decision concerning the production of a hybrid CD-Rom (Mac/PC) have been issued by the eCAADe council and this product will be available in the course of spring 2001. It was also agreed to add the already digitally produced proceedings (1997-2000) to this CD-Rom.

Approx. 400 individual contributions from the rough material were separated by means of the "Distiller" for the offline-version. The Annual General Meeting of eCAADe-members in Weimar 2000 agreed to the proposal of the eCAADe-Council to have restricted access to those full pdf-papers, i.e. exclusively limited to eCAADe-members. CUMINCAD was adapted to meet those requirements, i.e. that the e-mail address is to be entered prior to transmission a pdf-contribution. Then a check is made, and provided the specific e-mail address is registered in the member database, the desired document is furnished as attachment.

Using CUMINCAD

A period of two years allows for a review of how such a service is used. The analysis is based on web server's log files from July 1999 till mid May 2001. From the log files such activities of both universities were excluded which involved data update, maintenance and development, as well as activities of Web robots making sure that CUMINCAD appears in the indexes such as AltaVista or Google. Except for two static html files all hits go to dynamic CGI-generated pages delivering the content from the database. Most information provided by off the shelf log analysis software was of limited or no use at all.

During the period of 22 months the CUMINCAD site served approx. 330.000 requests (500 per day), two thirds of which stemming from robots or from the two authoring institutions. The remaining 110.000 requests came from 20.000 different workstations in 30.000 different sessions (45 sessions per day). The majority of these people seem to be just passing by not exploring the site. 3900 visitors looked at more than 5 CUMINCAD pages and this is the number of users we can claim that CUMINCAD has. While keeping the service free, since late January 2001 it is required that users who want to see an abstract should register and provide his/hers email address. In the last 4 months 150 have registered - about 1.2 per day (figure 2).

Over time, CUMINCAD usage is stable as shown in the figure. Peaks came after big announcements. A rather small number of hits this year is probably a result of restrictions placed on CUMINCAD use. Since registration is required, looking at abstracts accounts for 22% of all requests, however only roughly half of them registered and thus able to see abstracts. An even greater restriction is placed on access to the full texts - users must be members of eCAADe; also only 10% of the entries is available as full text. Since

February 2001, visitors wanted to look at 1163 abstracts but succeeded to see 694. They wanted to download 70 different papers (from 350 available) in 134 requests, however, only 8 full texts were ultimately retrieved, as only eCAADe membership entitles to such. 60% of CUMINCAD users are from the US, whereas between 5 and 2% of the users come from countries such as United Kingdom, Germany, Italy, Netherlands, Canada, France, Australia, Spain, Belgium and Slovenia.

Figure 3 shows shares of the ten most frequently used types of requests. The most frequently used request is Search (42%) followed by viewing of records (Show and Go, 23%). 22.000 times, views of main home page (index.htm) 18.000 times and so forth. The next most popular command is BrowseAZ (7%) that enables user to browse by keywords (25%) or by authors (65%). This shows that when browsing, users are more interested in seeing which authors are involved. Initial studies of the click-through patterns show that browse by author is a good way to avoid unsuccessful searching (figure 3).

The top 50 search terms are cumincad (searched for 600 times), virtual, cad, test, virtual reality, architecture, case based resoning (sic!), sketch, presentation, EAEA, Aarnaes, a, Jakimowicz, internet, vr, autocad, amadeo, management, Gero, representation, philosophy, animation, ecaade, form, Oxman, interface, Heylighen, future, fractal, ahuja, construction, design, and, daylight, Hendricx, education, structural, ecadee Martens, engineering, environmental, educational, dynamic, analysis, caad, early stage, teaching, virtual design studio, Asanowicz (search for 90 times).

The most frequently opened papers were written by Aarnaes, Abadi, Abbott, Jung, Achten, Af Klercker, Gross, Anders and Martens. Of these papers, all but “Gross, Mark D (1999) Drawing, Seeing, and Reasoning: The Added Value of Computer Aided Architectural Design” and “Jung, Th., Do, E.Y. and Gross, M.D. (1999) Immersive Redlining and Annotation of 3D Design Models” are favored by the way CUMINCAD is organized. Being written by an author with a last name starting in “A”, promotes the paper to the top of search results listing. The Martens and Turk paper about CUMINCAD is referenced in the service’s home page.

By looking at those log files we learned that we should reconsider the restrictions on the end users; we should probably make sure that the search results are not sorted by last name but by some other criteria.

Further extension plans

Due to the current expansion three stages of availability will be provided: anonymous access will remain, only issuing access to bibliographic data,

Figure 2. Analysis of log files.

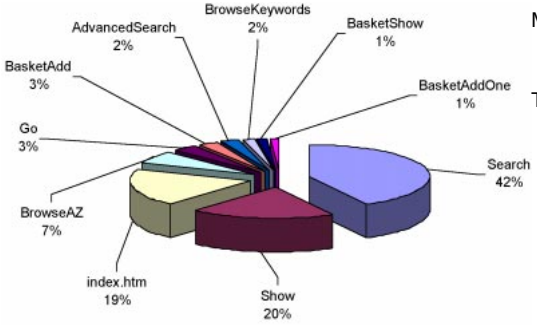
however. E-mail address entry makes the user a "friend" also having access to the summaries. At present full access is only available to eCAADe-members and grants access to pdf-documents. This grading makes for future use in a differentiated manner (e.g. hits-analysis).

Additionally, an exchange of ideas with the CAAD-associations (ACADIA, CAADRIA and SIGRADI) is to be encouraged possibly leading to a mutual access for members. ACADIA like eCAADe has a comprehensive collection of paper-based Proceedings. Subsequent procuring seems readily possible apart from some few exceptions (1998 resp. 1993-1999 still available). Anyhow, the (part) digitalization could seem meaningful. The situations of CAADRIA and SIGRADI might prove better as the present Proceedings probably are available in a digital format and thus mere simplified conversion or indexing work would be necessary.

Conclusions

This contribution issues an extensive description of the frame conditions for the digitalization of Conference Proceedings representing in its entirety a great asset for a CAAD-association. The wide availability of eCAADe-Proceedings in the period of 1983 to 2000 was accomplished by means of an offline-version (CD-Rom) and an online-version (CUMINCAD), financially absolutely affordable. Considering the working effort involved it could have been rather tempting only to scan-in the total package without text recording, then, however, the file-sizes in on-line use would have proved quite problematic and, furthermore, a full-text search would not have been possible. Regarding this aspect the digital Proceedings differ considerably from their paper-based originals. Doubtlessly, a digital publication can be designed according to specific appearance requirements from the very beginning provided the paper-based form is only to act as a mere byproduct. Apart from the more than 2000 pages 3.600 additional follow-up pages conference material from the various sources have been made digitally available. Thus the stage for further CAAD research work is well prepared (Turk et.al., 2001) and spectacular devices to this end will not have to be invented over and over again.

Figure 3. What users do in CUMINCAD.



References

Björk and Turk, 2000: A Survey on the Impact of the Internet on Scientific Publishing in Construction IT and Construction Management, ITcon, Vol. 5, pg. 73, http://itcon.org/.

Kalay, Y.E.: 1991, CADLINE, A Bibliographic Database on CAAD-Related Publications. Buffalo: State University of New York.

Martens, B. and Z. Turk: 1999, Working Experiences with a Cumulative Index on CAD: "CUMINCAD", in: Architectural Computing from Turing to 2000 [eCAADe Conference Proceedings] Liverpool (UK), pp. 327-333.

Martens, B. and Z. Turk: 2000, The Creation of a Cumulative Index on CAD: "CUMINCAD" in ACADIA Quarterly 19:3, pp. 18-19.

Turk, Z. et.al: forthcoming 2001, The Topics of CAAD - A Machine's Perspective, in: CAAD Futures 2001 Conference Proceedings