

## Cumincad.es: a first step towards multilingual digital libraries in CAAD

Bob Martens, Ziga Turk

*Vienna University of Technology, Austria and University of Ljubljana, Slovenia*

*b.martenstuwien.ac.at, ziga.turk@itc.fgg.uni-lj.si*

This paper presents developments concerning a digital library called CUMINCAD.ES, which includes CAD-related publications in Spanish language. CUMINCAD is an acronym for Cumulative Index on CAD and in this library only publications in English language are archived. While English may be the lingua franca of scientific communication, publications in other languages is important to strengthen local cultures, invent professional terminology, enable technology transfer to those who are not very proficient in English, such as practitioners and students and to provide an incentive for authors to publish in their native languages. CUMINCAD.ES (<http://cumincades.scix.net>) is a repository complementing CUMINCAD providing a basis for the investigation of multilingual repositories. As this is work in progress, the state of the art will be reviewed as well as the potential for further developments.

**Web-based Bibliographic Database; Searchable Index; Retrospective CAAD Research**

### Introduction

Today, most scientists accept the need to publish in English in order to reach a worldwide audience. In most cases this language has, however, been learned by scientists as a second language and the expression in another language than the mother tongue suffers from home-made translations by the authorship. Doubtlessly, a language can serve as a barrier and non-English scientific output has not the same chances to receive wide recognition. Although English has become a kind of universal language in science, still a remarkable output is produced in national languages. The development of a series of web-based multilingual repositories would provide non-English publications a chance for wider dissemination and, furthermore, support the growth of regional scientific communities. Digital publications can be made available at a low price and also back-digitalization does not cause immense expenses to serve the community.

The series of SiGraDi-conferences started in 1997 and so far more than 500 papers have been published. It, however, is getting increasingly difficult to keep track of the previous output. In Venezuela a series of biannual regional conferences started in 1999 called "ConVeACA". As a result of two conference meetings, 56 papers were published. Conference papers are not the only type of entry aimed at. Also theses and dissertations, as well as academic studies and reports etc. are of interest.

### Repository for Publications in Spanish related to CAAD

CUMINCAD.ES (Cumulative Index on CAD - <http://cumincades.scix.net>) was set up in 1999 and serves the Spanish reading CAAD-community as an important source of archived scientific information. A first stage of development concentrated on a complete recording of CAD-related Conference Proceedings (SiGraDi and ConVeACA), which included also summaries and full-text versions. Especially doctoral degree students benefit very much from a collection of structured information. At the time of writing approx. 600 records have been created and nearly 100% of these entries provide also of a full-text version in pdf-format. Smart acquisition techniques were implemented, so that manual work entering the initial fund of records could be reduced.

Nowadays a substantial part of the scientific output can be characterized as so-called "grey literature" (limited number of copies, absence of publisher or ISBN-number, missing archival copies in libraries) in terms of studies, working papers and reports, etc. Proper "archival" for this type of output obviously is granted low priority and furthermore, does not aim at general and quick availability. Also, these materials

have in common that knowledge about their existence is not secured at all and even with the knowledge of bibliographical key data retrieval is difficult (time- and cost-consuming). However, these "publications" could deliver a substantial input to the development of current and future research work. It can be assumed that on a regional level enough scientific output could potentially be collected in a language specific repository, such as CUMINCAD.ES. Within Europe especially e.g. the German and French scientific communities would certainly allow for the setup of similar specialized repositories (CUMINCAD.GER and CUMINCAD.FR).

The CUMINCAD-repository consists at the time of writing of 5.942 records. In total 2.324 papers were published in 55 proceedings in the period 1981-2002 by the CAAD-associations (eCAADe, ACADIA, CAADRIA, SiGraDi and CAADfutures). Furthermore a substantial number of corresponding papers in pdf-format (nearly 50%) could be attached to the records. The annual production of conference papers in English in the field of CAAD has increased significantly in the last decade. For instance in 2001, 355 papers were published in the framework of conference proceedings. However, this seems to be a number which will not grow substantially in the near future. The current extension on an annual base - without CAAD futures [biannual conference] - is approx. 300 papers. CUMINCAD.ES cannot "compete" with these quantities, but the number of stored records is far beyond a critical mass.

SiGraDi-conference papers are also recorded in CUMINCAD if an abstract in English is available. However, the paper title may be in Spanish or Portuguese and in this case an English translation between brackets is added.

### **Grey Literature. Theses and Dissertations**

A thesis or a dissertation is by its very nature produced in very limited quantities and in many cases the only copy available is the archival copy deposited in the library. In the course of a request, cost and delay factors are a significant deterrent. The relatively restricted access to print theses is the predominant reason for their under-utilization. Making the full-text available from any computer desktop across the web would greatly increase knowledge, access and availability of such a significant resource.

An extension with theses and dissertations is planned for the extension of CUMINCAD.ES in the near future. Concerning non-English dissertations and theses listing in some (central) "Union Catalog", where this kind of academic work could be searched is badly needed. Presently, the user has to visit individual university libraries one by one and to get acquainted with a variety of interfaces. This is time-consuming and makes sense only if a specific piece of work has to be retrieved (i.e. university and author are known). The following selected European examples – in alphabetical order - give an impression about a situation with a rich diversity:

Austrian Dissertations Database (<http://www.arcs.ac.at/dissdb/welcome>)  
Delft University of Technology (<http://www.library.tudelft.nl/dissertations/>)  
Dissertationen Online  
([http://www.educat.hu-berlin.de/diss\\_online/biblio.html](http://www.educat.hu-berlin.de/diss_online/biblio.html))  
Eindhoven University of Technology  
(<http://www.tue.nl/bib/ftproefelders.html>)  
ETH-Zurich (<http://e-collection.ethbib.ethz.ch/diss/>)  
Lund University Dissertation Abstracts (<http://www.lub.lu.se/>)  
RWTH Aachen (<http://www.bth.rwth-aachen.de/ediss/ediss.html>)  
Theses in GB and Ireland (<http://www.theses.com/>)

These examples show that a unifying catalogue with metadata including a large number of dissertations and theses (at least by means of bibliographical data extended with a summary) would make sense.

A call for submission by way of the existing CAAD-list-servers could set the stage for the fruitful acquisition of theses and dissertations to CUMINCAD.ES. In this context particularly “younger” dissertations and theses have a good chance to be available in pdf-format, as digital data may be archived.

### **Outlook and Conclusions**

In the sense of a self-organizing repository, the development of CUMINCAD.ES relies on the input from both individuals and associations. As a repository is steadily growing, individual users are invited to record their bibliographies in the corresponding area. The inclusion of e-prints and preprints seems wise and the storage has to be arranged in different categories, so that users of these information packages are well informed about the status of a publication. In fact an individual user could use this repository environment all over the world and this independently from a specific computer. A repository containing a critical mass of initial content such as CUMINCAD.ES would issue an index of authors (with e-mail addresses) and thus support the establishment of direct contacts with authors for further (relevant) submissions. The basic requirement for every submittal would be the delivery of the corresponding full-papers in form of pdf-files as well as the Spanish summaries.

Currently, the web-interface for CUMINCAD.ES is in English. However, a translation of the woda-database tool in a short term would prove meaningful. The work for translation can be estimated with 2-3 days work excluding implementation.

Future work should be related to content analysis of the stored entries as the work concerned with setting-up of the entries-collection is not considered an aim in itself. Also efforts towards automated classification and translation would be of interest.

### **References**

- Martens, B.; Bjoerk, B.-C.; Turk, Z.: 2002, Open, Self Organising Repository for Scientific Information Exchange - The SciX Project, *SiGraDi 2002 Conference Proceedings*, Caracas (Venezuela), pp. 116-120
- Martens, B.; Turk, Z.; Cerovsek, T.: 2001, Digital Proceedings: Experiences regarding Creating and Using, *Architectural Information Management* [Proceedings 19th eCAADe-Conference Helsinki], pp. 25-29

