

# Information System on Literature in the Field of ICT for Environmental Sustainability

Martin Schreiber

Leuphana University of Lueneburg, Sustainability Sciences Department & Computing  
Centre,

Scharnhorststr. 1, D-21335 Lueneburg, Germany (schreiber@uni.leuphana.de)

## Abstract

The literature database *EnviroinfoLit* (lit.ict-ensure.eu) of the ICT-ENSURE project provides substantial scientific papers in the field of ICT for environmental sustainability to the scientific community and to program managers working in this field. The information system that is being developed comprises resources available in the field, including conference proceedings (EnviroInfo proceedings since 1998), workshop proceedings, and other scientific publications. The literature database *EnviroinfoLit* contains besides the meta data the full text of the literature and provides different access through navigational structures, standard and fuzzy search routines. The articles matching the search are available for download as PDF files.

The ICT-ENSURE information system is described in detail in another contribution to this conference proceeding: "Werner Geiger, Richard Lutz, Christian Schmitt: A Pan-European Information System on Environmental Informatics Research Programmes and Projects"

## Keywords

Environmental Informatics; Literature database; European Research Area (ERA); ICT for Environmental Sustainability; ICT-ENSURE

## 1. INTRODUCTION: THE ICT-ENSURE PROJECT

The ICT-ENSURE project ([www.ict-ensure.eu](http://www.ict-ensure.eu)) aims to establish a web information system on research programs in Europe and their results as well as a literature database for publications in the field of environmental sustainability and environmental informatics in general.

The ICT-ENSURE project (**I**nformation and **C**ommunication **T**echnologies - **E**nvironmental **S**ustainability **R**esearch) was applied by a consortium consisting of the University of Technology Graz, the International Society for Environmental Protection (ISEP) Vienna, and Forschungszentrum Karlsruhe (Karlsruhe Research Centre). The project is funded by the 7th Research Framework Programme of the European Union (FP7) and runs from May 2008 until April 2010.

The ICT-ENSURE project was presented in detail at the EnviroInfo 2008 conference in Lueneburg and the EnviroInfo 2009 conference in Berlin, both cities of Germany. According to Tochtermann [Tochtermann et al., 2008] the key objectives of the project are:

- a comprehensive overview on the situation of ICT for environmental sustainability research in Europe
- establishment and extension of a network of experts and communities
- a concept for the creation and further development of SISE (Single Information Space in Europe for the Environment) representing the European environmental landscape

The literature database is part of the web based information system and focuses on the publication of the conference series „EnviroInfo“ and other literature in this research field.

## 2. MOTIVATION

Today, in times of digital libraries and a fast and ubiquitous Internet, scientists may directly access larger data inventories than ever before in the history of science. Large inventories of specialized literature, e.g. the proceedings of the EnviroInfo conference, however, cannot be accessed at all. Proceedings are contained in libraries in the form of monographies only and, hence, certain contributions to the proceedings and in particular their full texts are not accessible. This would be desirable, however, as current developments, research projects, and specialized projects are not only described by journals, but also by these proceedings that may provide scientists with constantly updated information.

In the environmental area, even specialized technical databases, such as the technical environmental library ULIDAT ([Lohse, 1994], doku.uba.de) of the German Environmental Authority, do not provide extensive and up-to-date access to the proceedings to the EnviroInfo conferences. The findings are provided with descriptors, but abstracts can be found occasionally only and full texts are not available. *EnviroinfoLit* is closing this gap by presenting a database on environmental literature with full text and search routines on this full text.

Data mining techniques that are used by the search machine giant Google for the web and by Google Books in the printing area have resulted in a significant increase in the material available, but so far, conferences and workshops in the field of environmental informatics have not been acquired systematically and completely.

This gap shall be closed by the literature database developed within the ICT-Ensure project. It provides the community with access to results in the fields of information science for environmental protection, sustainable development, and risk management. This field is represented by the Technical Committee of GI (German Society for Information Science) that also organizes the international EnviroInfo conferences. At these conferences that have taken place annually since 1986, numerous scientific papers have been produced:

EnviroInfo Conferences	22
Reviewed Papers	3,000
Workshops	98
Pages Documentation	30,000
Authors	7,500

Table 1: Output of the EnviroInfo conference series

Thus it would be of great value for the scientific community to have a system to access all this literature and which will be updated continually.

## 3. STRUCTURE AND TECHNOLOGY

The structure of the database is quite similar to that of a regular literature database. For the specific scope of this database though there is additional information on conferences, authors and their institutions included. This information is necessary to create links between different articles and affiliations of the same author.

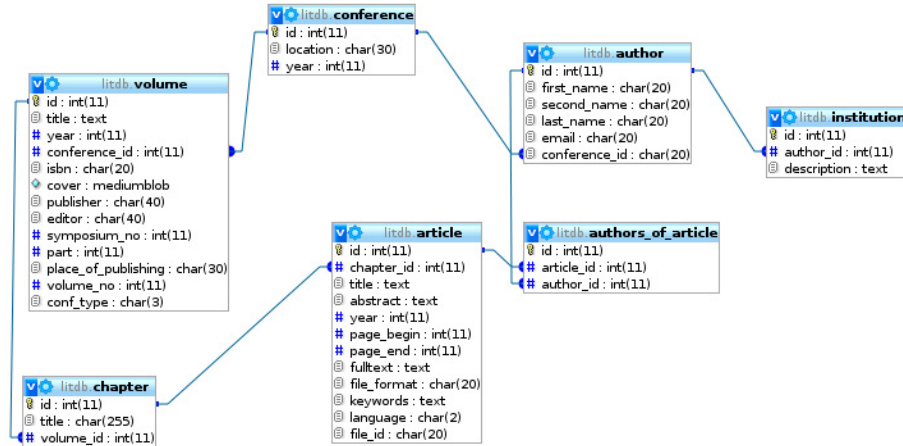


Figure 1: Table diagram of literature database EnviroInfoLit.

Since the collecting and preparation of the data was estimated as a very time consuming process, it had to be started right at the beginning of this project. For this purpose a rapid prototype of a database was built with the proprietary database product *Filemaker Pro* from FileMaker Inc. This database was appropriate for storing and validating the collected literature data right at the beginning. In parallel the development of the final system started with *MySQL* as database management system and *Java Servlet Technology* as programming environment, because the ICT information systems is also based on these technologies. With these two well-known open source products the system is open and also meets the requirements of the EU to use only open source software.

#### 4. CONTENT

To supply literature in this field to the complete extent in full texts and online, an abstract, the full text, meta-data, and a PDF file of every article had to be generated. As the 22 proceedings and workshop volumes were published by various publishers, the data are not available in a standardized format. Since 1998, the proceedings have been published in digital form, but sometimes in different formats. Before 1998, no digital versions of the proceedings were published, such that digitization by scanning with subsequent optical character recognition (OCR) was required.

By now all proceedings of the EnviroInfo conference series since 1995 were included in full text format in the database. Additionally all proceedings of the working group „Umweltdatenbanken / Umweltinformationssysteme“ since 2006 have also been added to the database.

Hence, far more information will be provided than by a conventional literature database. Access to texts is often aggravated by the fact that the proceedings are sometimes out of stock, since usually, a small number is printed only.

#### 5. NAVIGATION AND SEARCH

The literature database holds monographs with an inherent structure. The conference series *EnviroInfo* publishes each year proceedings of the conference in one or two volumes. Each volume is structured hierarchically with chapters and articles. This structure can be used to provide a special, hierarchical access to the articles with menus and submenus. Thus an article can be chosen by selecting a conference year, after this the volume, after this the chapter and finally the article. This access can be of some value, if there is only a vague reminder on the title and the author of an article, but a good reminder of the conference and of the track in which the talk was held.

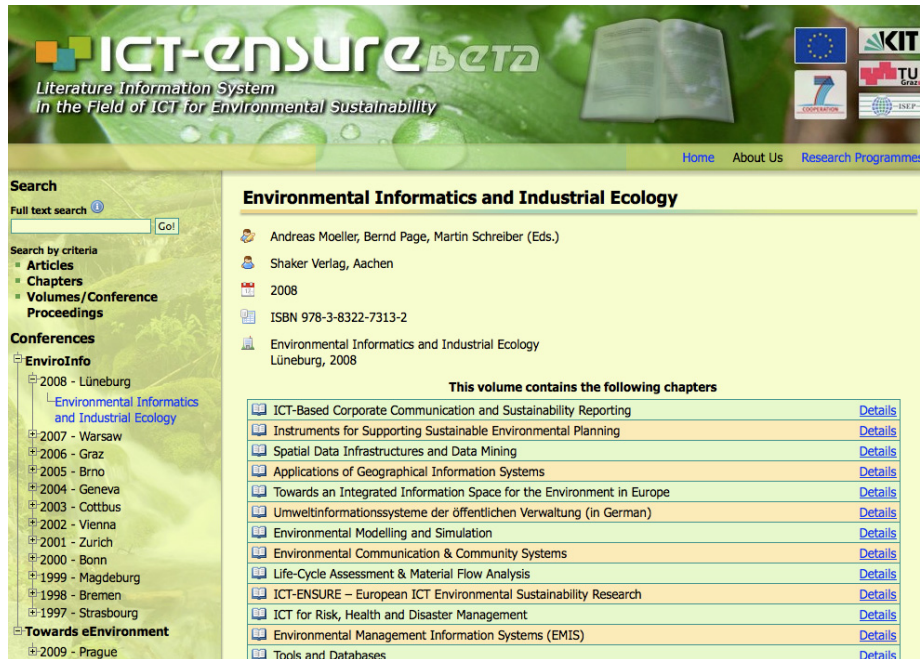


Figure 2: Screenshot of a conference proceeding with chapter information for further navigation.

Of course there are also the standard search routines for each field and the full text search for the abstract and the article itself. Search criteria can be combined with usual logical operators or jokers.

The full text search is implemented with *Lucene* [Hatcher, 2009], an open source java search-framework. *Lucene* generates the index and performs the search.

One special feature is a search using the Levenshtein Distance [Levenshtein, 1966] which some times is called *editing distance*. The Levenshtein Distance is a notion from information theory and denotes a metric for measuring the difference between two strings. A Levenshtein Distance of two means that two editing changes have to be made with one of the strings to get the other one. In this case a search string will match to words even if up to two characters are different.

The bibliographic data of the results of a literature search may be exported flexibly. For this purpose, the standard export format tab-return or, as a bibliographic format, the exchange format of EndNote or the RIS format may be chosen. The RIS format is on the way of becoming a de facto standard. The file format has a simple structure<sup>1</sup> and many bibliography programs like EndNote, Reference Manager, Citavi and digital libraries like SpringerLink, ACM Digital Library, IEEE Xplore and ScienceDirect are supporting it. In the RIS format, the bibliographic data are qualified with tags and can be processed further more easily. Figure 2 shows a screenshot of the article view of one found record of the literature database.

<sup>1</sup> Each entry for the reference starts with two letters, two spaces and a dash and is tagged with this two letters. The beginning of the citation of this article would be:

TY - CONF

AU - Schreiber, Martin

TI - Information System on Literature in the Field of ICT for Environmental Sustainability

...

For further references see: [http://en.wikipedia.org/wiki/RIS\\_%28file\\_format%29#cite\\_note-0](http://en.wikipedia.org/wiki/RIS_%28file_format%29#cite_note-0)

The screenshot displays the ICT-ENSURE BETA website. At the top, the logo 'ICT-ENSURE BETA' is prominent, with the subtitle 'Literature Information System in the Field of ICT for Environmental Sustainability'. Navigation links for 'Home', 'About Us', and 'Research Programmes' are visible. A search bar is located on the left, and a sidebar lists 'Conferences' by year and location, including '2008 - Lüneburg' and 'Towards eEnvironment'. The main content area features an article titled 'Combination of Optimization Methods and Material Flow Analysis for Improvement of Operational Material Use (KOMSA): Concept and its Implementation'. The article is attributed to Lambrecht, Hendrik and Zimmermann, Mike. It includes an abstract, a page number '310 - 318', and options to download the citation or access the full text.

Figure 3: Screenshot of an article with bibliographic data, abstract, and function for downloading the bibliographic reference or the full text of the article as PDF file.

The EnviroInfoLit literature database serves as an important part of the web based information system of the ICT-ENSURE project. Together with the data on research activities in the field of environmental informatics and ICT for Environmental Sustainability it constitutes a substantial source of data on research activities and literature in Europe.

## 6. OUTLOOK

Many data collections and database, which are acquired in projects, are not sustainable because they are not continued after the projects end. The Technical Committee of GI “Environmental Informatics – Informatics for Environmental Protection, Sustainable Development and Risk Management” which organizes the conferences each year will ensure to continue the update of the literature database.

## 7. REFERENCES

- Hatcher, Erik; Gospodnetic, Otis: Lucene in Action – A guide to Java search engine. Greenwich, 2009.
- Levenshtein, V. I.: Binary Codes Capable of Correcting Deletions, Insertions, and Reversals, *Soviet Physics-Doklady*, 10(8), 707–710, 1966.
- Lohse, Siegbert: Die öffentlich zugänglichen Datenbanken des Umweltbundesamtes - Umweltinformation aus der Praxis für die Praxis. In: Informatik für den Umweltschutz, 8. Symposium, Hamburg 1994. Eds. L. M. Hilty, A. Jaeschke, B. Page und A. Schwabl. Hamburg 1994, pp. 227-232, 1994.
- Tochtermann, K., Granitzer, G., Pillmann, W. and Geiger, W.: ICT-ENSURE – A 7th Framework Program Support Action for Building the European Research Area in the Field of ICT for Environmental Sustainability. In: EnviroInfo 2008 – Environmental Informatics and Industrial Ecology, 22th Symposium, Lüneburg, September 10 – 12, 2008. 456-463, 2008.