

SL/AM-IT: A CD-ROM-BASED, INTERACTIVE BIBLIOGRAPHY ON SERVICE LIFE AND DURABILITY

SL/AM-IT: CD-ROM Bibliography

D.J. VANIER, M.A. LACASSE and D.R. PAYER

Institute for Research in Construction, National Research Council Canada, Ottawa, Canada

Durability of Building Materials and Components 8. (1999) *Edited by M.A. Lacasse and D.J. Vanier.* Institute for Research in Construction, Ottawa ON, K1A 0R6, Canada, pp. 2547-2556.

© National Research Council Canada 1999

Abstract

This paper describes a technology transfer technique to handle large *corpora* of technical information. The paper describes and illustrates a sophisticated Information Technology tool to interactively browse over 1500 abstracts and full text documents in the domain of service life and asset management. The first goal of the project is to provide full access to this compendium of information in the topic domain. The second goal of the project is to illustrate a methodology for data storage, data retrieval and technology transfer for other technology domains. The product described in this paper is a compact disk-read only memory (CD-ROM)-based, interactive bibliography on service life prediction, durability and asset management. There are two innovative features of this compendium: (1) a keyphrase index, that was created automatically using sophisticated indexing tools, and (2) an intuitive, interactive interface to large volumes of technical information. The paper describes the problems of accessing the existing large volumes of data and information related to service life prediction, durability and asset management, outlines sophisticated IT tools currently available for indexing and searching large *corpora* of technical information, details the interface and background data structure of the SL/AM-IT (Service Life/Asset Management-Information Technology) methodology and product, and discusses advantages of such techniques in the domain of service life and asset management and other similar technical domains.

Keywords: Bibliography, service life, durability, CD-ROM, Compact Disk, keyphrases, interactive, Internet.



1 Introduction

This paper describes a technology transfer technique to handle large *corpora* of technical information. The paper describes and illustrates a sophisticated Information Technology (IT) tool to interactively browse over 1500 abstracts and full text documents in the domain of service life and asset management.

The first goal of the project is to provide full access to this compendium of information in the topic domain. The second goal of the project is to illustrate a methodology for data storage, data retrieval and technology transfer for other domains.

The product described in this paper is a compact disk-read only memory (CD-ROM)-based, interactive bibliography on service life prediction, durability and asset management. There are two innovative features of this compendium: (1) a keyphrase index, that was created automatically using sophisticated indexing tools, and (2) an intuitive, interactive interface to large volumes of technical information. The paper describes the problems of accessing the existing large volumes of data and information related to service life prediction and asset management, outlines sophisticated IT tools currently available for indexing and searching large *corpora* of technical information, details the interface and background data structure of the SL/AM-IT (Service Life/Asset Management-Information Technology) methodology and product, and discusses advantages of such techniques in the domain of service life and asset management and other similar technical domains.

2 Problem description

2.1 Background information

There has been considerable research work accumulated in the combined fields of service life and asset management; however, much of this information is spread over a wide selection of documents, sources and countries. In the past 40 years, there has been considerable effort by researchers to create this compendium of knowledge; but there is no one central location for the accumulation of this valuable storehouse of information. Journals, conference series, monographs, on-line databases, Internet World Wide Web (WWW) sites and some other CD-ROM are available; however, for many organizations, these references are difficult to locate and many can be expensive to acquire, if they can be located in the first place.

The current search tools for locating this information are not standardized, and many are extremely sophisticated, requiring considerable training to master and normally expertise in library sciences. In addition, there are a number of administrative and legal issues involved in making such information available, including availability of the source documents and copyright laws.

Clearly, in a mature research domain such as service life prediction, durability and asset management there is a considerable *corpora* of technical information, and there are knowledgeable users who require immediate access to this *corpora*, and additionally, many of the copyright owners of the information are public bodies who

promote progressive technology transfer. All that remains is to provide a mechanism to publish this *corpus*, so it is easily accessible to a wide audience of users.

2.2 Technology transfer agenda for CIB and RILEM

The work plan for CIB W080/RILEM 175-SLM (Service life methodologies) follows closely the developments made in the previous joint CIB/RILEM working group on "Prediction of service life of building materials and components" (CIB W080 1999; Jernberg *et al.* 1997). It is based on the elaboration of key areas of knowledge in the service life domain that together provides practitioners with the necessary tools to guide their decisions. The five areas include:

1. Information Technologies in Construction: an Integration Tool in service life prediction
2. Service Life and Durability
3. Selection of Service Life Prediction Methods
4. Understanding Factors Causing Degradation
5. Characterisation of Degradation of Building Materials and Components.

Since 1982, there have been three consecutive joint CIBW80/RILEM Committees, including 71-PSL, 100-TSL and 140-TSL. In this time, members of numerous countries participating in the joint activities have done considerable work. Indeed, this group has been responsible for a series of useful working documents and related literature published by CIB and RILEM and as well, has co-ordinated the efforts required to bring about seven international symposia related to durability and service life issues. The number of significant contributions collectively presented in these conference proceedings provide a substantial depth of knowledge to the field. Full exploitation of this vast body of knowledge for the benefit of manufacturers of building materials and components, designers, specifiers, constructors, as well as asset and property managers, requires the development of suitable guides and related information for use by these various practitioners.

It is the aim of the current CIB W80 Working Commission (WC), working jointly with RILEM 175 SLM, to help develop the necessary guides, methods and techniques that will enable practitioners in selecting the appropriate tools to predict service life. In fact, many of the future contributions of the WC may very well find a place within international standards, as the current working commission actively participates in providing technical expertise to the relevant ISO Working Group developing a standard on Design Life of Buildings. To achieve this aim, it is proposed that existing service life techniques be further developed in the context of emerging information technologies. Hence, the focus of the WC is the integration of existing prediction and service life techniques, tools and methods, using information technologies being developed for the construction industry. Furthermore, stochastic and reliability methods, typically used in the aerospace and automotive industries, will be reviewed as potential means of addressing particular aspects of service life prediction (SLP).

2.3 Existing corpora of information

CIB W80/RILEM 175 SLM, and the proceeding WC's have had an active program in technology transfer for a number of decades. For the past 25 years, many organizations have put their names behind a series of triennial conferences entitled the *Durability of Building Materials and Components* (DBMC), including the National Research Council Canada (Sereda and Litvan 1978), the National Institute of Standards and Technology (Frohnsdorff and Horner 1981) VTT (Sneck and Kaarresalo 1984), National University of Singapore (Lee 1987), Building Research Establishment (Baker *et al.* 1990), Building Research Institute (Nagataki *et al.* 1993), and the Royal Institute of Technology (Sjöström 1996). In addition, a number of journals related to this topic have come and gone, including *Durability of Building Materials* (Elsevier 1987). Other congress proceedings related to this topic domain have also been published, including CIB's *Building Technology, Design and Production* (Sjöström *et al.* 1983) and RILEM's *Durability of Construction Materials* (Maso 1987). These publications form the *corpora* of information for the SL/AM-IT CD-ROM.

2.4 Copyright issues

Fortunately, the copyright to the described *corpora* of information is held by a number of organizations that are interested in making their knowledge more widely available. The authors made contact to secure copyright to the abstracts for the SL/AM-IT CD-ROM either with the proceedings' publishers or host organizations. In the case of the previous seven DBMC conferences, the host organizations and the proceeding publishers were only too happy to see their information resurface in another form, and made their abstracts available. In the case of 8th DBMC, the National Research Council of Canada (NRCC) holds the copyright, and as such, made access to the full text of the documents a possibility. In the future, it is hoped that subsequent DBMC conference organizers or publishers of related journals will make their articles a part of future releases of SL/AM-IT.

2.5 Technology transfer

There are a number of solutions available to address this identified need for rapid access to *corpora* of technical information related to the domain of service life and asset management; some of these are IT-based. New IT tools and techniques now make it possible to create sophisticated applications that would have been extremely expensive to create even a short time ago. The Internet protocol that is in common use on the World Wide Web (WWW) is such a tool (Vanier and Turk 1994). The rapid evolution and universal acceptance of the Internet and the HyperText Markup Language (HTML), provide opportunities for capturing, storing and disseminating information of a technical nature, and in particular would be particularly useful when dealing with the *corpora* of knowledge on service life and asset management. The current implementation of SL/AM-IT is a CD-ROM; however, owing to the use of HTML in the project, there are plans in place to make this *corpora* available on the Internet.

3 Materials and methods

The two goals of the project identified earlier were: (1) to provide full access to specific corpora of knowledge, and (2) to illustrate the methodology for other applications and domains. Two innovative features of SL/AM-IT are: (1) a keyphrase index, that was created automatically using sophisticated indexing tools, and (2) an intuitive, interactive interface to large volumes of technical information. In summary, the system had to be comprehensive, quick and intuitive (user-friendly) for the users, and it had to be easy-to-construct for the developers.

3.1 Process description

Figure 1 outlines the stages followed to bring the data together for the CD-ROM. Starting from the left in Figure 1, the hardcopy documents were scanned using “off-the-shelf” technology, Optical Character Read (OCR’d) using commercial software, and proof-read by (human) editors; thus creating electronic copies of the abstracts and associated information. The two lead authors, in their role as organizers of the 8DBMC, had access to the electronic files for all the conference contributions, so this formed part of the *corpora* of information for the CD-ROM.

These electronic files were processed, as illustrated in the centre of Figure 1, to create the electronic indexes required for the database searching.

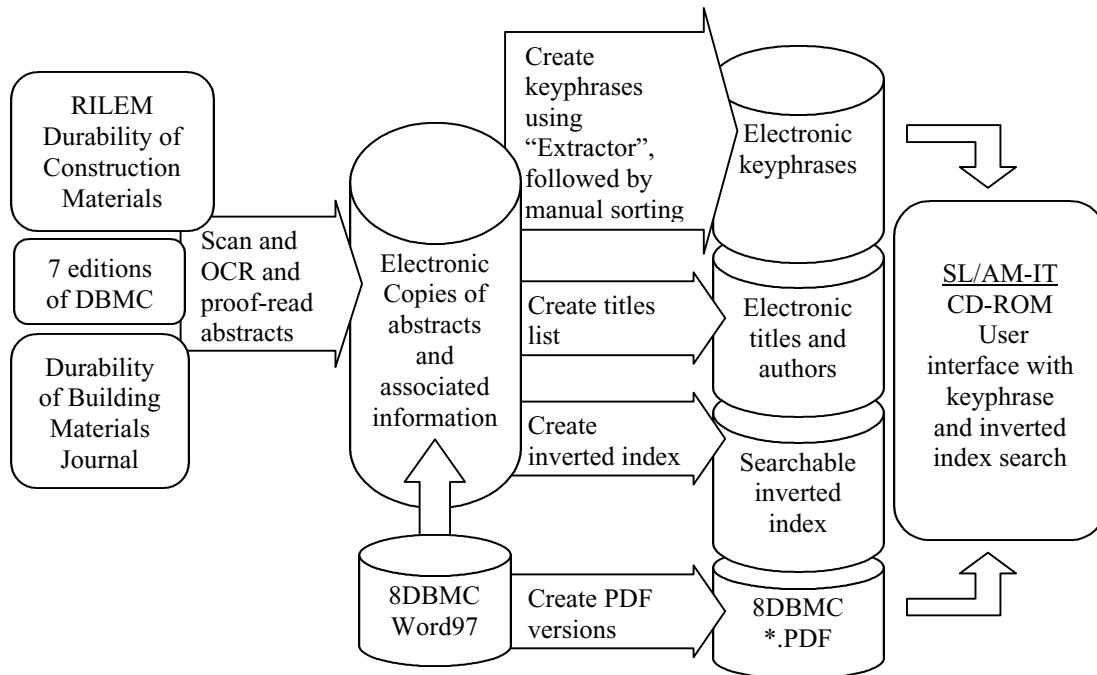


Fig. 1: Process and implementation

3.1.1 Extractor software

The Institute for Information Technology of the NRCC has developed a software product that uses the full text version of documents to create keyphrases (Extractor 1999). This program was used successfully in a related project (Kosovac and Vanier 1999), and the authors saw many opportunities for such a technology in

the storing and dissemination of technical information. The general theory behind the “Extractor” methodology is that automated tools can systematically and objectively create keyphrases from full text documents. One of the advantages of using this automatic generation of keyphrases, such as the “Extractor” methodology, is the objectivity in the selection of keyphrases; normally, authors always have some form of bias in their selection of keyphrases. Computer generation of terms is also precise, repeatable, fast and simple (Turney 1997). “Extractor” has been successfully benchmarked against a number of other “indexing” software (Turney 1997).

To create the desired keyphrases for indexing and searching for SL/AM-IT, the individual abstracts and full text documents from the *corpora* were individually indexed using “Extractor”; this produced six keyphrases per document. This list of keyphrases was combined with the keyphrases supplied by the authors, and a domain expert purged the resulting list for duplicates and similar keyphrases. The resulting short list formed the “pick list” of keyphrases for the *corpora* searching.

3.1.2 Title and author lists

Title and author lists were also created for searching; thus allowing the user to specifically identify authors from the “Master Service Life List”, shown in Figure 2.

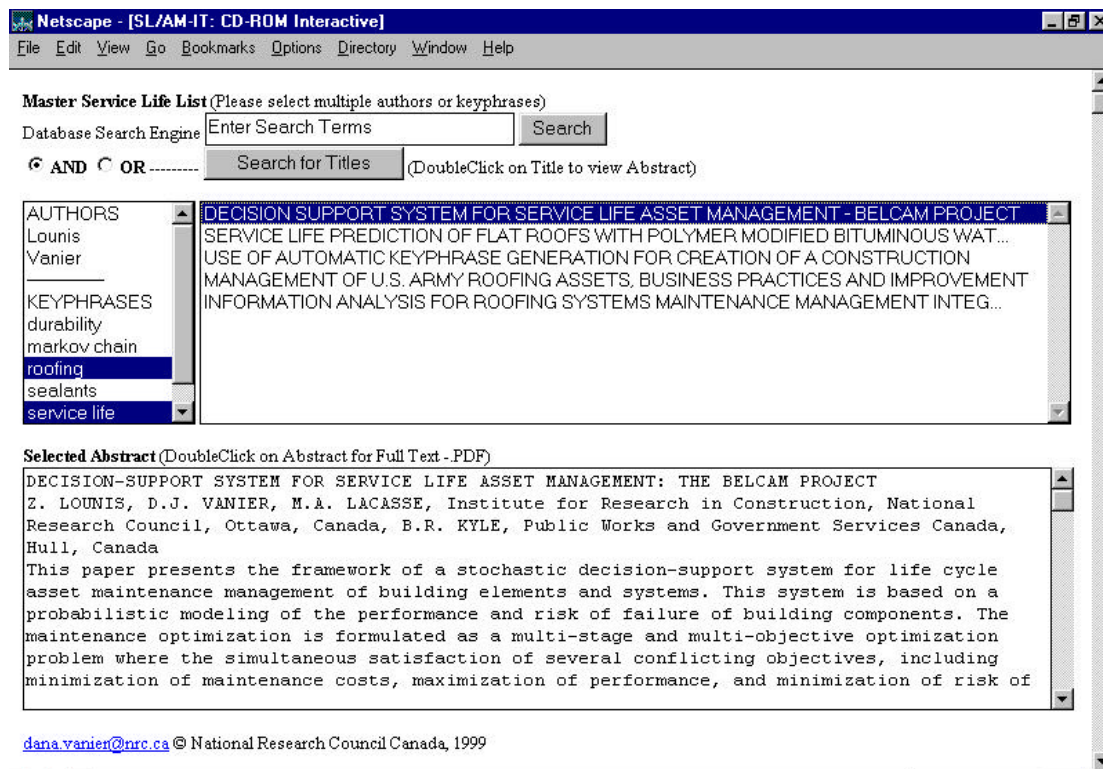


Fig. 2: CD-ROM: Hypertext user interface

3.1.3 Inverted index

A standard inverted index of the text documents forming the *corpora* of knowledge was created for rapid searching. An inverted index creates a table of terms used in the *corpora* along side their exact location in the documents. The user

simply enters the desired term for searching and the contents of the *corpora* are quickly scanned. This type of “search engine” is a standard feature of the Internet.

3.1.4 Portable Document Format (PDF) files

As illustrated in Figure 1, the Portable Document Format (PDF) files for the 8DBMC contributions form part of the *corpora* of SL/AM-IT. The text versions of these documents were also indexed using “Extractor”.

3.2 SL/AM-IT interface

The graphical user interface was designed to provide a number of access points to the database, including the keyphrase information provided by the author, the keyphrases provided by Extractor, as well as the inverted index and the full text of the PDF files. Features of the user interface are described in the following sections.

3.3 User interface design

Figure 2 illustrates the search input screen for SLAM-IT. This is a typical WWW interface using standard HTML. The interface can be used with any number of Internet browsers, allowing complete portability of the SL/AM-IT data and interface. The search options are found at the top of the screen in Figure 2, where the user can use the inverted index to select specific “free form” terms or the keyphrase options.

The middle portion of Figure 2 shows the interface where the user can select a number of keyphrases and/or author names. In the example shown, the keyphrases “roofing” and “service life” were selected. The user can request either “AND” or “OR” Boolean operations to be performed; the results of the “AND” operation are shown in the centre right of Figure 2. Five articles form the results of searching on both those keyphrases. To view the abstract for the first listing entitled “Decision-Support System for Service Life Asset Management”, the user double-clicks on that output line and the abstract appears at the bottom of Figure 2. The three input fields described above are all scrollable; allowing the user to browse over the data any number of times until a suitable article is located. If the full text version of the document of interest is available on the CD-ROM, then the user clicks on the abstract to view the PDF file for that article, as shown in Figure 3.

3.4 System design

Figure 4 provides an illustration of the system design for the SL/AM-IT CD-ROM. The “Author/keyphrase” lists, shown at the left of Figure 4, have many-to-many relationships to the “Title Lists”. That is, there can be many authors and many keyphrases for each paper, as well as, each author and each keyphrase can appear on a number of papers.

In the centre of Figure 4, each “Title List” entry points to one abstract citation (including the listing of the name of the publication, article title, list of authors, author affiliations, and author-created keyphrases). There can be PDF full text documents for some of the abstracts in the “Abstract List”, as shown on the right of Figure 4. This software design permits a simple and easy-to-manage mechanism to locate files using standard HTML protocols.

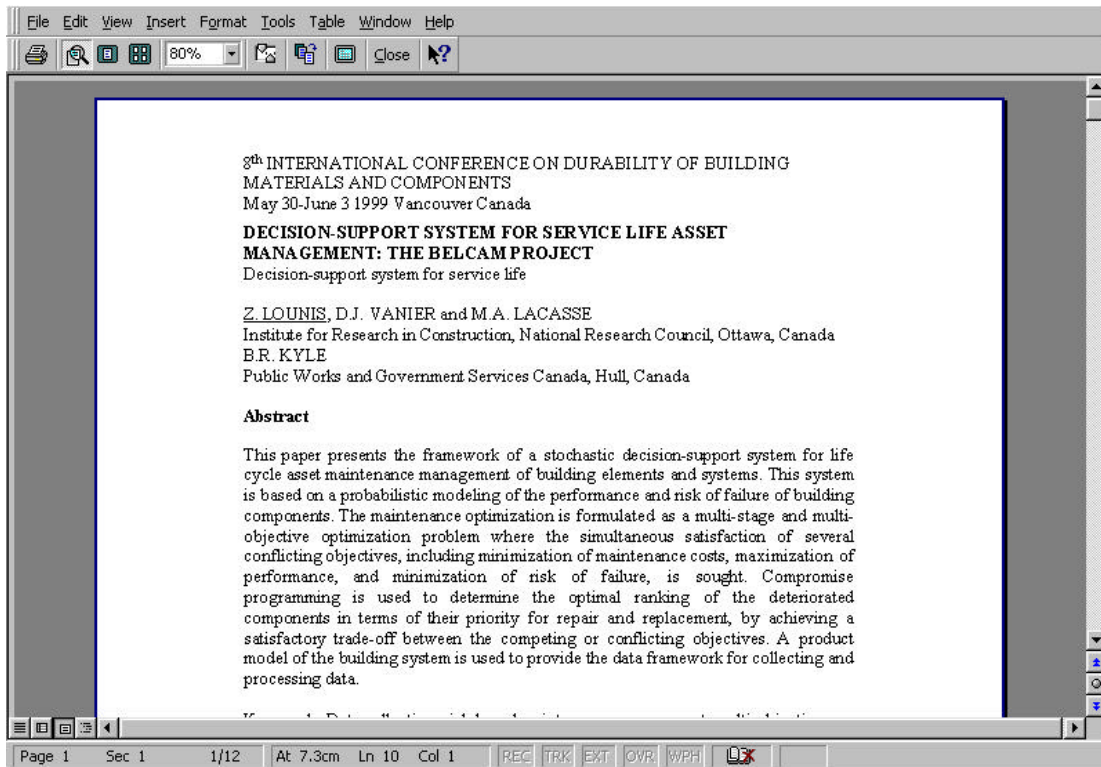


Fig. 3: Full-text, printable and searchable version of the document

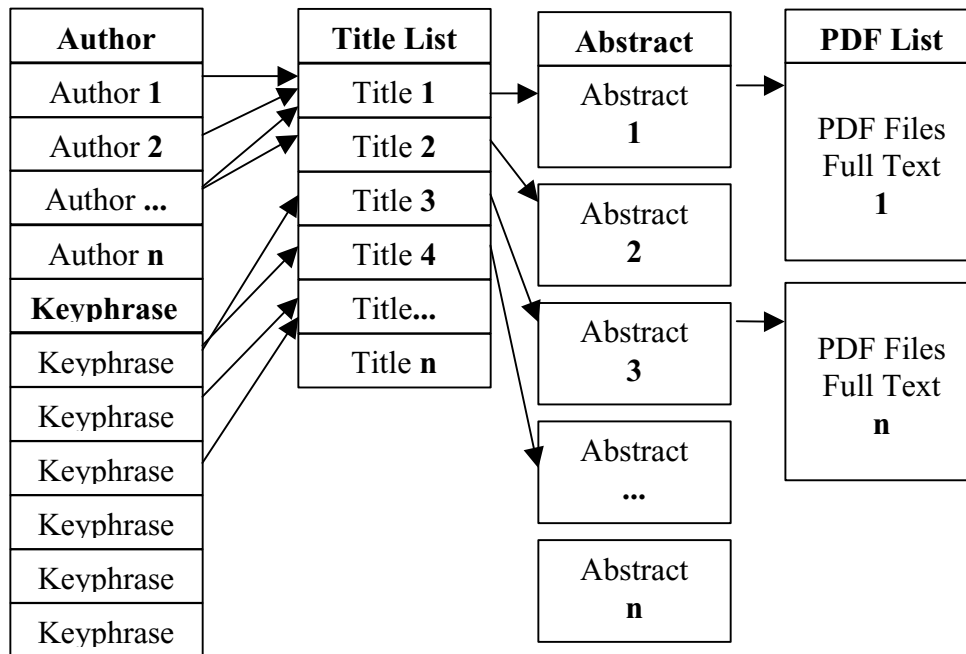


Fig. 4: Data structure

4 Discussion

The SL/AM-IT CD-ROM meets both goals established in the project; it allows complete access to large *corpora* of data and information related to service life and asset management, and its methodology] can be modified to suite similar *corpora*.

The SL/AM-IT CD-ROM addresses many of the search and retrieval problems identified at the start of the project: (1) it provides a repository for considerable information in the area of service life prediction and asset management and the methodology can be used for most *corpora* of knowledge; (2) the combination of the keyphrases and the inverted index provides quick and accurate retrieval of information for both novice and expert users, and (3) because of the portable nature of the HTML protocol, both CD-ROM and Internet applications are readily available.

“Extractor” provides objective, unbiased keyphrases that can be used by novice and expert alike for searching of the *corpora* of knowledge. The interactive user interface is simple-to-use and provides both the keyphrase search capabilities as well as inverted index searching. In addition, the user interface is easily recognized by, and is quite familiar to, a large percentage of Internet users.

The system design and data structure are extremely easy to duplicate and to create for any domain *corpora*. All of the software used in the processing can be purchased commercially, including “Extractor”. All the files used in the operation are text files and all the lists were created using standard database techniques and word processing software. Although there was considerable manual intervention in the creation of all these files, many of the operations could be automated.

One of the advantages of this HTML-based system design is the possibility to continually update the *corpora* of knowledge in the future by including future DBMC conference proceedings and other full text documents, alongside full text articles from previous conferences and related journals, where only abstracts exist currently.

5 Conclusions

This paper describes SL/AM-IT, a technology transfer tool using IT to serve large *corpora* of technical information to users. The two goals were achieved in the course of this project: (1) provide full access to a compendium of information, and (2) illustrate a methodology for data storage, data retrieval and technology transfer for other domains. There are two innovative features of SL/AM-IT: (1) a keyphrase index was created automatically using automated indexing tools, and (2) an intuitive, interactive interface accesses large volumes of information. SL/AM-IT is a valuable contribution to the service life and asset management, and to CIB W80/RILEM 175-SLM; it can be seen as a *corpora* of knowledge to be expanded in the future.

6 Acknowledgements

The authors would like to that Dr. Peter Turney and the Institute for Information Technology of the NRCC for the use of the “Extractor” software and their assistance during the course of this project.

7 References

- Baker, J.M., P.J. Nixon, A.J. Majumdar and H. Davies, Eds. (1990) *Durability of Building Materials and Components*, Proceedings of the Fifth International Conference held in Brighton, U.K., 7-9 November.
- CIB W080 (1999) <http://www.nrc.ca/irc/cibw80/>
- Elsevier (1987) *Durability of Building Materials* Vols. 1 to 5 (1982 - 1987), Elsevier, Amsterdam, The Netherlands
- Extractor (1999) Extractor Home Page, <http://extractor.iit.nrc.ca>, Institute for Information Technology, National Research Council Canada, Ottawa, Canada.
- Frohnsdorff, G. and B. Horner, Eds. (1981) *Proceedings of the 2nd International Conference on Durability of Building Materials and Components*, 14-16 Sep., National Bureau of Standards, Washington, D.C., USA.
- Jernberg, P., C. Sjöström and M. A. Lacasse (1997) *State-of-the-art report, TC 140-TSL: Prediction of service life of building materials and components*, Materials and Structures, Supplement March, pp. 22-25.
- Kosovac, B. and D.J. Vanier (1999) "Use of Automatic Keyphrase Generation for Creation of a Construction Thesaurus", Accepted by 8th International Conference on the Durability of Building Materials and Components, 30 May-3 June, Vancouver.
- Lee, S.L., Ed. (1987) *Durability of Building Materials and Components*, Proceedings of the Forth International Conference, Singapore, 4-6 Nov.
- Maso, J.C., Ed. (1987) *Durability of Construction Materials*, Proceedings of the First International Congress held by RILEM and organized by AFREM (Versailles, France, September 7-11, 1987), Chapman & Hall, London, UK, 1447p.
- Nagataki, S., T. Nireki, and F. Tomosawa, Eds. (1993) *Durability of Building Materials and Components*, 2 Vols., Proceedings of the 6th International Conference, Omiya, Japan, Oct., E & FN Spon, London, UK.
- Sereda, P. and G. Litvan, Eds. (1978) *First International Conference on Durability of Building Materials and Components*, ASTM STP 691, American Society for Testing and Materials, Philadelphia, PA.
- Sjöström, C., Ed. (1996), *Durability of Building Materials and Components* 7, 2 Vols., Proceedings of the 7th International Conference, Stockholm, Sweden, May, E & FN Spon, London, UK.
- Sjöström, C., H-A. Löfberg and A. Zingmark, Eds. (1983), *Building Technology, Design and Production*, Vol. 2 of: Proceedings of the 9th CIB Congress, Stockholm, Sweden, 1983.
- Sneck, T. and A. Kaarresalo, Eds. (1984) *Third International Conference on the Durability of Building Materials and Components*, 4 Vols., Espoo, Finland.
- Turney, P. (1997) *Extraction of Keyphrases from Text: Evaluation of Four Algorithms*, Report ERB 1951, Institute for Information Technology, National Research Council Canada, Ottawa, Canada.
- Vanier D.J. and Z. Turk (1994) "Internet Opportunities for Distributed Construction Information", *International Journal of Construction Information Technologies* 2(3), pp. 45-65.