COMPUTER ASSISTED COMPILATION OF STATE-OF-THE-ART REPORT

Devika Aptagiri, Sudhanshu Bala Satapathy and A.R.D. Prasad

Documentation Research and Training Centre, Indian Statistical Institute, Bangalore

ABSTRACT

Deals with the steps involved in compilation of a State-Of-The-Art Report (SOTAR). Discusses a methodology used for computerised compilation of SOTAR. Applies a simple computer assisted technique of using of CD-ROM databases in compiling the report. Gives a detailed description of the steps in which the work was carried out.

INTRODUCTION

Increasingly, information/knowledge is increasing to be viewed as a basic national resource. Therefore information transfer and dissemination has been recognized as an essential element and input for all research and development activities. The manifold activites, complex and continuous explosion of knowledge at an exponential rate has resulted into explosion of the variety of recorded information which requires efficient handling for proper and timely utilisation.

Library and information science professionals who are primarily responsible for collecting, organising and disseminating information are forced to devise new information services and products to cope up with the situation. In this connection various information products are designed. State-Of-the-art Report is one such product of information analysis and consolidation.

State-of-the-art-Report is an exhaustive, systematic and sometimes critical review of published and unpublished materiral on a specific subject. It is a review concentrating on the most up-to-date information and literature in a given subject or topic. It is a tool for ascertaining the latest trends in a field of study and the extent of its development.

2. NEED FOR STATE-OF-THE-ART REPORTS

Research is carried out in different fields of study constantly. This adds to the information available on the area and results in its development. New theories and products arise out of research activities. Some concepts may become obsolete or there may be a new interpretation of old theories and practices. This implies that every field is dynamic as developments are incorporated into it. It is necessary that these developments are traced and recorded in a source of information. State-of-the-art-reports (SOTAR) are compiled for the purpose of giving a report on the developments in a field. It serves as an important source of information for researchers who have to define the scope of research topic and also seek to know the latest developments in area of their interest.

4. STEPS IN COMPILATION SOTAR

Until recently SOTAR preparation was a job of subject specialist. But the advances made in the field of library and information science especially in the technology of organising and displaying the information, has lead to the library and information science professionals playing a dominant role in the preparation of State-of-the-Art-Reports.

There are various activities involved in compilation of SOTAR. Traditionally, the work begins with defining the scope of the report, identification of sources and collecting necessary information from the sources and then consolidation into a state-of-the-art report. Different stages involved in the manual preparation (1) of SOTAR are as follows.

4.1 Determination of scope of the SOTAR :

The scope of the report has to be determined according to the needs of the users of the report. Scope may be defined by the broad subject or using a string of keywords. Also the depth of the subject required by the target users should be determined. When describing the latest developments in a subject, a certain amount of jargon cannot be avoided. The exact coverage therefore would depend upon the requirements of the end-user of such a product. In other words, scope determines what is to be included and what is to be excluded.

sources :

Information sources mark the development in a subject. Usually the most recent findings are reported in primary periodicals. These form a major source of information in compilation of SOTAR. However, the other sources like monographs, reviews, reference books, technical reports,

proceedings of seminars and conferences, patents and text books also cover details about the field and serve as input to the report. Non-documentary sources are the most important means of communicating the latest developments in any field area through informal exchanges in seminars and conference or through unpublished handouts that may be circulated within a circle.

A compilar of a state-of-the-art report will have to depend largely on secondary sources like the abstracting and indexing journals in order to trace the relevant items that were published. So the first step after determining the scope of the report to be compiled would be to trace the secondary and tertiary sources which would direct to the primary sources available on the subject. It would not be amiss to also go through reports already compiled on related topics or broader subjects areas.

4.3 Appraisal of information collected :

The information collected from various sources needs to be anlayzed to suit the purpose at hand. A state-of-the-art report should be crisp and to the point. This warrants a careful analysis and selection of items to be included. The scope of the report dictates its contents. Generally the significant milestones in the development of the field are noted down. The information that traces the trneds in the field is included.

4.4 Arrangement of ideas in a helpful sequence :

Ideas can be arranged in different ways. Choice of particular sequence should be based on its helpfulness to user. The objective of SOTAR is to record the latest developments in a field. The most helpful sequence of concepts in keeping with the objective would be a chronological arrangement of the ideas tracing the subject form its inception and following the trends and developments.

4.5 Integration of the information into an organized text :

The information collected from the various sources after analysis and arrangement of ideas has to be integrated into an organized text. The result of integration should conform to the principle of unity of ideas. In this integration, the objective of the SOTAR should be kept in view. The flow of ideas should be logical.

5. COMPUTERIZATION OF STEPS IN SOTAR COMPILATION :

Information technology has a great impact on the activities in libraries and information science. Information technology consists of techniques in information handling for efficient and effective information management. It encompasses activitieslike generation, processing, storage, retrieval and dissemination.

Documentation of information has been influenced by new developments in information technology. Computers have influenced many fields and documentation could not quite naturally remain unaffected. In fact, librarians have b very hopeful by observing the computer applicatio different documentation work. At present the developm in information technology has crystallized further and we now able to see a more clear picture of computer applica in library and information work.

5.1 Procedure for compilation of SOTAR :

A methodology has been developed using computers in compilation of SOTAR. Availability of CD-ROMs in the fi of library and information science has made it possible simplify the task of information collection. All the referenc were downloaded from the LISA and ISA CD-ROM using appropriate keywords. As therer is no standardised form both LISA and ISA record structure, the program LISACDS.EXE AND ISA.EXE are used on LISA and IS outputs respectively. The records in standard format a onverted to ISO-2709 format using program Pygmalia CDS/ISIS allows importing records that are in ISO-27 format. A consolidated CDS/ISIS database is thus create As both ISA and LISA include many common referenc there will be duplications in the CDS/ISIS databases. The duplicates are removed manually after taking a sorted outp Records are analysed for their advance and decision taken regarding their inclusion or deletion. Finally Informati is integrated into a coherent text. Following steps we adopted for preparation of SOTAR:

Steo 1. Determination of scope :

The State-of-the-art report was prepared on the subje 'Multimedia Technology'. Multimedia is a combination different media like text, graphics, audio, video, animatio and still picture. Multimedia technology refers to a combinatio of technologies that enables the PC to utilise sound animation, graphics and video.

Step 2. Collection of information from computerised databases :

The sources of information used to generate the SOTA were chiefly CD-ROM databases i.e. LISA (Library an Information Science Abstracts) and ISA (Informatio Science Abstracts). LISA and ISA are publications bibliographic reference to articles in the field library an information science respectively. They include annotat abstracts to the articles. They are updated annually.

Step 2.1 Downloading of bIbliographic records :

LISA (Library and Information Science Abstracts) and I (Information Science Abstracts) CD-ROMs were search under the keywords or descriptors of the particular fie 'Multimedia', 'Hypertext' and 'Hypermedia'. The sea interfaces to LISA and ISA are simple. Search quer stated by the fields or descriptors using boolean operat The search is restricted by 'english' as the language records to bedownloadedand date of publication betw 1990 to 1995. The downloaded bibliographic referen contain the fields, author, abstract, source, source information, publication year and descriptor.

Records of LISA and ISA are not in a standard format (Appendix I and II). Standardisation method has been adopted to convert the records to a standard format.

Step 3. Standardisation of bibliographic records :

All the bibliographic records with full citation are downloaded. But the output of LISA is different from that of ISA.

For example the LISA output was,

Title in English:
Author LN:
Author FN:
Source: etc:

ISA output of the bibliographic records are in this form:

TI:
AU:
PY:
SO: etc.

Therefore to achieve uniformity betwen the two databases, two programs LISACDS and ISACDS, written in C language are used. After execution of ISACDS on ISA database and LISACDS on LISA database. The outputs are integrated into a common file.

Features of LISACDS.EXE and ISACDS.EXE

Lisacds.exe and Isacds.exe are programs written in C language, developed at DRTC. The programs take care of repeated fields such as author and descriptor fields, and assigns separate postings in the output for each occurence of data element in a field.

For example:

If two or more authors are there for a single article the posting will be

AU: First author
AU: Second author

In LISA output the author names are printed out as.

forename, FN: and lastname, LN:

For example:
LISA output:

Author LN: Ranganathan Author
FN: S.R.

LISACDS.EXE output:

AU: Ranganathan, S.R.

The programs also resolve the fields which are occuring in more than one line into a single line as required by the program pygmalian. Pygmalian is used to produce records in ISO 2709 format.

The output of LISACDS.EXE and ISACDS.EXE are given in Appendix III.

Step 4. Conversion to ISO-2709 Format Using Pygmalian:

The program pygmalion (developed at DRTC), was used for conversion the records into ISO 2709 formats. The pgoram PYG.EXE asks for field names and respective tag numbers which are to be converted into ISO-2709 format. The field names along with the tag numbers are entered. The program prompts the steps and the records were converted the records to ISO-2709 format. Output of Pygmalian is given in Appendix IV.

Step 5. Importing Records to CDS/ISIS :

The field definition table, worksheets and display formats for the CDS/ISIS database were created using the same tag numbers used in Pygmalian. Using the master file service all the records in ISO-2709 format were imported to CDS/ISIS.

Step 6. Removal of Duplicates :

There were a number of duplicates after merging both LISA & ISA Databases. For removing duplicates different fields like title, author, source of information are taken. The print format for this was

v1*0.15,c20,v2*0.10,c50,v4*0.10,c65,f(mfn,5,0)

Records were sorted aplphabetically. The master file numbers of the rcords which had same author, title, source of information were noted. Ultimately, the duplicate records were removed manually by using editing service of CDS/ ISIS.

Step 7. Identification of subject headings using VCDs :

The aid of Vocabulary control devices is essential for identifying the subject headings for presentation of information. A subject thesaurus may be used for the purpose. If a thesaurus is already existing in the subject area of the SOTAR it may be used directly. On the contrary, if it is not available a thesaurus may be constructed where input is in form of Prolog facts which state the broader term, narrower term and synonymous relations to terms. The input and output of the thes.ari programs are as given in appendices V and VI.

Step 8. Analysis of information :

The information thus obtained was anlayzed using the editing service of CDS/ISIS. Irrelevant records or irrelevant descriptors or lines found in the records were eliminated.

Step 9. Arrangement of concepts :

Using the sorting services of CDS/ISIS, the records were

sorted under subject headings and sub-topics. Under each subject headings and sub-topics, minute topics are arranged chronologically.

Step 10. Integration into organized text :

This steps involves the description of the individual concepts obtained through the above mentioned process in order to form a coherent text. Since the input was downloaded from CD-ROMs the records were isolated without continuity of thoughts. Therefore trends were related and collaboration was given to the facts wherever necessary to make narrative report. In addition various text books have been consulted to arrive at the chapter headings and the order the chapters should be arranged. In each chapter the broadest terms are considered as section headings. In other words, the keyterms play the role of sections, sub-section, sub-sub-section etc. If NT1 corresponds to section heading, NT2 may correspond to sub-section and so on. A general idea of the chapter under consideration and the order the BT1 are to be presented should is to be acquired from knowledge of text books. In addition, under each sectionn heading we have presented the definition of the term (Section heading) so that it provides a kind of introduction to the concept. This is followed by the list of abstracts arranged chronologically by CDS/ISIS print and sort services. The major job in the preparation of SOTAR is to edit the abstracts under each sub-section. This editing obviously involves a lot of intellectual/manual work and cannot be substituted by the present day technology.

6. CONCLUSION

The methodology items from described above has been found to be quite adoptive to many State-of-the-art repor compiled later. In manual compilation hundreds of abstracts have to be collected. The identification and actual selectio of periodicals is a very tedious process. Also all source required for the purpose may not be available. The tim factor involved in the manual compilation was reduce greatly by adopting the steps described above.

7. REFERENCES

NEELAMEGHAN (A)
Presentation of ideas in technical writing.

Ed1. Sharada Ranganathan Edownment for Library Scienc Series No. 7. Bangalore. 1975.

PRASAD (A R D) and KARISIDAPPA (C R). Declarativ programming and thesaurus construction. ISALIC Bulleti Vol 38 (4). 1993. Pp 161-67.

PRASAD (K N).
Ability of information worker for information analysis an consolidation. Library Science with a slant to documentatio and information studies. Vol 22 (3). 1985. Pp 153-60.

RAB (S F).
The changing frontiers of library information science Common welath publishers; New Delhi 1994.

SATAPATHY (S B).
Multimedia Technology : A State-of-the-art report. DRT Project II. DRTC. Bangalore. 1995.

APPENDIX - I

OUTPUTA OF LISA

Database Title : Library and Information Science Abstracts

Title in English : The use of title and cited title as document representation for automatic classification.

Author LN : Kwok
Author FN : K.L.
Source : Inf. Proc. Man.
Source Info :11 (8/12) 1975, 201-206, 4 table, 9 refs.

Abstract :

The use of title and cited title wsords as document representation is explored. It offers a method intermediate between the use of title and abstract of a document and that of citation identities, retaining some advantage of both. Compared with title and abstract, it leads to more compact and uniform document representation with a high concentration of indicative words, gives more consistent coupling strength to profile with results agreeing well with the method employing citations, and offers a more consistent ability for inter-group differentiation when the groups are close to each other. Compared with the use of citation, it gives results with less specificity and operationally requires an extra step to input and analyse the full citation titles. However, the group profiles derived from title and cited titles are words and can be used to classify documents that have descriptive abstract but no or few citations.

Language : English
Publication Year : 1975
Subject : Cited titles
Titles
Source materials
Extraction
Terms
Technical Services
Information storage and retrieval
Information work
Subject indexing
Automative Subject indexing

APPENDIX II

OUTPUT OF ISA

TI : The use of title and cited titles as document representation for automatic classification.

AU : Kowk, K.L
SO : Inf. Proc. Man.
SI : 11 (8/12) 1975, 201-206. 4 tables. 9 refs

PY : 1975
LA : English
AB :

The use of title and cited title words as document representation is explored. It offers a method intermediate between the use of title and abstract of a document and that of citation identities, retaining some advantages of both. Compared with title and abstract, it leads to more compact and uniform document representation with a high concentration of indicative words, gives more consistent coupling strengths to profiles with results agreeing well with the method employing citations, and offers a more consistent ability for inter-group differentiation when the groups are close to each other. Compared wtih the use ofcitation, it gives results with less specificity and opperationally requires an extra step to input and analyse the full citation titles. However, the group profiles derived from title and cited titles are words and can be used to classify documents that have descriptive abstracts but no or few citations.

MESH : Cited titles
Titles
Source matereials
Extraction
Terms
Technical Services
Information storage and retrieval
Information work
Subject indexing
Automatic subject indexing.

APPENDIX - III

OUTPUT OFLISACDS.EXE AND ISACDS.EXE

TI : Theuse of title and cited titles as document representation for automatic classification.
AU : Kwok, K.L.
SO : Inf. Proc. Man.
SI : 11 (8/12) 1975, 201-206. 4 tables. 9 refs

AB :

The use of title and cited title words as document representation is explored. If offersa method intermediate between the use of title and abstradt of a document and thatof citation identities,retaining some advantages of both. Compared with title and abstracts, it leads to more compact and uniform document representation with a highconcentration of indicative words, gives more consistent coupling strengths to profiles with results agreeing well with the method employing citations, and offers a more consistent ability for inter-group differentiation when the gorups are close to each other. Compared with the use ofcitations, it gives results with less specificity and operationally requires an extra step to input and analyse the full citation titles.However, the group profiles derived from title and cited titles are words and can be used to classify documents that have descriptive abstracts but no or few citations.

LA : English
PY : 1975
DE : Cited titles
DE : Titles
DE : Source materials
DE : Extraction
DE : Terms
DE : Technical services
DE : Information storage and retrieval
DE : Information work
DE : Subject indexing
DE : Automatic subject indexing

TI : Towards automatic indexing: automatic assignment of controlled-language indexing and classification from free indexing.
AU : Field, B.J.
SO : J. Docum.
SI : 31 (4) Dec 75, 246-265. 4 illus. 3 tables, 14 refs

AB :

A number of techniques have been studied for the automatic assignment of controlled subject headings and clasifications from free indexing. These techniques involve the automatic manipulation and truncation of the free-index phrases assigned to a document and the use of a manually-constructed thesaurus and automatically-generated dictionaries together with statistical ranking and weighting methods.These are based on the use of a statistically generated 'adhesion coefficient' which reflects the degree of association between the free-indexing terms, the controlled subject headings, and the classifications. By the analysis of a large sample of manually-indexed documents the system generates dictionaries of free-language and controlled-language terms together with their associated classifications and adhesion coefficients. Having learnt from the manually-indexed documents the system uses these dictionaries in the subsequent automatic classification procedure. The accuracy and cost-effectiveness of the automatically- assigned subject heading and classifications has been compared with that of the manually system. The results were encouraging and the costs comparable to those of a manual system.

LA : English
PY : 1975
DE : British Library
DE : INSPEC
DE : Electrical Engineering
DE : Physics
DE : Computerized information services
DE : Information services
DE : Classification
DE : Technical services
DE : Information storage and retrieval
DE : Information work
DE : Subject indexing.

APPENDIX - IV

OUTPUT OF PYGMALIAN

01483000000000000021700000000

#The use of title and cited titles as document representation for automatic classification. #Kwok, K.L.#Inf. Proc. Man.#11 (8/12(1975, 201-206. 4 tables, 9 refs#. Thse use of title and cited title words as document representation is explored. It offers a method intermediate between the use of title and abstract of a document and that of citation identities, retaining some advantages of both. Compared with title and abstract, it leads to more compact and uniform document representation with a high concentration of indicative words, gives more consistent coupling strengths toprofiles with results agreeing well with the method employing citations, and offers a more consistent ability for inter-group differentiation when the groups are close to each other. Compared with the use citations, it gives results with less specificity and operationally requires an extra step to input and analyse the full citation titles. However, the group profiles derived from title and cited titles are words and can be used toclassify documents that have descriptive abstracts but no or few citations. #1975 #Citedtitles #Titles # Source materials# Extraction# Terms# Technical Services# Information storage and retrieval# Information work# subject indexing# Automatic subject indexing##.

APPENDIX V

INPUT FOR THES.EXE

bts(HypertextSoftware',hypercard',notecard',hyperwriter',hyperpad',hypergraph',knowledgepro',authorware pro',memex',architext',hyperdocument',hyperties',HAM','hypercale'])

syn ('Hypertext Software',['Hypertext package'])

bts ('hypercard',['Hypertalk language']).

bts ('Information Systems',['"multimedia information system']).

bts ('User Interface',['graphical user interface','menu driven interface','icon driven interface','command driven','natural language interface','window driven']).

syn ('User Interface',['front end','gate way']).

syn ('Hypertext Abstract Machine',['HAM']).

bts ('graphics', ['computer graphics','graphical interface'].

LA : English
PY : 1975
DE : British Library
DE : INSPEC
DE : Electrical Engineering
DE : Physics
DE : Computerized information services
DE : Information services
DE : Classification
DE : Technical services
DE : Information storage and retrieval
DE : Information work
DE : Subject indexing.

APPENDIX - IV

OUTPUT OF PYGMALIAN

01483000000000000021700000000

#The use of title and cited titles as document representation for automatic classification. #Kwok, K.L.#Inf. Proc. Man.#11 (8/12(1975, 201-206. 4 tables, 9 refs#. Thse use of title and cited title words as document representation is explored. It offers a method intermediate between the use of title and abstract of a document and that of citation identities, retaining some advantages of both. Compared with title and abstract, it leads to more compact and uniform document representation with a high concentration of indicative words, gives more consistent coupling strengths toprofiles with results agreeing well with the method employing citations, and offers a more consistent ability for inter-group differentiation when the groups are close to each other. Compared with the use citations, it gives results with less specificity and operationally requires an extra step to input and analyse the full citation titles. However, the group profiles derived from title and cited titles are words and can be used toclassify documents that have descriptive abstracts but no or few citations. #1975 #Citedtitles #Titles # Source materials# Extraction# Terms# Technical Services# Information storage and retrieval# Information work# subject indexing# Automatic subject indexing##.

APPENDIX V

INPUT FOR THES.EXE

```
bts(HypertextSoftware',hypercard',notecard',hyperwriter',hyperpad',hypergraph',knowledgepro',authorware
pro',memex',architext',hyperdocument',hyperties',HAM','hypercale'])

syn ('Hypertext Software',['Hypertext package'])

bts ('hypercard',['Hypertalk language']).

bts ('Information Systems',["multimedia information system']).

bts ('User Interface',['graphical user interface','menu driven interface','icon driven
interface','command driven','natural language interface','window driven']).

syn ('User Interface',['front end','gate way']).

syn ('Hypertext Abstract Machine',['HAM']).

bts ('graphics', ['computer graphics','graphical interface'].
```

APPENDIX VI

THES.EXE OUTPUT

3-D GRAPHICS

BT1 : Computer Graphics
Cls under : Computer Graphics
animation
desktop graphics
fractal graphics
morphing
raster graphics

AGGREGATION HIERARCHIES

BT1 : Object Oriented database
Generalisatiion hyerarchies

AMATEUR FILMS

BT1 : motion picture
Cls under : motion picture
animation
cartoon films
puppet films

AMETEUR

BT1 : cinematography
BT2 : animation
BT3 : Computer Graphics
BT3 : motion pictures
Cls under : cinematography
animated cartoon
color
lighting
sound
Stereoscopic

ANALOG RECORDING

BT1 : recording
Cls under : recording
digital recording