DESIDOC Bulletin of Information Technology, Vol. 19, No. 6, Nov 1999, pp. 9-15

Software for Ranking Indian Scientists Based on Citations to Their Publications

SC Saxena, Sonia Relan and SS Murthy

ABSTRACT

The paper describes the software developed at DESIDOC for ranking Indian scientists and technologists based on the citations to their publications as covered in Science Citation Index (SCI). The software provides facilities for generating reports on highly cited authors, highly cited journals, highly productive institutions, weightage of authorship etc. quickly. This measure could be one of the important parameters for rating the performance of scientists for giving awards, promotions or other career incentives and for making appointments to important positions.

1. INTRODUCTION

Citation counting is one of the often used methods of ranking the performance of scientists and technologists in their subject fields. Presently, Science Citation Index (SCI) published by Institute for Scientific Information (ISI), Philadelphia is the best source usually available for carrying out such citation studies but subscription of SCI is very expensive and data is also spread on many CD-ROMs. By using the SCI-CDs, one can find citation counts of different papers and other contributions made by scientists and technologists, and of their institutions and of the journals publishing these contributions. However, the search software provided in the SCI-CDs only facilitates search name after name for finding the citation counts of their publications and for making comparison among them. Such approach is a very time consuming and laborious process. In India, a few institutions like Indian National Scientific Documentation Centre (INSDOC), New Delhi; National Institute of Science, Technology and Developments (NISTADS), New Delhi; Indian National Science Academy (INSA), New Delhi; etc carried out some bibliometric studies and they have also faced the same problems. Since it would be more convenient to have listings of rank-wise contributions of cited authors, institutions and journals, it was decided to develop an application software for this purpose.

For developing a software for quickly and almost automatically getting the desired citation counts, a project proposal was submitted to National Science and Technology Management Information System (NSTMIS), Dept of Science and Technology (DST), New Delhi by Defence Scientific Information and Documentation Centre (DESIDOC), Delhi. The project was sanctioned by NSTMIS to DESIDOC in August 1994. The software was developed as a part-time activity of the authors (in addition to the regular acitivities in their respective jobs) and was completed by the end of 1997 and it was demonstrated in early 1998. The software provides an automated method for determining and listing the authors, and their institutions and the journals publishing their contributions in the order of their ranking in terms of their citation counts.

2. METHODOLOGY

First, a close study was made of the structure of the SCI database and the search facilities available in the SCI search software. Then detailed specifications were drawn for the development of the software which would provide the following functions:

(a) Identify the contributions by Indian scientists covered in the SCI database.

(b) Download the bibliographic details of these contributions.

(c) Arrange the down loaded details in alphabetical order by author, by title, by name of journal, by specific subject, by institution, etc.

(d) Generating the list of maximum cited authors, journals, maximum productive institutions, etc., both subject-wise and across different subject fields in the order of decreasing number of citations.

(e) Append new data from various updated SCI-CD-ROM discs.

(f) Create various indices on various fields like author, title, journal name, corporate source, year, etc. to provide for faster search.

After completing the system analysis, the software was developed in C++, as it is an object-oriented programming language with facility for fast retrieval from large databases. After the development of software, the data from SCI-CDs was down loaded. The data on Indian scientists from SCI-CD was searched by giving "India" as the search term, which is present in the field "corporate author" in the record of SCI. Except this key, there is no way to download all the records on Indian scientists from the SCI at one time. The downloaded records contain the following fields, viz., author, title of article, language, address of the scientists (corporate source), journal name, year of publication, type of document and cited author/reference details. Then using the software developed, the downloaded data was rearranged in the defined database structure and various indices were created on different fields for conducting faster searches on individual fields or by combining two or more fields using Boolean logic operators.

3. LIMITATION OF THE SCI-CD SEARCH SOFTWARE

The following limitations of the SCI-CD search software were observed during downloading the data from SCI-CDs :

(i) Publications of Indian scientists who did not give the addresses of their affiliated organisation in India, could not be downloaded, because India doesn't appear as a search term in the corporate author field.

(ii) Publications of the foreigners who cited Indian scientists' work could not be downloaded because in their addresses also India doesn't appear in the corporate author field.

(iii) Publications of foreign scientists who gave addresses of affiliated organisations in India, were downloaded because India appeared in their addresses.

3.1 Limitations of the SCI-CD for this study

(1) Coverage of Indian contributions in SCI-CD is very low because SCI covers only 4500 journals and out of which the Indian journals are only twelve. For example, a subset of SCI-CD on chemistry citation Index, which was also used for this study, covers only four Indian journals i.e. Indian Journal of Chemistry Section A—Inorganic Bio-Inorganic, Physical Theoretical & Analytical Chemistry, Indian Journal of Chemistry Section B—Organic Chemistry Industry Medicinal Chemistry, Journal of Scientific & Industrial Research and Proceedings of the Indian Academy of Sciences—Chemical Sciences.

(2) The structure of SCI-CD is so complex that it is not possible to break the links among the citations when data is downloaded selectively. All the fields and subfields can not be downloaded separately because they all are inter linked.

(3) Lack of standardisation in some fields. For example, if user wants to search or download records for a particular institution say, Indian Institute of Technology, Delhi, it is not possible from the database to download

because of lack of standardisation in using the names of the institutions and their abbreviations. Secondly, place name of the institutions is not indexed. Because of these problems the correct ranking of the institutions in terms of their productivity become difficult in many cases.

4. SOFTWARE TESTING

The software developed in C++ was tested at DESIDOC by downloading about 9400 records of Indian authors (scientists) from SCI-CD database covering data for the period Jan-June 1993. Listings were created for answering some sample questions as :

(a) who are the most cited Indian authors in physical chemistry, and their ranking?

(b) which are the institutions whose authors(scientists) are most cited and what are the ranks of the institutions?

(c) which are the journals whose papers are most cited and what are the ranks of these journals.

Results of the searches conducted showed good performance of the software and meeting the objectives of the project.

The project was reviewed in February 1997. It was decided in the project Review meeting of NSTIMS that DESIDOC should purchase a subset of science citation index *viz.* Chemistry citation index for further testing the software. DESIDOC purchased this subset in 1997 and downloaded the Indian contributions from 1991 to 1997 from 7 CD discs of chemical citation index by conducting search on corporate source field. The downloaded data was rearranged and searches were conducted. The results of the searches conducted showed good performance on this sample also.

5. IMPORTANT OUTPUTS OF THE SOFTWARE

The software provides facilities for manipulating the downloaded data in different ways and generating various outputs. Some of the important reports generated by the software are as follows:

(i) Ranking scientists, journals, institutions– discipline-wise
   - – highly cited authors (Appendix–2)
   - – highly cited journals (Appendix–3)
   - – highly productive institutions.

(ii) Weightage of authorship.

6. EXPECTED BENEFITS FROM THE SOFTWARE

The software provides a quick and automated method for ranking scientists and technologists for various purposes such as according recognition of their work and giving awards, and other career-based incentives. This software could be distributed by the Department of Science and Technology (DST) who sponsored this project, to the institutions receiving the SCI-CD for providing information on the contributions of Indian scientists, institutions and highly reputed journals, etc. It would also be useful to the scientists and institutions studies and in the mapping of science in India. This software can also be used for providing useful information as mentioned above, even for relating other countries. In addition, the data downloaded can be accumulated over the years and used for many bibliometric studies.

The authors are Scientists in Defence Scientific Information and Documentation Centre (DESIDOC), Metcalfe House, Delhi - 110054.

Science Citations Retrieval System

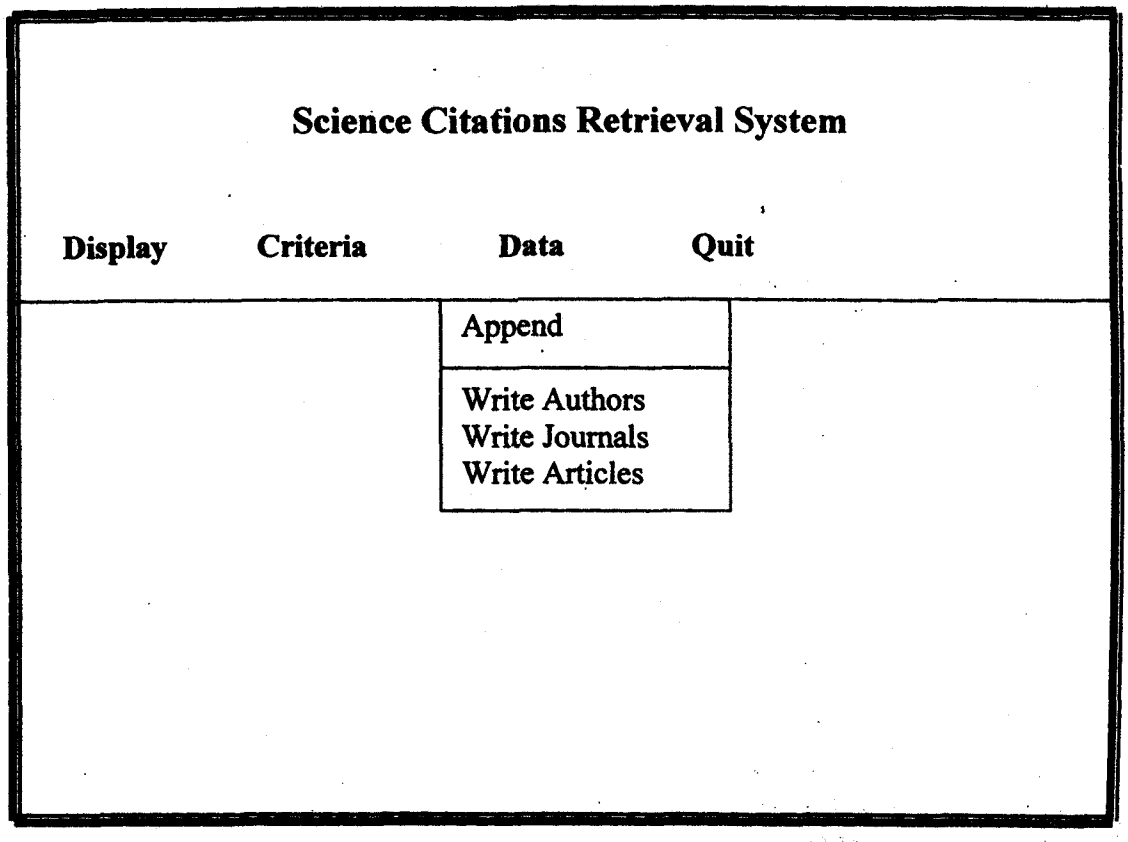
Display **Criteria** **Data** **Quit**

Chronological

Science Citations Retrieval System

Display **Criteria** **Data** **Quit**

Record No.
TGA No.
Title
Language
Author
Address
Journal Name
Year
Document Type
Cited Authors
Combined
Max Cited Ath
Max Cited Jnl
Max Cited Art


Science Citations Retrieval System
Display
Criteria
Data
Quit
Append
Write Authors
Write Journals
Write Articles

Science Citation Index

Highest Cited Authors

First 25 ranks

Citations	Name
537	VOGEL AI
430	SHELDRICK GM
322	KHAN MMT
221	NAKAMOTO K
192	COTTON FA
172	LEVER ABP
152	MEHROTRA RC
145	RAO CNR
139	COREY EJ
134	MATHUR P
130	DEWAR MJS
129	CHOUDHARY VR
118	KUMAR A
115	MEHTA G
113	AMINABHAVI TM
111	JOHNSON CK
110	SINGH K
101	DASH AC
100	BROWN HC
97	PEARSON RG
96	SINGH B
93	BHATTACHARYA S
91	PERRIN DD
90	GEARY WJ
89	FIGGIS BN

Science Citation Index

Highest Cited Journals

First 19 ranks

Citations	Name
8136	J AM CHEM SOC
4206	J CHEM PHYS
3923	INORG CHEM
3811	J PHYS CHEM US
2758	J ORG CHEM
2506	TETRAHEDRON LETT
1990	PHYS REV B
1776	J CHEM SOC CHEM COMM
1466	CHEM PHYS LETT
1370	TETRAHEDRON
1300	INDIAN J CHEM A
1298	PHYS REV LETT
1173	J CHEM SOC
1127	J APPL POLYM SCI
1117	J INDIAN CHEM SOC
1108	NATURE
1090	PHYTOCHEMISTRY
1084	INDIAN J CHEM B
1046	ANAL CHEM