CYBERMetrics '99

Proposal for a research agenda for cybermetric studies

Isidro F. Aguillo
CINDOC-CSIC
isidro@cindoc.csic.es

Seventh International Conference on Scientometrics and Informetrics. Colima, 1999

INTRO

"The aim of this Agenda is to provide an insight of the new possibilities that offers the application of the scientometric and informetric methods to the Internet R&D resources. Special attention is given to obtain new indicators for describing the presence of S&T institutions in the World Wide Web, using citation analysis techniques"

Seventh International Conference on Scientometrics and Informetrics. Colima, 1999

Objects of study

There are three main groups of resources in the Internet for cybermetric studies:

OBJECTS

- **Exchange Fora:** The messages of electronic mail transmitted publicly to multilateral fora: *Usenet News and Lists*
- **R&D Webspace:** The accessible resources of the hypertextual Web
- **"Invisible Internet":** Data bases, electronic journals and virtual communities

Exchange Fora

Promising areas for the quantitative analysis of the of the electronic mail might be:

F
O
R
A

- **Differential use and dynamic evolution of the fora**
  - **Absolute and relative (temporal) evolution: visibility, impact**
  - **Size of the virtual communities**
  - **Analysis of messages circulation**
  - **“listeners” and writers**
  - **Invisible colleges**
- **Informetric description of the models:**
  - **Unidirectional communication versus multidirectional**
  - **Opened versus closed**
  - **One subject versus multidisciplinary fora**
  - **International versus national**
  - **Graphical representations of these models**

FORA

- **Studies of users: Individuals**
  - **personal perception**
    - visibility of researcher in databases (or number of times cited in search engines)
    - quantitative and formal analysis of personal pages (links received)
- **Studies of users: Survey**
  - **active membership**
    - number of fora
    - motivation of the participation (professional or not)
    - participation frequency
  - **percentage of successfully uses**
- **Studies of users: Global**
  - **classification**
    - geographic origin and languages
    - institutional allocation
    - topic of the participation
  - **typology of the contributions**
    - “noise": impact, typology and respect of conduct codes

R&D Webspace

WEBESPACE

The quantitative description of the processes of scientific communication in the Web covers:

- **Identification and analysis of the institutional presence of the organisations, public and private, dedicated to the R&D in the Web**
  - **Creating data bases describing this presence, providing data in a way that allow a comparative analysis with the situation outside Internet**
- **Study of the production and scientific productivity of institutions, groups and researchers through publications in the Web**
  - **Building Reports that reflect by means of quantitative indicators the quality, visibility and impact of the scientific activity**

WEBESPACE

- **Illustration of the relationships and evolution of contiguous disciplines and the multidisciplinary studies**
  - Producing “maps” from the multidimensional analyses, reducing to main components and grouping correspondences...
  - Calculating “distances” and “relationships” among disciplines
- **Discovery of the co-operation links between different scientific teams, at national and international level**
  - Creating tables that show the exchanged connections to make analysis of this co-operation
- **Description of the mechanisms of scientific communication in the Web**
  - Defining tools and procedures for periodic automatic archiving of websites and
  - Analysing their dynamics and evolution
    - relating the results to the introduction or adoption of new S&T methods or techniques

WEBESPACE

- **Evaluation of the patterns derived from the use of the Web by the researchers**
  - **Making surveys of representative groups to know how the information sources are used**
  - **Explaining the patterns of consumption of the webpages**
    - "hits" (visits received) analysis
- **To undertake the study of the impact of the Web R&D information in other areas**
  - **Describing the patterns of usage of the STM information by other groups not directly involved with research**
    - Economic and industrial impact
    - Political impact
    - Cultural impact
    - Social (citizen oriented) impact
  - **The role of the academic and research part of the Web in the spreading and transmission of the scientific and technical knowledge**

WEBESPACE

- **Debate on the impact of the different institutions in the scientific progress**
  - **Evaluating the presence and contribution from different actors in the generation processes and transmission of I+D results**
    - academic organisations (Universities)
    - research institutes
    - specialised organisations (Hospitals)
    - support centres (Libraries and centres of documentation)
    - publishers
    - producers and/or suppliers of electronic information goods and services (especially STM data bases)
    - Managers and executors of R&D policy
    - other local, regional, national or international public bodies
    - private sector, such as industrial companies (pharmaceutical, petrochemical, electronics) and SME (information brokers)
    - Scientific societies, associations and foundations)
    - Media (specially those in electronic format)

W E B E S P A C E

- **Analysis of the formal aspects of the STM websites**
  - Describing the physical and logical size and the coverage of these sites
  - Evaluating their formal quality in key aspects like authors and responsibility, regularity, legibility
  - Calculating their hipertextuality in terms of richness, density and diversity of links
  - Quantifying the objects and the multimedia character (number, type and size of the files)
- **Correlation of these results with the analysis of the contents**
  - Considering the impact and the penetration of the metadata
    - individual elements
    - different schemes
  - Analysing the occurrences and co-occurrences
  - Comparing the different graphical methods used for the description and the internal search or navigation systems

WEBESPACE

- **Relevance of the information recovery systems in relation to the consumption and needs of the scientific activity**
  - **Studying the general and specialised indexes**
    - methods used in its compilation
    - organisation and taxonomies
  - **Testing the flexibility, power, speed, precision, coverage, update and overlapping of the different search engines**
  - **Emphasising the advantages and disadvantages of the use of multi- and metasearchers**
- **Design of new information products of interest for the scientific community**
  - **Developing new criteria for quality evaluation of resources using quantitative indicators**
  - **Designing new automatic/semiautomatic methods for the building of indicators using the main characteristics of the Web**
    - global/hypertext network

W E B E S P A C E

- **Study of the quality evaluation (including "peer-review") processes by means of the use of the citation analysis in the Web**
  - Assuming that the processes that lead to the establishment of a hypertextual link, *in academic environment*, are similar to those giving rise to the bibliographical citation process
  - Recognising that the methods of the scientometric analysis based on citation are also valid in the STM Webuniverse for the cybermetric hypertextual analysis
  - Studying the links made from conventional publications to reviewed and not-reviewed electronic resources, analysing in the first case their importance on the calculation of the bibliometric indicators
  - Generating new ratios and indices using hypertextual links following the model of the Science Citation Index
  - New reports offering comparative results from both techniques and possible merged indexes combining data
  - Publishing these results in several electronic formats

"Invisible Internet"

A group of largely heterogeneous items requiring different methods for their quantitative analysis:

- **Scientific publications in regulated formats**
  - **Electronic journals**
- **Other mechanisms of access to scientific information**
  - **Repositories of "preprints" science**
  - **Archives of computer programs and files**
  - **Patents**
  - **Graphical representations, including dynamic and/or the interactive ones**
- **Virtual digital libraries formed by groups of records:**
  - **Bibliographics**
    - impact of Z39.50
  - **Electronic full texts (semantic and syntactic analyses)**
  - **Other alphanumeric databases**

I N F R A N E T

I N F R A N E T

- **New habits of communication and publication**
  - **New sources**
    - institutional versus personal pages
  - **Quality**
    - formal publication: The future of peer review
    - informal publication: Pre-prints repositories
  - **Citation patterns**
    - relationships between paper and electronic formats
  - **Other sources of "irregularity" in the e-world**
    - Formal aspects
    - Volatility versus final versions (updating and revisions not allowed)
  - **Globalisation**
- **From the the virtual communities to the Web Portals**
  - **Structure and composition of the virtual communities**
    - Horizontal versus Vertical Communities
    - Dynamics: Coverage and exhaustiveness
    - Frequency of updates

QUESTIONS?

"Cybermetrics, International Journal of Scientometrics, Informetrics and Bibliometrics"

http://www.cindoc.csic.es/cybermetrics/