Parallel Session 16: Are Internet expectations being accomplished?

CONTENT AND LINK NETWORK ANALYSIS OF FIVE SCIENCE COMMUNICATION WEBSITES IN LATIN AMERICA

Fabio Castro Gouveia[1] & Eleonora Kurtenbach[2]

[1]*Museu da Vida – Fundação Oswaldo Cruz – Brazil. Av. Brasil, 4365, 21045-900 - Rio de Janeiro, RJ, Brazil. Tel: +55-21-3865-2103, Fax: +55-21-3865-2121, E-mail: fgouveia@coc.fiocruz.br*

[2]*Departamento de Bioquímica Médica – ICB – UFRJ – Brazil. Universidade Federal do Rio de Janeiro – Departamento de Bioquímica Médica/ICB/CCS, 21941-590 - Cidade Universitária, RJ, Brazil. Tel: +55-21-2562-6753, Fax: +55-21-2270-8647, E-mail: kurten@bioqmed.ufrj.br*

Abstract

One of today's most famous search engine (Google®) is primarily based on the PageRank® methodology described by Brin and Page (1998). As part of this methodology of indexing pages there are some features that could be of great help for those looking to evaluate the impact of a website over the Internet. The links pointing to a particular URL (link network) can be obtained with a simple query search. Here we present an analysis of the link network of five Latin American science communication websites, exploring the applications of this methodology and what to remind of during this evaluation.

Keywords: Internet; Link Network; Search Engine

Introduction

One of the first things we teach to a new Internet user is how to search and find specific information. It is known that over 85% of Internet traffic is driven directly or indirectly by search sites. Each one of these sites uses a methodology of indexation, and we can divide them into two categories: directories and search engines (Hu *et al.*, 2001).

Search engines create their database automatically using "spider" programs. The resulting database is then indexed using an algorithm to sort the search results. After the publication of Page & Brin (1998), their search engine (Google) has become the most used worldwide.

Google uses an algorithm called PageRank in order to sort its results. It is based on the fact that the more one site is referenced by other sites, the greater its relative relevance. This algorithm greatly improved the accuracy of search results.

Method

The database that Google has can be accessed using special search tags. One of them is the "link:" tag that shows the list of pages in the database that point to a particular web address (link network).

To obtain that the steps should be as follows:

- Open a Google session

- Set the preferences of *number of results* to the maximum (100)

- Do the search using the tag *link*: + the desired URL

- Save each of the pages

- Consolidate the pages in one document

In general the links pointing to a particular website can be placed in one or more of the categories listed in Table 1.

Type of Link	Comment
Links from the same site or from the same institution	Structural – to be discarded from the evaluation
Links from institutions, associations or networks	Relation to other institutions on the web
Links from news or from events hold or supported	Ability to communicate new activities over the web
Links from visitors or groups that enjoy the content of the site	Result from the strategy of the website
Links from directories and link lists	Increase visibility and access
Links from documents or references	Documents on the web or personal pages

Table 1 – Types o links pointing to a website

We selected five websites of institutions involved in science communication in Latin America to perform a link network analysis. They were arbitrarily chosen as examples for this study. Briefly the results obtained are listed in the next session.

They will be better explored during the open plenary with other general information.

Results

>Fundación CIENTEC, in Costa Rica, had a list of 98 links from which 40 were from the same site. Besides that it had a good list of institutions linking to the website and some from special events http://www.cientec.or.cr/

>Maloka, a Science Centre in Colombia, had a total of 100 links of which 70% were from list of links or directories. The website is very well indexed and have links from visitors which is very important for a science centre website http://www.maloka.org

>Papalote, a children museum in Mexico, presented a similar amount of links (92) and a similar link network result from Maloka http://www.papalote.org.mx/

>Programa Explora, a science popularisation program in Chile, had a total of 101 links, but half of them were from the same site or institution. The rest of them were from events and directories http://www.explora.cl/

>RedPOP (Science Popularisation Network from Latin America and the Caribbean) from UNESCO had the modest link network of all five sites. With only twenty links from which ten had to be discarded, the rest were from directories and supported events. For a website of a network that recently promoted a congress and have several institutions this can be considered a really modest result http://www.redpop.org/

Conclusions

We conducted the present study with five arbitrarily selected websites as examples for a link network analysis. We could observe that some of these sites were not referenced by their potential community while others have a good link network.

For Latin America standards, a list of about 100 to 200 links for a site dealing with science communication is quite common. Different languages represent different communities, and nearly the same happens about different countries. Contents in English are more capable of producing a greater network, since they are the majority of pages on-line. Search Engines can't index all the pages on the Web, so the link network obtained is a subset of the real link network on the Internet.

We think that link network retrieval should be part of the routine evaluation of a website.

References

Brin, S & L. Page (1998). The Anatomy of a Large-Scale Hypertextual Web Search Engine. Computer Networks and ISDN Systems, 30, 107-117.

Hu, W.C., Y. Chen, M.S. Schmalz, & G.X. Ritter. (2001). An overview of the World Wide Web search technologies, In: *Proceedings of 5th World Multi-conference on System, Cybernetics and Informatics* (pp. 356-361), Orlando, Florida.

4