R. Poonkothai & Dr. A. Ganesan

A COMPARATIVE STUDY BETWEEN SUBJECT GATEWAYS AND SEARCH ENGINES

R.Poonkothai, * Dr.A.Gansean

Abstract

The Web is quickly becoming the world’s fastest growing repository of data. People are increasingly going to the Internet before they go to the library. Subject gateways doing for Internet information resources what librarians do for books. Using a subject gateway instead of a general search engine can result in the return of more genuine and relevant web pages from our search. They have been particularly cited as a helpful starting-point for Internet beginners to explore the resources available in a particular subject.

Keywords: Subject Gateways, Search Engines, Internet

INTRODUCTION

Learners and researchers have a potentially vast range of resources available to them through the Internet and associated technologies. But, as the Internet increases in size and complexity, so do the associated usability and navigational issues. A range of strategies has been used to manage online information and provide different searching protocols, navigational aids and maps. Nonetheless, information overload persists. This information overload problem is well recognized and a number of structured resources and environments attempt to address it, such as search engines, information gateways and portals, digital libraries and virtual learning environments. In reality, users may need to adopt more of a ‘mix and match’ strategy, selecting individual features from relevant sources to meet their specific needs.

What is Search Engine?

Search engines are the card catalogs of the web. They search for keywords by using "spiders" or "Robots" (electronic software) that crawl through web space from link to link, collecting and indexing the words on millions of web pages and feeding them into searchable databases. Search Engines for the general web do not really search the World Wide Web directly. Each one searches a database of the full text of web pages selected from the billions of web pages out there residing on servers. When we search the web using a search engine, we are always searching a somewhat stale copy of the real web page. When we click on links provided in a search engine's search results, we retrieve from the server the current version of the page.

If a web page is never linked to in any other page, search engine spiders cannot find it. The only way a brand new page - one that no other page has ever linked to - can get into a search engine is for its URL to be sent by some human to the search engine companies as a request that the new page be included. All search engine companies offer ways to do this.

What is subject Gateways?

The Web is quickly becoming the world’s fastest growing repository of data. People are increasingly going to
the Internet before they go to the library. Subject gateways “doing for Internet information resources what librarians do for books”. An Internet search tool to help people find information on the Internet e.g.: electronic journals software datasets electronic books mailing lists / discussion groups (and their archives) articles / papers. Subject Gateways are Internet services which support systematic discovery of high quality information provide links to resources accessible via the Internet, are based on resource descriptions provide browsing and searching access to the resources via a subject structure The key ingredient ...the semantics that only the human factor can bring: subject specialists library / information professionals.

Common gateway features
1. The host institutions of each gateway have a mandate to be information providers, but they are not required to be information creators.
2. Each gateway has a distinctive logo and a reflective name, thereby positioning them for a significant Web presence. Distinctive branding also bears fruit as a common recognition technique.
3. Each gateway has selected and utilized a standard metadata schema for describing the resources incorporated into the gateways. The metadata is applied by librarians or educators with experienced knowledge of the disciplines.
4. The resources in each gateway have all been selected according to pre-determined criteria, published at each gateway site as part of a content coverage policy. A gatekeeper function, often a mix of computer and human intervention, ensures adherence to the selection criteria.
5. The coverage policies have made electronic resources, both 'born digital' and digitised, the highest priority for inclusion in the gateways; The resources are supported by similar architectures, based on distributed creation and maintenance of their metadata, with a centralised facility for access.

Why gateways better than search engine?
Using a subject gateway instead of a general search engine can result in the return of more genuine and relevant web pages from our search. If we do wish to use a search engine, then we will have to think more carefully about appraising the information we retrieve from the web before we can rely on it as a basis for our study or research. Subject gateways are portals to web pages in a particular subject area. They usually provide a search interface which allows us to find relevant web pages. An added advantage is that they usually link to web pages which are of reliable quality.

Search engines, like the Web, are ever changing. What holds true today for a search engine may not do so in the future. A judgment at a point in time may not hold true for very long in the changing environment that is the World Wide Web. Search engines do not index all the documents on the Web, nor do they all index the same sites or in the same fashion.
A good search robot combined with full-text indexing but poor retrieval software will give rise to high recall and low precision. The main criticism of automated tools is that they tend to overload users with irrelevant, misleading results. Documents which share the same relevant words but not necessarily shared relevant context.

The key difference between subject gateways and the popular automated large-scale Web indexing systems such as Google is the quality of the results which the end-user receives. This is dependent on the nature of the cataloguing process. For example, we searched Google and OMNI on the term "epilepsy" on June 2007; from the results, we observed the descriptions of the first two hits:

**Comparative study of Search Engine and Subject gateway**

<table>
<thead>
<tr>
<th>Google</th>
<th>OMNI (Organising Medical Networked Information)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Epilepsy is a common chronic neurological disorder that is characterized by recurrent unprovoked seizures. [1] [2] These seizures are transient signs and/or ...</td>
<td>1. Surgery for epilepsy</td>
</tr>
<tr>
<td>1. In-depth epilepsy and seizure information for patients, families, and the professionals who care for them.</td>
<td>This leaflet, aimed at parents, provides information about epilepsy, and how it can be helped by surgery. It explains who would be considered for surgery, tests and scans that would be required, types of operation, risks, the success rate, and follow up treatment. Published on the Web by Great Ormond Street Children's Hospital (GOSH) and the Institute of Child Health (ICH).</td>
</tr>
<tr>
<td>2. In-depth epilepsy and seizure information for patients, families, and the professionals who care for them.</td>
<td>2. Improving services for people with epilepsy: Department of Health action plan in response to the national clinical audit of epilepsy-related death. This document Improving services for people with epilepsy: Department of Health action plan in response to the National Clinical Audit of epilepsy-related death was published by the Department of Health (DH) in February 2003. This action plans aims to improve standards of care and support for people with epilepsy and their families. It is hoped that this will be the start of a process of improving awareness and understanding of Sudden Unexpected Death in Epilepsy.</td>
</tr>
</tbody>
</table>

Here, we can see that the resources catalogued in OMNI have been described in a "human readable" fashion, whilst the entries in Google are presented more as "raw data". OMNI points directly to the home page or start point of a resource while Google often points to pages without context, leaving the user to find their own way. Google indexes individual *pages*, not *resources*. As an illustration of the difference between a page and a resource, consider that an online textbook could consist of many web pages, hyperlinked together via a table of
A Comparative study between Subject Gateways and Search Engines

Contents. Subject gateways such as OMNI, on the other hand, catalogue at the resource level, and will therefore describe resource composed of many pages in a much more coherent fashion.

Lastly, the resources described in a subject gateway are likely to have been hand-picked and catalogued with a particular audience in mind. The OMNI gateway, for example, includes only biomedical resources of interest to the higher education and research community, and catalogues resources with this in mind. Thus, resources included in OMNI are indexed and classified in a similar way to books in an academic medical library, and are selected so that they are at an appropriate level for students, researchers, lecturers, etc. This tailored and selective approach is not possible for a service such as Google, which successfully serves a much broader community.

CONCLUSION

Subject gateways are undoubtedly very useful Internet search tools. While they are more limited in the extent of their coverage than search engines, they are valued by users for the higher quality and relevance of their results. They have been particularly cited as a helpful starting-point for Internet beginners to explore the resources available in a particular subject. The subject gateways have established a new trend in collaboration, which can only revitalise library service provision in the future.

REFERENCES
4. The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture by John Battelle
5. Evaluate Search Engines: http://www.clubi.ie/webserch/search_tools/index.htm/