Search engine user behaviour: How can users be guided to quality content?

Dirk Lewandowski
University of Applied Sciences Hamburg
dirk.lewandowski@haw-hamburg.de

The Bloomsbury Conference: a joint EUSIDIC - CLSIG Conference
London, 31 March 2008
## Agenda

<table>
<thead>
<tr>
<th>Situation</th>
</tr>
</thead>
<tbody>
<tr>
<td>User behaviour in Web search engines</td>
</tr>
<tr>
<td>Quality content in search engines</td>
</tr>
<tr>
<td>Conclusion: What libraries can learn from search engines</td>
</tr>
</tbody>
</table>
## Agenda

### Situation

- User behaviour in Web search engines
- Quality content in search engines
- Conclusion: What libraries can learn from search engines
Situation

- **Web context**
  - Search engines dominate web searching
  - Major problem: Relevance.

- **Scientific context**
  - Students use general Web search engines for finding scientific content.
  - Search engine vendors created scientific search engines
    - Google Scholar
    - Windows Live Academic
    - Scirus
  - Where do libraries (and their information sources) stand?
<table>
<thead>
<tr>
<th>Agenda</th>
</tr>
</thead>
<tbody>
<tr>
<td>Situation</td>
</tr>
<tr>
<td><strong>User behaviour in Web search engines</strong></td>
</tr>
<tr>
<td>Quality content in search engines</td>
</tr>
<tr>
<td>Conclusion: What libraries can learn from search engines</td>
</tr>
</tbody>
</table>
Users use relatively few cognitive resources in web searching.

<table>
<thead>
<tr>
<th>Queries</th>
</tr>
</thead>
<tbody>
<tr>
<td>– average length: 1.7 words (German language queries; English language queries slightly longer)</td>
</tr>
<tr>
<td>– Approx. 50 percent of queries consist of just one word</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Search engine results pages (SERPs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>– 80 percent of users view no more than the first results page (10 results)</td>
</tr>
<tr>
<td>– Users normally only view the first few results („above the fold“)</td>
</tr>
<tr>
<td>– Users only view up to five results per session</td>
</tr>
<tr>
<td>– Session length is less than 15 minutes</td>
</tr>
</tbody>
</table>

- Users are usually satisfied with the results given.
- Users expect that all information systems to be as easy as Google (and to produce good results for their queries, too).
Search engines deal with different query types.

**Query types (Broder, 2002):**

- **Informational**
  - Looking for information on a certain topic
  - User wants to view a few relevant pages

- **Navigational**
  - Looking for a (known) homepage
  - User wants to navigate to this homepage, only one relevant result

- **Transactional**
  - Looking for a website to complete a transaction
  - One or more relevant results
  - Transaction can be purchasing a product, downloading a file, etc.
<table>
<thead>
<tr>
<th>Agenda</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Situation</strong></td>
</tr>
<tr>
<td>User behaviour in Web search engines</td>
</tr>
<tr>
<td><strong>Quality content in search engines</strong></td>
</tr>
<tr>
<td>Conclusion: What libraries can learn from search engines</td>
</tr>
</tbody>
</table>
Search engines *do* rate web pages for quality.

- **There’s more than just text matching**
  - Link popularity algorithms (PageRank, HITS, etc.)
  - Page freshness
  - User location

- **Adjusting the results to the query**
  - Broad queries are best answered with broad pages, specific queries are best answered with specific information.
  - There are “must have” pages for some queries.

- **Problem: All these work on a per page basis only.**
Search engines offer more than a results list to a query.

**Composition of a typical SERP**

- **Web results**
  - Gordon Brown - www.dealline.co.uk
    - Find, compare & buy. Browse reviews on 1,000s of products.
  - Top news articles for gordon brown
    - Is this useful? Yes | No
    - Gordon Brown's Carla Bruni kiss
    - Nicolas Sarkozy and Gordon Brown poised...
    - President Sarkozy embarrasses Gordon ...
  - Biography of the Rt Hon Gordon Brown MP, Prime Minister and First Lord ...
    - Biography of the Rt Hon Gordon Brown MP, Prime Minister and First Lord of the Treasury. Mr Brown was born in 1951 and educated at Kirkcaldy High School and Edinburgh University ...
    - www.number-10.gov.uk/output/Page12037.asp · Cached page
  - Gordon Brown Associates - Chester-le-street estate agents, Low Fell ...
    - Solicitor specialising in property, financial and legal services. Includes property search, details of services and contact information.
    - www.gordon-brown.co.uk · Cached page

- **Additional databases**
  - Images, Video, News, Maps, MSN, More

- **Sponsored results** (ads)
- **One box results** (e.g., news)
- **Web results**
Search engines offer additional content from specialised web databases.

- **Additional databases**
  - News
  - Video
  - Blogs
  - Q&A
  - Books
  - Scientific
  - ...

So what is quality content in search engines?

- **Authoritative web pages**
  - Official sites (e.g., biography of George W. Bush on whitehouse.gov)
  - Topical authorities (e.g., IMDB.com for films)
  
  From regular web crawl, results are part of the web results list.

- **Relevant documents from additional databases**
  - News, video, etc.
  
  Additional databases are created through focused crawling.

- **Pointers to external databases**
  - Content itself is not indexed by the search engines
“Universal Search“ ranks results from all databases into the web results list.
**Agenda**

<table>
<thead>
<tr>
<th>Situation</th>
</tr>
</thead>
<tbody>
<tr>
<td>User behaviour in Web search engines</td>
</tr>
<tr>
<td>Quality content in search engines</td>
</tr>
</tbody>
</table>

**Conclusion: What libraries can learn from search engines**
Libraries and information vendors can learn from web search engines.

- New library search solutions seem to focus on Web 2.0, but the search problem is still not solved!
- The libraries’ problems with different collections and databases are similar to the search engines with their collections.
- It’s not only about offering quality content, but also about guiding the user to it.
- Users need one access point to all the content offered.
- Results from all the library’s collections and databases should be shown within one results list.
- Therefore, the OPAC must become a scientific search engine.
- This goal can only be achieved with libraries, database vendors, and OPAC vendors working together.
Thank you for your attention.

Prof. Dr.
Dirk Lewandowski

Hamburg University of Applied Sciences
Department Information
Berliner Tor 5
D - 20099 Hamburg
Germany

www.durchdenken.de/lewandowski
E-Mail: dirk.lewandowski@haw-hamburg.de