Search engines: Google alternatives and why you would want to use them

Dr. Nuria Plattner

UNIVERSITY LIBRARY BERN
Nuria Plattner, Aline Frank, Michael Horn, and Silvan Christen
www.unibe.ch/ub/sciencelibrary
The Internet - an infinite universe

Travel through the world wide web

- The internet can be described as a huge collection of unstructured information
- Communication and information through the internet is enabled by unified protocols
- The information in the internet is only valuable if it can be searched efficiently

@ internet-map.net: site sizes proportional to traffic
Searching the internet

Search engine working principles

- WWW can be seen as huge database containing URLs
- Each URL contains a web page with its content
- Naïve content search: just search for a given content/keyword in all pages
- A search engine in contrast builds its own database (page index) of webpage content by crawling the web
- Relevant content for a given search is then identified and ranked based on algorithms

Symbolbild: pikrepo.com
Searching the internet

The page rank algorithm

- Requirement: algorithm able to identify the most relevant search results
- Manual indexing not scalable
- Basic idea: page ranking based on links between pages
- Basic ranking procedure of Google Pagerank

$$PR_x = \sum_{v=0}^{V} \frac{\text{links on page } v \text{ to } x}{\text{links on page } v}$$

Google architecture 1998 [1]

Searching the internet

Where does a search engine search?

- Search engines only provide results of pages listed in their index
- Some webpage types can not be crawled
- Many pages are not being included in the index, examples:
  - company internal networks
  - non-public databases
  - pages with no links to them
  - censored content (search engine and country dependent)
  - …

@ https://phys.org/news/2015-05-deep-web-scientists.html
Search engine biases

Are the search results engine-dependent?

• Results between search engines can differ substantially due to various reasons:
  • Efficiency and algorithms for indexing
  • Choices what to include in index
  • Algorithms used to rank content
  • Personalization: results not only differing between search engines, but also between users using the same engine

Filter bubble symbolic picture: https://www.flickr.com/photos/gforsythe
Search engines and personalized data

Search results, privacy and filter bubbles

- The use of personalized data can improve search results
- Personalized searches can lead to filter bubbles
- Personalized data can be used for other purposes: privacy and formation of opinion at risk

Search engine examples:

**collecting your data:**

- Google: www.google.com
- Bing: www.bing.com
- Yahoo: www.yahoo.com

**not collecting your data:**

- DuckDuckGo: https://duckduckgo.com
- Qwant: www.qwant.com
- Swisscows: https://swisscows.com
Consequences of search engine ranking

Example: Election outcomes

• Results of five double-blind experiments, using a total of 4,556 undecided voters

• Voters represent diverse demographics in the United States and India

• Results show that biased search rankings can shift the voting preferences of undecided voters by 20% or more

Clicks on search results and time allocated to web pages as a function of search result rank, aggregated across three experiments[1]

Evaluation of search engine bias

Example: search term suggestion

- Evaluation of Google, Bing and DuckDuckGo query suggestions over a period of 4 months for 629 German politician names
- Analysis of search term suggestions reveals biases dependent on search engine

Venn diagram with the sizes of the sets of overlapping and unique search terms for each of the three search engines

Query suggestion examples: a) Bing b) DuckDuckGo
# The search engine zoo

## General-purpose search engines

<table>
<thead>
<tr>
<th>Search Engine</th>
<th>Website</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google</td>
<td><a href="http://www.google.com">www.google.com</a></td>
</tr>
<tr>
<td>Bing</td>
<td><a href="http://www.bing.com">www.bing.com</a></td>
</tr>
<tr>
<td>Yahoo</td>
<td><a href="http://www.yahoo.com">www.yahoo.com</a></td>
</tr>
<tr>
<td>Baidu</td>
<td><a href="http://www.baidu.com">www.baidu.com</a></td>
</tr>
<tr>
<td></td>
<td><a href="http://www.baiduenglish.com">www.baiduenglish.com</a></td>
</tr>
<tr>
<td>Yandex</td>
<td><a href="http://www.yandex.com">www.yandex.com</a></td>
</tr>
<tr>
<td>Meta search engine:</td>
<td><a href="http://www.dogpile.com">www.dogpile.com</a></td>
</tr>
<tr>
<td></td>
<td><a href="http://www.searchencrypt.com">www.searchencrypt.com</a></td>
</tr>
<tr>
<td></td>
<td><a href="http://www.gibiru.com">www.gibiru.com</a></td>
</tr>
<tr>
<td>Google without tracking:</td>
<td><a href="http://www.startpage.com">www.startpage.com</a></td>
</tr>
<tr>
<td>DuckDuckGo</td>
<td><a href="https://duckduckgo.com">https://duckduckgo.com</a></td>
</tr>
<tr>
<td>Qwant</td>
<td><a href="http://www.qwant.com">www.qwant.com</a></td>
</tr>
<tr>
<td>Swisscows</td>
<td><a href="https://swisscows.com">https://swisscows.com</a></td>
</tr>
<tr>
<td></td>
<td><a href="http://www.yandex.com">www.yandex.com</a></td>
</tr>
</tbody>
</table>
Special-purpose search engines

Scientific search and useful information

www.wolframalpha.com
provides historical information
by date, unit conversions,
stock data e.t.c. It contains
examples by topics WolframAlpha

search content (images, videos,
music...) you can reuse under a
creative commons licence
https://search.creativecommons.org

Search for scientific articles and authors

Google Scholar
https://scholar.google.com

Microsoft Academic
https://academic.microsoft.com

CiteSeerX
http://citeseerx.ist.psu.edu/index
Discussion / Conclusions

Web search engines

• Web search engines build their own database (index) of webpages and their content by crawling the web

• Web pages are ranked based on algorithms which in some cases use personalized data

• Various search engines exist, some protecting user privacy

• There are general-purpose and specific search engines, e.g. for scientific content

• Biased content and filter bubbles can be the result of using a single search engine; the use of different search engines is advised
Thanks for your attention

Questions?