



Search Engines

Information Technology and Social Life

March 2, 2005

Ask difference between a search engine and a directory



Search Engine History

- A search engine is a program designed to help find files stored on a computer, for example a public server on the World Wide Web, or one's own computer. The search engine allows one to ask for media content meeting specific criteria (typically those containing a given word or phrase) and retrieving a list of files that match those criteria. (Wikipedia)
- A search directory is a directory on the Web that specializes in linking to other web sites and categorizing those links. Web directories often allow site owners to submit their site for inclusion. editors review submissions for fitness.
- Primarily a phenomenon of the Web; Archie and Veronica for FTP and Gopher
- Early search engines were lists or collections of links
- Lycos - 1st commercial endeavor 1994
- WebCrawler, Hotbot, Excite, Infoseek, Inktomi, AltaVista, Ask Jeeves
- InfoPeople Search Tools Chart - <http://www.infopeople.org/search/chart.html>
- How Search Engines Work - <http://www.learnthenet.com/english/animate/search.html>



Yahoo

- Stanford grad students David Filo and Jerry Yang
- Headquartered in Sunnyvale, CA
- Yet Another Hierarchical Official Oracle
- Started in mid-90s; IPO April 1996
- Addition of mail, instant messaging, Web hosting, etc.
- Yahoo-originated as a directory, later added search engine functionality - used Google technology until Feb. 2004
- 2002 bought Inktomi, 2003 acquired company that owned AltaVista and AlltheWeb.
- 3 billion page views per day



Google

- Larry Page and Sergey Brin - Stanford students - 1996
- Company founded in 1998; headquartered Mountain View, CA
- Named as a derivation of Googol, a 1 with 100 zero's after it.
- Originally named BackRub - checked back links
- Link popularity and Page Rank
- Eric Schmidt later joined as CEO (worked for Novell and Sun)
- IPO- August 2004, Internet Auction; \$85 per share, currently \$188
- Many new features in works, News, Images, Scholar, Gmail, etc. - employees can spend up to 20% of their time working on new products
- Owns Blogger
- 2004 - handled 80% of all search requests
- Philosophy - don't be evil
- Google turned a profit every year since 2001 and earned a profit of \$105.6 million on revenues of \$961.8 million during 2003.
- Microsoft increasing efforts for Web search at msn.com

Page Rank - www.google.com/technology - PageRank relies on the uniquely democratic nature of the web by using its vast link structure as an indicator of an individual page's value. In essence, Google interprets a link from page A to page B as a vote, by page A, for page B. But, Google looks at more than the sheer volume of votes, or links a page receives; it also analyzes the page that casts the vote. Votes cast by pages that are themselves "important" weigh more heavily and help to make other pages "important."

Important, high-quality sites receive a higher PageRank, which Google remembers each time it conducts a search. Of course, important pages mean nothing to you if they don't match your query. So, Google combines PageRank with sophisticated text-matching techniques to find pages that are both important and relevant to your search. Google goes far beyond the number of times a term appears on a page and examines all aspects of the page's content (and the content of the pages linking to it) to determine if it's a good match for your query.



Pew Search Engine Report

- 84% of Internet users use search engines
- 92% confident with their searching ability
- 68% say search engines are fair; 19% don't think so
- 44% say they only use one search engine
- 62% unaware of paid vs. unpaid results distinction
- More than half of searchers do so for fun as well as important things



Pew Center Search Report

- More men than women use search engines (88% vs. 79%); 40% of men search daily, only 27% of women
- Men more confident about searching abilities than women; more men know about paid/unpaid distinction
- Younger users more likely to use search engines (89% under 30); 67% over 65
- Younger users are very confident in their search skills