



An Introduction to Search Engine Optimization

Carnegie Mellon University
Pittsburgh, PA
October 12, 2009

Presented by
Mary Ecsedy
Managing Partner
Circuit Riders, LLC
Pittsburgh, PA
412.422.1611
www.circuit-riders.net



*“Information That Cannot Be Located
Might As Well Not Exist.”*

- Nancy Mulvaney
Editing Books



Outline

- I. Information Storage and Retrieval – an Ancient Problem**
- II. Brief History of the Web and Search Engines**
- III. SEO Basics**
- IV. Getting Started**
- V. The Business of SEO/SEM**
- VI. SEO in the Real World – Case Studies**
- VII. Future Trends**
- VIII. Question & Answer Period**



Information Storage and Retrieval – an Ancient Problem

- In the beginning, human knowledge was stored in oral and visual culture, within reach of individual memory.
- c. 4,000 years BCE: Cuneiform writing was invented by the Sumerians. They immediately ran into storage and retrieval problems with uncounted numbers of clay tablets, and developed meta-data lists to help them keep track of important texts.
- c. 300 BCE: The Library of Alexandria was founded, eventually accumulating several tens – or hundreds – of thousands of scrolls. The organization of the material in the Alexandrian library was apparently modeled on Aristotle's own private library, which is considered the first systematically organized research library.



Information Storage and Retrieval – an Ancient Problem

- c. 800 AD: Library catalogs are introduced in the House of Wisdom - books were organized into specific genres and categories, making it easier to find texts on a specific subject.
- c. 1789: The first cards used for catalogs occurred in France during the revolution. Books and manuscripts were confiscated and brought to literary depots at several locations in Paris. The staff at each depot recorded on cards the basic information about each item. These cards were then bound up in bundles and sent to the Paris Bureau de Bibliographie.
- Library card catalogs were originally hand-written cards. These were replaced by typewritten cards, then machine-printed cards, and finally replaced by online catalogs in the 1980's and 90's.



Circuit Riders, LLC

Information Storage and Retrieval – an Ancient Problem



Library of Congress, 19th century – books waiting to be cataloged...



Information Storage and Retrieval – an Ancient Problem

- Computing and digital technologies have helped, but they've also made the problem of storing and retrieving information infinitely worse by creating so much of it, and in such a viral format.
- It's easier to create a million digital copies of a book than to transcribe one illuminated manuscript by hand. So not only do we have more data, but we have more bad data.
- Sorting through all the junk to find information that is useful, authoritative, and relevant is one of the challenges of our times on both a personal and a global level.



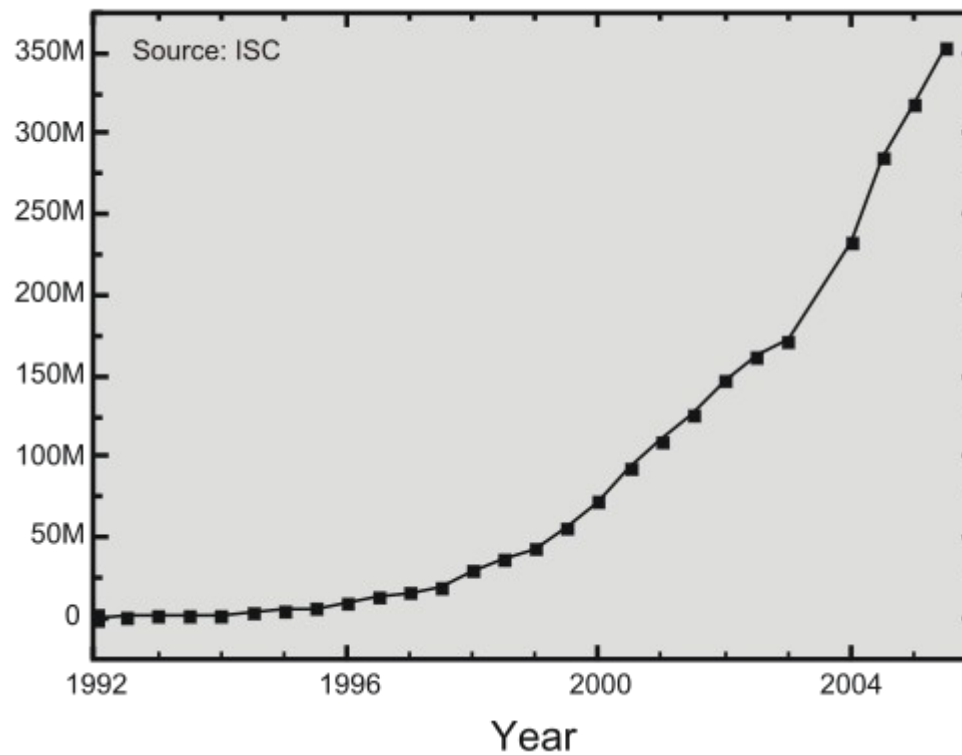
The Internet and the World Wide Web

- 1991. The list of web servers edited by Tim Berners-Lee and hosted on the CERN webserver included about 30 servers.
- January 1995. IBM's Superbowl commercial introduced the Internet to the masses.
- May 1995. Commercial restrictions ended when NSF handed off its sponsorship of the Internet backbone
- By 2000 there were over 50,000,000 hosts. By mid-2004 there were over 350,000,000.



The Internet and the World Wide Web

Number of Internet Hosts



Source: http://en.wikipedia.org/wiki/History_of_the_internet#Opening_the_network_to_commerce



Search Engines – a Brief History

- 1990: ARCHIE, the first tool for searching the (pre-web) Internet, was created. The program downloaded the directory listings of all the files located on public anonymous FTP (File Transfer Protocol) sites, creating a searchable database of file names; however, Archie did not index the contents of these sites. Veronica and Jughead soon followed.
- 1991. Gopher was created. It was a protocol designed for distributing, searching, and retrieving documents over the Internet, and was a predecessor of the World Wide Web. Gopher imposed a strong hierarchy on information stored on it.



Search Engines – a Brief History

- November 1993: Aliweb, the world's first search engine appeared in November 1993. No robot. Depended on being notified by website administrators of the existence at each site of an index file in a particular format.
- 1994. One of the first “crawler” search engines, WebCrawler, let users search for any word in any web page, which became the standard for all major search engines.
- 1994. Lycos was developed at CMU. In April, "Jerry and David's Guide to the World Wide Web" was renamed "Yahoo!" Yahoo's strategy was to create a human-indexed, hierarchical directory to the web. It was superior, but labor-intensive, and became too expensive to keep up with the explosive growth of information online.



Search Engines – a Brief History

- SPAM was an early problem
- 1996, Netscape struck a deal with 5 of the major search engines, where for \$5Million per year each search engine would be in a rotation on the Netscape search engine page. These five engines were: Yahoo!, Magellan, Lycos, Infoseek and Excite.
- 2000: Google rose to prominence with an innovation called PageRank. This iterative algorithm ranks web pages based on the number and PageRank of other web sites and pages that link there, on the premise that good or desirable pages are linked to more than others. This breakthrough was important in the battle against spammers, and Google is now a dominant force in our world.



Search Engines – How They Work

- **Web Crawling.** Store information about the web, which is retrieved using an automated web “crawler” (aka “spider” or “bot”). A crawler follows every link it finds.
- **Indexing.** The contents of each page are analyzed, and findings are stored in the search engine's index database. Google stores all or part of the source page (referred to as a cache) as well as information about the web pages, others, such as AltaVista, store every word of every page they find.
- **Searching.** When a user enters a query into a search engine, the engine examines its index and provides a listing of best-matching web pages according to its criteria, usually with a short summary and sometimes parts of the text.



Search Engine Optimization Basic Concepts

The usefulness of a search engine depends on the relevance of the result set it gives back. The best thing you can do as a website developer is to give them what they're looking for: pages that are rich in information that is relevant to the user.

- Organic and Paid Adword searches
- Search Engine Optimization (SEO)
- Search Engine Marketing (SEM)
- White Hat and Black Hat techniques
- Short and Long tail keywords



Search Engine Basics – Source Matters

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
  "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">

<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en" lang="en">

<head>
<title>Circuit Riders, LLC &#8211; Business Website Specialists</title>
<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1"/>
<meta name="copyright" content="Circuit Riders, LLC"/>
<meta name="author" content="http://www.circuit-riders.net"/>
<meta name="description" content="Circuit Riders is a business website development and
consulting company specializing in aligning business websites with business goals."/>
<meta name="keywords" content="business website consulting, business website development,
business goal website alignment, search engine optimization pittsburgh, seo pittsburgh,
website development pittsburgh, website risk management"/>

<link rel="stylesheet" type="text/css" href="cr.css"/>
</head>
...
  <h1>Business Website Specialists</h1>

  <p>Circuit Riders is a website development and consulting
    company specializing in the business of website
    ownership. We help our clients cut costs and increase
    revenue through strategic business and website planning and
    integration.</p>
```



Getting Started

- Develop the Goals and Requirements of your Search Engine Plan. The plan will be different depending on the answers to these and many other questions:
 - What is the main goal for the website?
 - Is it a design portfolio, a basic business brochure, or a high-volume/high-sales online store?
 - Is time a factor?
 - Is it important to generate a lot of traffic quickly?



Getting Started

- Do the Language Research & Analysis
 - What search keywords and phrases do you want to be found under?
 - What are the SERPs (Search Engine Result Pages) for those terms? Who are they? What do they do or sell?
 - Review the source code of the websites in the top search positions, “above the fold” on page 1, for your terms.
 - What language are they using in their TITLE, and META description and keyword tags? What is their h1? What language do they use in their first paragraph? What are their navigation links called? What are their alt tags? How is their writing?
 - Use Google's Keywords tool for doing basic research even if you aren't going to pay for adwords.



Getting Started – cont'd.

- Establish Your Baseline
 - What is your current ranking, if any, within the top 100 hits on Google, Yahoo, Bing, etc. for the keywords and phrases you have determined are important to you?
 - Keep a record of your research and your results.



Getting Started – cont'd.

- Implement Best Practices
 - Provide content-rich web pages.
 - Update your site every 4 – 6 weeks at a minimum.
 - Follow Google's Webmaster guidelines.
 - Don't try to trick the search engines and users into coming to your site with keyword stuffing, link farms, or other annoying practices that make life hell for the rest of us.



Getting Started – Essential SEO Checklist

- Write clean, standards-compliant HTML and CSS source code
- TITLE, META description and keywords contain key language
- h1 and first paragraph contain key language
- alt tags on images and title tags in links contain key language
- Sitemap
- RSS/MRSS
- Submit to search engines for indexing
- Update with something new every 4 – 6 weeks



Getting Started – SEO & Graphic Design Issues

- Flash. The trouble with Flash is that it's image-intensive, but it does contain text that can be seen by the search engines. You can use Flash and not hurt your SEO, but do your research first.
- JavaScript is awesome, but like Flash and image files, Google and other bots can't see it. You can work around this by providing simple HTML links on a site map. You can also use CSS to achieve many interesting effects on various elements of your page without writing a line of JS.
- Remember: if you're working on a site where traffic really doesn't matter, you can do whatever you want design-wise.



Getting Started – SEO & Graphic Design Issues

- Navigation link images.

Two issues: scalability and SEO. When you use images for your navigation links you have total design control, and sometimes that trumps other considerations. However, if you want to add a new navigation link, or change the text, you are going to have to create a new image. If SEO is an issue, remember to add an alt tag to the image. Your blind users will be grateful.

This also becomes an issue over the lifespan of a website. It can become very difficult and expensive to replace image navigation links. Nobody may know what font was used by the original designer, or they may not have it installed on their machine, or there may be so many links that change is prohibitively expensive.



Getting Started – a Tangent About Color

- Color Management & Browser Resolution, OR:
“Why don't the colors look the like they do on my monitor?”

The problem is that many graphic designers are working with applications that are color-managed, such as Adobe Illustrator and Photoshop. Unfortunately, the only browsers that are color-managed are Safari and Firefox 3. So most users will see different colors than you do. When you are creating images for website display, use the sRGB color profile. The reason is that sRGB is the closest profile to the way monitors display colors by default. Other color profiles have a wider gamut than sRGB, so the colors can't display the same as you see them in PhotoShop or Illustrator on your monitor. Or work with the colors you want, but also view them in sRGB so you know what users will see.



Getting Started – Statistics

The main purpose of SEO is basically to collect and understand the meta data for a website: Who's coming in, what are they looking for, where are they coming from, when do they visit, how long do they stay, what links do they follow?

Web stats can provide many answers to these and other questions; without it your SEO is basically just an intelligent guess.

You can implement basic SEO and follow best practices and take it no further, and you will do yourself no harm. But to really feel the pulse of a site and understand your users and customers, you need to look at the numbers and understand what they mean. It's a form of field research.



Getting Started – Statistics

- Web server logs
- Google Analytics
- Others (MSN, Yahoo, Bing, Woopra, etc.)



Getting Started – Google Analytics

- Pros

- Free
- Better interface than most server log tools
- Industry standard
- Extremely useful
- Highly recommended

- Cons

- Does not retain the keywords when the visit is from images.google.com
- 24 hour refresh lag
- Sometimes they make changes and forget to tell everyone



Business Aspects of SEO/SEM

- If you build it they may come, but how will you know?
- Critical to Advertising. Advertising dollars spent online are increasing exponentially.
- Critical to Sales & Marketing. Understand your website traffic: prospects, customers, competitors.
- Critical to online sales.
- Critical to maintaining and updating your site.



SEO in the Real World – Case Studies

- The pet supply store and the pay-per-click scam. Our introduction to SEO.
- CNHA and the Snakes of Utah
- Utah State Parks Task Force – Edge of the Cedars
- Search & Rescue
- Project Remedies Inc.



Future Trends

- Constant change is the only certainty.
- Stay on top of the industry as much as possible. Everything is changing very quickly, and the best way to keep up with it is to read a little every week. Set aside time for basic industry research – even an hour or two every week will keep you up-to-date with the big picture.
- Learn how to learn. The learning curve goes vertical instantly and never slacks off.
- Don't be paralyzed with worry over not being on top of the latest-greatest whatever. Follow best practices at all times and you will be OK.



*“It's better to know the country
than to be the best cowboy.”*



References

- **Circuit Riders Resources Page:** <http://www.circuit-riders.net/resources.php>
- **History of the Card Catalog:**
http://liswiki.org/wiki/History_of_the_card_catalog
- **History of the Internet:** http://en.wikipedia.org/wiki/History_of_the_internet
- **A Little History of the World Wide Web:** <http://www.w3.org/History.html>
- **Web search engine history:**
http://en.wikipedia.org/wiki/Web_search_engine#cite_note-0
- **Google's Business Solutions:** http://www.google.com/intl/en_us/services/var_2.html
- **Google's Webmaster Tips:** <http://www.google.com/webmasters/>



References cont'd.

- **Some SEO BLOGS:**
<http://organicresponse.com/>
<http://www.seo-theory.com/>
<http://www.seomoz.org/>
- **SEO Conference in March 2010:** <http://www.searchenginestrategies.com/newyork/>
- **Other Analysis Tools:**
<http://www.hittail.com>
<http://www.woopra.com>