GOOGLE SEARCH ALGORITHM & SEOs

Asis Panda
Search Engines Timeline

- ~1991- Tim Berners-lee maintained a list of webservers hosted on the net. No doubt as the numbers of servers grew exponentially it became impossible to keep track.
- 1991- Archie, downloaded the directory listings automatically.
- 1993- A range of cataloging search engines came in to existence like WWWWanderer and Aliweb, which used web robot and being notified by admins of each site.
- 1994- Web Crawlers were introduced. It runs a “full text” crawler based search, which can search any text at any webpage... it was highly desirable.
- 1994~1995- Lycos became the first commercially successful search engine on the net. soon others like Magellan, Excite, Infoseek, Inktomi, Northern Light, and AltaVista joined the suit.
- ~2000- Here comes Google. Google used its Page rank algorithm to become the top ranking search engine the world has ever seen.
- ~ Now- Google is still the #1 search engine. But has gone through a host of modifications in its algorithm.
A typical search engine uses the following three methods:

1. Crawling
2. Indexing
3. Searching

A search engine retrieves information from a website which it does from the html, it is done by the web crawler. The pages are then indexed basing upon the data obtained from the html like from meta tags, descriptions, etc. Indexing accelerates the search by reaching to the relevant information as quickly as possible.

When a user enters a query into a search engine (typically by using keywords), the engine examines its index and provides a listing of best-matching web pages according to its criteria, usually with a short summary containing the document's title and sometimes parts of the text.
How Google Search Works?

1. The web server sends the query to the index servers. The content inside the index servers is similar to the index in the back of a book—it tells which pages contain the words that match any particular query term.

2. The query travels to the doc servers, which actually retrieve the stored documents. Snippets are generated to describe each search result.

3. The search results are returned to the user in a fraction of a second.
Google has three distinct parts:

- **Googlebot**, a web crawler that finds and fetches web pages.
- The **indexer** that sorts every word on every page and stores the resulting index of words in a huge database.
- The **query processor**, which compares your search query to the index and recommends the documents that it considers most relevant.
GoogleBot: google's web crawler

- Googlebot is Google’s web crawling robot, which finds and retrieves pages on the web and hands them off to the Google indexer. It’s easy to imagine Googlebot as a little spider scurrying across the strands of cyberspace, but in reality Googlebot doesn’t traverse the web at all. It functions much like your web browser, by sending a request to a web server for a web page, downloading the entire page, then handing it off to Google’s indexer.
- Googlebot consists of many computers requesting and fetching pages much more quickly than you can with your web browser. In fact, Googlebot can request thousands of different pages simultaneously. To avoid overwhelming web servers, or crowding out requests from human users, Googlebot deliberately makes requests of each individual web server more slowly than it’s capable of doing.
When Googlebot fetches a page, it culls all the links appearing on the page and adds them to a queue for subsequent crawling. Googlebot tends to encounter little spam because most web authors link only to what they believe are high-quality pages. By harvesting links from every page it encounters, Googlebot can quickly build a list of links that can cover broad reaches of the web. This technique, known as deep crawling, also allows Googlebot to probe deep within individual sites. Because of their massive scale, deep crawls can reach almost every page in the web. Because the web is vast, this can take some time, so some pages may be crawled only once a month.

To keep the index current, Google continuously recrawls popular frequently changing web pages at a rate roughly proportional to how often the pages change. Such crawls keep an index current and are known as fresh crawls. Newspaper pages are downloaded daily, pages with stock quotes are downloaded much more frequently. Of course, fresh crawls return fewer pages than the deep crawl. The combination of the two types of crawls allows Google to both make efficient use of its resources and keep its index reasonably current.
Google's Indexer

- Googlebot gives the indexer the full text of the pages it finds. These pages are stored in Google’s index database. This index is sorted alphabetically by search term, with each index entry storing a list of documents in which the term appears and the location within the text where it occurs. This data structure allows rapid access to documents that contain user query terms.
- To improve search performance, Google ignores (doesn’t index) common words called stop words (such as the, is, on, or, of, how, why, as well as certain single digits and single letters). Stop words are so common that they do little to narrow a search, and therefore they can safely be discarded. The indexer also ignores some punctuation and multiple spaces, as well as converting all letters to lowercase, to improve Google’s performance.
Google's Query processor

- The query processor has several parts, including the user interface (search box), the “engine” that evaluates queries and matches them to relevant documents, and the results formatter.
- Google considers over a hundred factors in computing a PageRank and determining which documents are most relevant to a query, including the popularity of the page, the position and size of the search terms within the page, and the proximity of the search terms to one another on the page.
- Indexing the full text of the web allows Google to go beyond simply matching single search terms. Google gives more priority to pages that have search terms near each other and in the same order as the query.
All that in some fractions of seconds!
How so FAST?

Google runs on a distributed network of thousands of low-cost computers and can therefore carry out fast parallel processing. Parallel processing is a method of computation in which many calculations can be performed simultaneously, significantly speeding up data processing.
The concept of page rank was very basic and simple. It mentioned that if a number of important pages linking to your page makes itself important and thus ranks higher than others. Example-if my website has a direct link from yahoo.com, blogger.com and baidu.com then it is likely to have a higher page rank than those who don’t. Google doesn’t only rely on Page Rank for its resource of search. By now it has made numerous changes to its search algorithm.
Google's Search Now

• Google search now doesn’t only depend on Page Rank for its search.
• Currently Google uses the following techniques as its base in its search algorithm
  ➢ Gauge of the trustworthiness of the Website
  ➢ Anchor text in the external links
  ➢ On-page keyword usage
  ➢ Page rank/link juice
• All above ranking factors together in mixed priority make up the 2009-2010 Google's search algorithm.
Importance Levels of Various Ranking Factors in Google's Algorithm Over Time

Ranking Criteria:
- Trust / Authority of Host Domain
- Anchor Text in External Links
- On-Page Keyword Usage
- Raw PageRank / Link Juice

(Note: This diagram is the personal opinion of Rand Fishkin from SEOMoz)
Search Engine optimization

There is definitely a need in business or in any genre of websites to appear higher in the search engine’s result page (SERP) of Google.

A place at the top of the SERP ensures a definite reach to all customers or target/intended users. Optimization can be done organically or by paid-terms. We do organic optimization which is ethically, free of cost and also yields a genuine result.

So how is it done...
SERP stands for Search Engine Results Page
Our position on SERP depends on following factors:

1. Webpage’s Page keyword density
2. Webpage’s keyword prominence for these keywords.
3. Link popularity-no. of other websites linking to you.
4. Anchor text – The outgoing hyperlinks in the text of the website
5. Your link and keyword relevance
Two types of SEOs

1. On-Page
2. Off-Page
On-page SEO is the process of optimizing the content of your website. This includes the text, images and links on your website. Anything uploaded to your site's domain is considered on page.
On-page SEO

1. Webpage layout factors relevant to SEO
2. Site structure
1.1 Webpage layout factors relevant to SEO

1. Amount of text in a page
2. Number of keywords on a page
3. Keyword density
4. Location of keywords on a page
5. Text format
6. Title tag
7. Keywords in links
8. Description meta tag
9. Keywords Meta tag
2.2 Site Structure

1. Number of pages
2. Navigation menu
3. Keywords in page names
4. Avoid subdirectories
5. One page-one keyword phrase
6. SEO and main page
Off-page SEO

3.1 Inbound links to sites are taken into account
3.2 Link importance (citation index)
3.3 Link Text (anchor text)
3.4 Relevance of referring pages
3.5 Increasing link popularity
   3.5.1 Submitting to general purpose directory
   3.5.2 DMOZ directory
   3.5.3 Press releases, news feeds, thematic resources
End