

## Evolution of Search

The Internet revolution drove every company towards creating a web site. Yahoo! was one of the first locations where this new information age was organized. Very much like an on-line yellow pages, this was directory organized into hierarchies based on the type of organization or services offered. This was and still is a manual process of submitting a web site page into this hierarchy and hoping the manual operators would update the directory in a timely manner. Very little automation was available and many new search engine submittal services showed up to facilitate this, all at a price.

The web site crawler, or spider, arrived later. At first, these new spiders were used to determine the viability of any one HTML page by checking the resulting links and determining if they were still active. The World Wide Web is a constantly changing series of individual documents accessible one page at a time using an URL or uniform resource locator.

An individual web page link or an entire domain location many disappear without any notification to those that link to it. The 'dead link' check was important maintenance while keeping each page up-to-date. The crawler soon became popular to look through a web site searching for certain words, or keywords. The location of a web site in the result set became known as the page rank, the more keywords found in a page meant the page was displayed first in the results. At first, most web sites crawling relied on the meta data keywords to categorize individual pages and the subsequent page ranking. Although Google was not the first search engine that used the crawler to populate a keyword database, it soon became the biggest.

The logic or algorithm behind the determining which web page to display first became a hot topic and Google's founders published their ideas at Stanford as a college project while creating the foundation for what Google is today. Google was one of the first search engines that did not rely on the meta data keywords for ranking, but used other factors including the HTML title and included visible text within the actual displayed page. Many webmasters became intrigued with reverse engineering these algorithms in an attempt to keep their pages at the top of the search results. In many ways, this changed the creation of web pages by reducing the amount of cute graphics on a page to the more Google friendly preferred text.

There were also many other improvements going on in the search business at the time. At one point, it was announced that there were more web sites and Internet users outside of the United States which created the need for multi-language search engines. Although there is no automated translation built into web site crawling, the ability to search and retrieve non-English results was a milestone and now an expected feature. Another improvement was the inclusion of spell checking and synonyms, many search engines returned messages that asked the user if the correctly spelled keyword or a synonym of the keyword wanted to be searched. Along the same lines, the number of results returned became so large that a search engines started offering the 'search in the search results' allowing the user to bring the result set to a more reasonable number. Finally, the need to improve search results led NASA into creating a new Contextual search engine where results are returned that relate to only a specific topic, the context of the keyword.

The search engine evolution will continue as new algorithms are integrated and tried, some will fail and never see a web page, while others may be quietly integrated and become part of the fabric of the Internet. Even the NASA Netmark Contextual Search algorithm, which is only available today through Xerox and Black Tulip Systems, may become another evolution that is integrated into Google and other Internet search engines in the future. One thing is true, without the search engines we use today, it would be almost impossible to navigate the over hundred million domains and billions of individual web pages.