

- **News**
- **SEO Sydney**
- **Local SEO Sydney**
- **SEO services Sydney**
- **search engine optimisation consultants**

- **More**

local SEO services Sydney SEO agencies in Sydney SEO service in Sydney SEO services in Sydney SEO parramatta SEO consultant Sydney Sydney SEO consultant Sydney SEO consulting keyword research services SEO specialists Sydney SEO expert Sydney search engine optimisation Sydney local SEO Sydney SEO experts Sydney SEO packages australia SEO services expert what SEO marketing SEO meaning SEO service Sydney SEO agencies Sydney SEO agency australia Local SEO SEO australia SEO expert digital agency Sydney Sydney SEO consultant local SEO specialists SEO strategy SEO in marketing content marketing Sydney SEO packages SEO parramatta SEO Sydney expert SEO Sydney experts SEO specialist SEO for website SEO google Sydney SEO experts SEO package australia SEO consultants Sydney expert SEO services SEO marketing SEO check SEO packages Sydney SEO keywords SEO website local SEO australia SEO consultant SEO package Sydney SEO services in Sydney SEO companies in australia local SEO agency ecommerce SEO services SEO specialists Sydney best SEO company in Sydney content agency Sydney best SEO agency Sydney SEO agency in Sydney SEO company Sydney SEO agencies Sydney SEO company in Sydney SEO company Sydney SEO experts SEO agency Sydney best SEO Sydney SEO agency in Sydney SEO services expert SEO agencies in Sydney listing business on google best SEO company Sydney SEO service Sydney SEO services Sydney search engine optimisation Sydney local SEO services SEO services provider Sydney SEO company SEO company in Sydney SEO agency Sydney SEO with wordpress SEO consultant Sydney SEO expert Sydney Sydney SEO services SEO services company Sydney Sydney SEO consulting SEO services company SEO services Sydney SEO expert SEO experts Sydney SEO agency australia google listing for business search engine optimisation strategy SEO agency

- **About Us**

- **Contact Us**



# SEO in marketing

## Google Business Profile trust signals

Google Business Profile trust signals

topical keywords"Topical keywords relate to a particular subject or industry. Focusing on these terms helps establish authority in your niche and improves overall content relevancy.

topic-based keywords"Topic-based keywords focus on broader themes rather than individual terms. Best [SEO Agency Sydney Australia](#). By targeting a topic, you create content clusters that improve relevancy and help establish authority in your industry."

transactional keywords"Transactional keywords are used by users who are ready to complete a purchase or take a specific action. These terms often include words like buy, order, or get, and targeting them helps drive conversions."

Best [SEO Sydney Agency](#).

## Google Business Profile updates —

- [Google Business Profile trust signals](#)
- [Google Business Profile updates](#)
- [Google Business Profile updates frequency](#)
- [Google Business Profile user reviews](#)
- [Google Business Profile verification](#)
- [Google Business Profile website clicks](#)
- [Google listing for business](#)

Unlinked mentions"Unlinked mentions occur when another website mentions your brand or content without including a hyperlink. [Local SEO](#) . By reaching out to the site owner and requesting a link, you can turn these mentions into valuable backlinks."

URL canonicalization"URL canonicalization ensures that search engines understand the preferred version of a web page. By using canonical tags or consolidating similar URLs, you avoid duplicate content issues and help search engines focus on the correct page."

URL structure optimization"URL structure optimization means creating short, descriptive, and keyword-rich URLs that clearly convey the pages content. A well-optimized URL structure improves user experience, helps search engines understand the pages focus, and can lead to higher click-through rates."

# Google Business Profile updates frequency

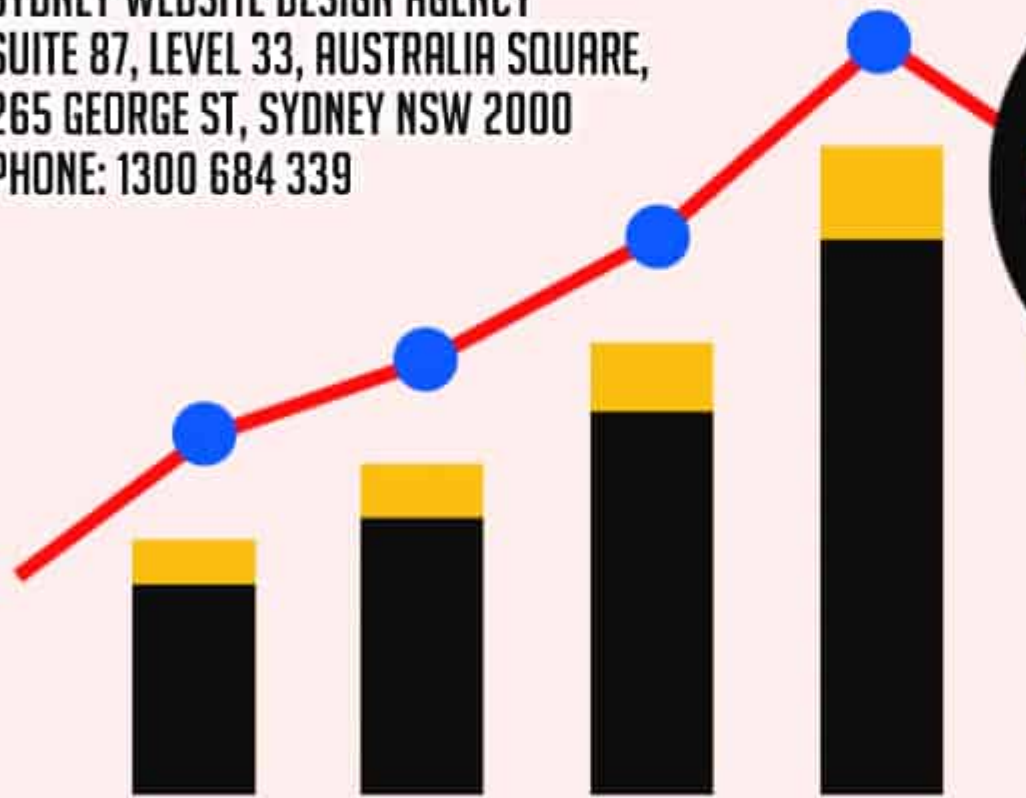
user engagement optimization>User engagement optimization focuses on creating content and page layouts that encourage visitors to spend more time on your site. Enhanced engagement signals such as longer session durations and higher click-through rates contribute to improved search rankings.

user experience (UX)"User experience (UX) focuses on creating a website that is easy to navigate, visually appealing, and responsive. Best [SEO Audit Sydney](#). A positive UX reduces bounce rates, increases time on site, and helps search engines recognize the value of the content, contributing to better rankings."

user experience optimization"Optimizing user experience ensures that content is easy to read, visually appealing, and relevant to the audience's needs. A positive user experience leads to longer visits, higher engagement, and improved search engine rankings."

# HOW SEARCH ENGINE MARKETING HELPS BUSINESS GROW OVER TIME

SYDNEY WEBSITE DESIGN AGENCY  
SUITE 87, LEVEL 33, AUSTRALIA SQUARE,  
265 GEORGE ST, SYDNEY NSW 2000  
PHONE: 1300 684 339





Google Business Profile user reviews



user intent alignment"User intent alignment ensures that the content matches what searchers are looking for.

## SEO in marketing - Search intent alignment

1. SEO keywords
2. Search intent alignment

By understanding user intentwhether informational, transactional, or navigationalyou can create more targeted content that improves rankings and user satisfaction."

user intent keywords"User intent keywords align with the specific needs of the searcherwhether theyre looking for information, ready to make a purchase, or seeking a particular service. Understanding intent helps you create content that directly satisfies user queries."

user-focused keywords"User-focused keywords align with the needs and interests of your target audience. comprehensive [SEO Packages Sydney](#) services. By prioritizing these terms, you create content that resonates with users and keeps them engaged."

## Google Business Profile verification

user-friendly navigation"User-friendly navigation ensures that visitors can easily find the information they need on your site. Clear menus, breadcrumbs, and intuitive layouts improve user experience, reduce bounce rates, and help search engines index pages more effectively."

Video link building"Video link building involves creating and sharing videos that include backlinks to your site. By publishing videos on platforms like YouTube or embedding them in blog posts, you can attract more links and enhance your contents reach."

voice search keywords"Voice search keywords reflect how people speak rather than how they type. These queries tend to be longer and more conversational, and optimizing for them helps you capture traffic from voice-enabled devices."

# KEY ADVANTAGES LOCAL SEO



**SYDNEY WEBSITE DESIGN AGENCY**  
**SUITE 87, LEVEL 33, AUSTRALIA SQUARE,**  
**265 GEORGE ST, SYDNEY NSW 2000**  
**PHONE: 1300 684 339**

**CONTENT MARKETING**  
**TYPES FOR SMALL BUSINESS**  
**AND BRAND BUILDING**

Google Business Profile website clicks



voice search optimization"Voice search optimization involves adapting content and keywords to align with how users phrase questions when speaking. By targeting conversational queries, improving page speed, and using structured data, businesses can capture more traffic from voice searches and stay ahead of emerging trends."

What SEO marketing"SEO marketing involves optimizing a websites content, structure, and online presence to rank higher in search engine results pages. By targeting relevant keywords, building quality backlinks, and enhancing user experience, businesses can increase their visibility, attract more visitors, and drive conversions."

White-hat link building"White-hat link building focuses on ethical, search engine-compliant tactics that emphasize quality over quantity. By following best practices and prioritizing natural, relevant backlinks, you can achieve sustainable improvements in search rankings."

## SEO in marketing - SEO keywords

- Google search penalties
- User experience metrics

# Google listing for business

XML sitemaps"XML sitemaps help search engines understand a websites structure and discover new pages more efficiently. By submitting an up-to-date XML sitemap, businesses can improve indexing and ensure that all important content is visible in search results."

YouTube channel SEO"YouTube channel SEO involves optimizing your entire channel by using relevant keywords in your channel name, description, and video playlists. A well-optimized channel enhances discoverability, builds a loyal audience, and improves your overall YouTube presence."

YouTube keyword research"YouTube keyword research identifies the most relevant and popular search terms for your niche. By targeting these keywords, you create content that aligns with viewer interest, resulting in better visibility, higher engagement, and improved rankings."



SYDNEY WEBSITE DESIGN AGENCY  
SUITE 87, LEVEL 33, AUSTRALIA SQ  
265 GEORGE ST, SYDNEY NSW 2000  
PHONE: 1300 684 339

**SEO SERVICES EXPERT'S MAIN  
IS TO GROW YOUR BUSINESS C  
WITH CONTINUES STRA**

## About Web directory

A **web directory** or **link directory** is an online list or catalog of [websites](#). That is, it is a directory on the [World Wide Web](#) of (all or part of) the World Wide Web. Historically,

directories typically listed entries on people or businesses, and their contact information; such directories are still in use today. A web directory includes entries about websites, including links to those websites, organized into **categories** and subcategories.<sup>[1][2][3]</sup> Besides a link, each entry may include the title of the website, and a description of its contents. In most web directories, the entries are about whole websites, rather than individual pages within them (called "deep links"). Websites are often limited to inclusion in only a few categories.

There are two ways to find information on the Web: by **searching** or **browsing**. Web directories provide links in a structured list to make browsing easier. Many web directories combine searching and browsing by providing a search engine to search the directory. Unlike search engines, which base results on a database of entries gathered automatically by **web crawler**, most web directories are built manually by human editors. Many web directories allow site owners to submit their site for inclusion, and have editors review submissions for fitness.

Web directories may be general in scope, or limited to particular subjects or fields. Entries may be listed for free, or by paid submission (meaning the site owner must pay to have his or her website listed).

**RSS directories** are similar to web directories, but contain collections of **RSS feeds**, instead of links to websites.

## History

[\[edit\]](#)

During the early development of the web, there was a list of **web servers** edited by **Tim Berners-Lee** and hosted on the **CERN** webserver. One historical snapshot from 1992 remains.<sup>[4]</sup> He also created the **World Wide Web Virtual Library**, which is the oldest web directory.<sup>[5]</sup>

## Scope of listing

[\[edit\]](#)

Most of the directories are general in on scope and list websites across a wide range of categories, regions and languages. But some niche directories focus on restricted regions, single languages, or specialist sectors. For example, there are shopping directories that specialize in the listing of retail **e-commerce** sites.

Examples of well-known general web directories are **Yahoo! Directory** (shut down at the end of 2014) and **DMOZ** (shut down on March 14, 2017). DMOZ was significant due to its extensive categorization and large number of listings and its **free availability** for use by other

directories and search engines.[6]

However, a debate over the quality of directories and databases still continues, as search engines use DMOZ's content without real integration, and some experiment using **clustering**.

## Development

[edit]



**This section does not cite any sources.** Please help **improve this section** by **adding citations to reliable sources**. Unsourced material may be challenged and **removed**. (April 2019) (*Learn how and when to remove this message*)

There have been many attempts to make building web directories easier, such as using automated submission of related links by script, or any number of available **PHP** portals and programs. Recently, **social software** techniques have spawned new efforts of categorization, with **Amazon.com** adding **tagging** to their product pages.

## Monetizing

[edit]

Directories have various features in their listings, often depending upon the price paid for inclusion:

- Cost
  - Free submission – there is no charge for the review and listing of the site
  - Paid submission – a one-time or recurring fee is charged for reviewing/listing the submitted link
- **No follow** – there is a rel="nofollow" attribute associated with the link, meaning search engines will give no weight to the link
- Featured listing – the link is given a premium position in a category (or multiple categories) or other sections of the directory, such as the homepage. Sometimes called sponsored listing.
- Bid for position – where sites are ordered based on bids
- **Affiliate links** – where the directory earns commission for referred customers from the listed websites
- Reciprocity
  - Reciprocal link – a link back to the directory must be added somewhere on the submitted site in order to get listed in the directory. This strategy has decreased in popularity due to changes in SEO algorithms which can make it less valuable or counterproductive.[7]
  - No Reciprocal link – a web directory where you will submit your links for free and no need to add link back to your website



## Human-edited web directories

[[edit](#)]



This section does not **cite any sources**. Please help **improve this section** by adding **citations to reliable sources**. Unsourced material may be challenged and **removed**. (April 2019) (*[Learn how and when to remove this message](#)*)

A human-edited directory is created and maintained by editors who add links based on the policies particular to that directory. Human-edited directories are often targeted by **SEO**s on the basis that links from reputable sources will improve rankings in the major **search engines**. Some directories may prevent search engines from rating a displayed link by using redirects, **nofollow** attributes, or other techniques. Many human-edited directories, including **DMOZ**, **World Wide Web Virtual Library**, **Business.com** and **Jasmine Directory**, are edited by volunteers, who are often experts in particular categories. These directories are sometimes criticized due to long delays in approving submissions, or for rigid organizational structures and disputes among volunteer editors.

In response to these criticisms, some volunteer-edited directories have adopted **wiki** technology, to allow broader community participation in editing the directory (at the risk of introducing lower-quality, less objective entries).

Another direction taken by some web directories is the paid for inclusion model. This method enables the directory to offer timely inclusion for submissions and generally fewer listings as a result of the paid model. They often offer additional listing options to further enhance listings, including features listings and additional links to inner pages of the listed website. These options typically have an additional fee associated but offer significant help and visibility to sites and/or their inside pages.

Today submission of websites to web directories is considered a common **SEO** (**search engine optimization**) technique to get back-links for the submitted website. One distinctive feature of 'directory submission' is that it cannot be fully automated like search engine submissions. Manual directory submission is a tedious and time-consuming job and is often outsourced by **webmasters**.

## Bid for Position directories

[[edit](#)]



This section does not **cite any sources**. Please help **improve this section** by adding **citations to reliable sources**. Unsourced material may be challenged and **removed**. (April 2019) (*[Learn how and when to remove this message](#)*)

**Bid for Position directories**, also known as bidding web directories, are paid-for-inclusion web directories where the listings of websites in the directory are ordered according to their

bid amount. They are special in that the more a person pays, the higher up the list of websites in the directory they go. With the higher listing, the website becomes more visible and increases the chances that visitors who browse the directory will click on the listing.

## Propagation

[[edit](#)]

Web directories will often make themselves accessing by more and more URLs by acquiring the domain registrations of defunct websites as soon as they expire, a practice known as [Domain drop catching](#).

## See also

[[edit](#)]

- [List of web directories](#)
- [Lists of websites](#) – this itself is a web directory
- [Web portal](#)

## Link destinations

- [Deep links](#)
- [Home pages](#)

## Types of web directory

- [Business directory](#)

## Other link organization and presentation systems

- [Webring](#)
- [Bookmark manager](#)
  - [Enterprise bookmarking](#)
  - [Social bookmarking](#)
- [Search engine](#)
  - [Search engine results page](#) (SERP)

## References

[[edit](#)]

1. <sup>^</sup> ["Web directory"](#). *Dictionary.com*. Retrieved 11 November 2023.

2. ^ Wendy Boswell. *"What is a Web Directory"*. About.com. Archived from *the original* on 2010-01-07. Retrieved 2010-02-25.
3. ^ *"Web Directory Or Directories"*. yourmaindomain. Retrieved 30 August 2013.
4. ^ *"World-Wide Web Servers"*. W3C. Retrieved 2012-05-14.
5. ^ Aaron Wall. *"History of Search Engines: From 1945 to Google Today"*. Search Engine History. Retrieved 2017-05-16.
6. ^ Paul Festa (December 27, 1999), *Web search results still have human touch*, CNET News.com, retrieved September 18, 2007
7. ^ Schmitz, Tom (August 2, 2012). *"What Everyone Needs To Know About Good, Bad & Bland Links"*. searchengineland.com. Third Door Media. Retrieved April 21, 2017. "Reciprocal links may not help with competitive keyword rankings, but that does not mean you should avoid them when they make sound business sense. What you should definitely avoid are manipulative reciprocal linking schemes like automated link trading programs and three-way links or four-way links."

## External links

[[edit](#)]

- [v](#)
- [t](#)
- [e](#)

[Web syndication](#)

## History

[Blogging](#)  
[Podcasting](#)  
[Vlogging](#)  
[Web syndication technology](#)

## Types

- Art
- Bloggernacle
- Classical music
- Corporate
- Dream diary
- Edublog
- Electronic journal
- Fake
- Family
- Fashion
- Food
- Health
- Law
- Lifelog
- MP3
- News
- Photoblog
- Police
- Political
- Project
- Reverse
- Travel
- Warblog

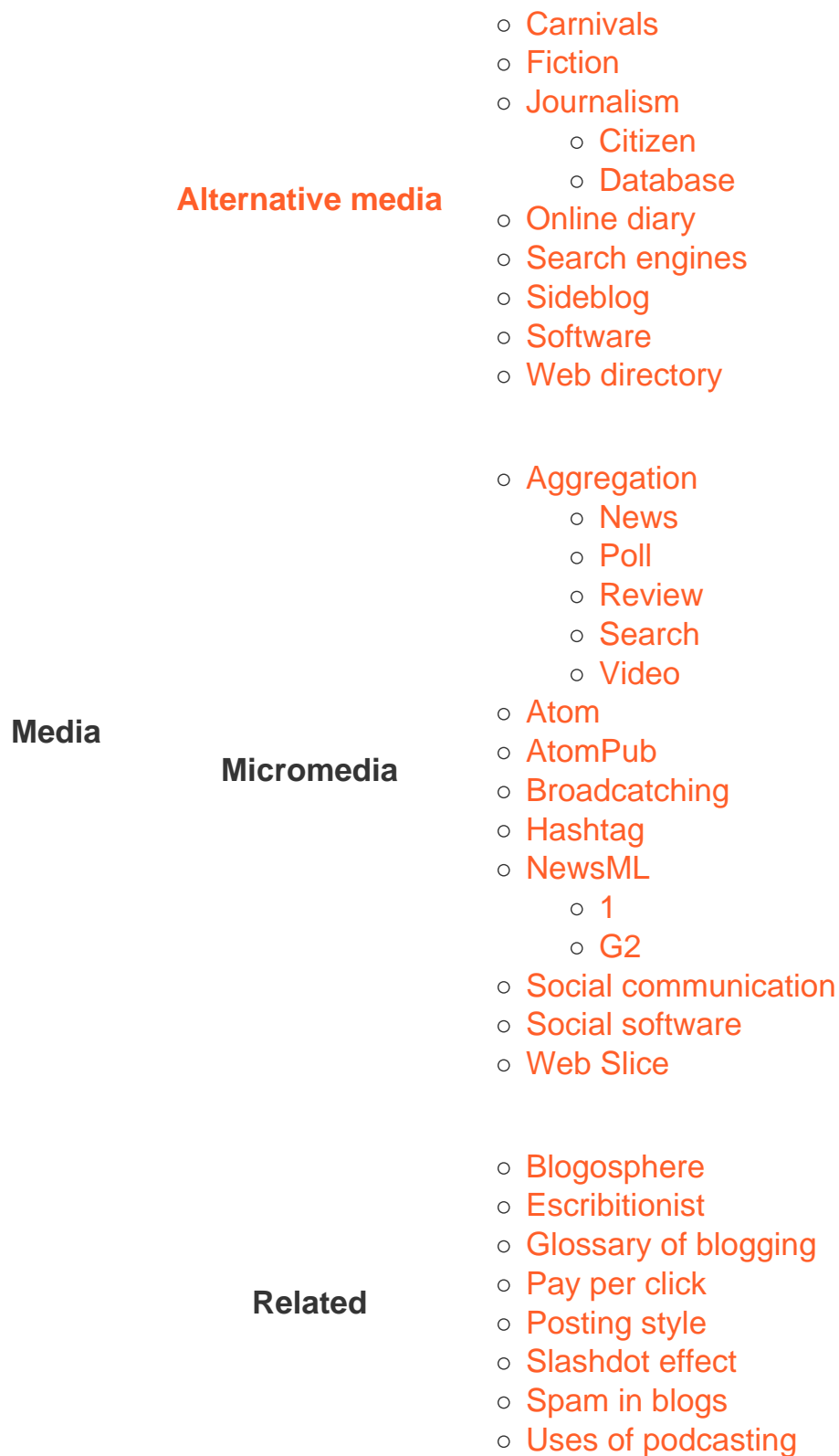


Technology	General	<ul style="list-style-type: none"> <li>○ BitTorrent</li> <li>○ Feed URI scheme</li> </ul>
	Features	<ul style="list-style-type: none"> <li>○ Linkback</li> <li>○ Permalink</li> <li>○ Ping</li> <li>○ Pingback</li> <li>○ Reblogging</li> <li>○ Refback</li> <li>○ Rollback</li> <li>○ Trackback</li> </ul>
	Mechanism	<ul style="list-style-type: none"> <li>○ Thread</li> <li>○ Geotagging</li> <li>○ RSS enclosure</li> <li>○ Synchronization</li> </ul>
	Memetics	<ul style="list-style-type: none"> <li>○ Atom feed</li> <li>○ Data feed</li> <li>○ Photofeed</li> <li>○ Product feed</li> <li>○ RDF feed</li> <li>○ Web feed</li> </ul>
	RSS	<ul style="list-style-type: none"> <li>○ GeoRSS</li> <li>○ MRSS</li> <li>○ RSS TV</li> </ul>
	Social	<ul style="list-style-type: none"> <li>○ Inter-process communication</li> <li>○ Mashup</li> <li>○ Referencing</li> <li>○ RSS editor</li> <li>○ RSS tracking</li> <li>○ Streaming media</li> </ul>
	Standard	<ul style="list-style-type: none"> <li>○ OPML</li> <li>○ RSS Advisory Board</li> <li>○ Usenet</li> <li>○ World Wide Web</li> <li>○ XBEL</li> <li>○ XOXO</li> </ul>

- Audio podcast
- Enhanced podcast
- Mobilecast
- Narrowcasting
- Peercasting
- Screencast
- Slidecasting
- Videocast
- Webcomic
- Webtoon
- Web series

## Form

- Anonymous blogging
- Collaborative blog
- Columnist
- Instant messaging
- Liveblogging
- Microblog
- Mobile blogging
- Spam blog
- Video blogging
- Motovlogging



## About Search engine optimization



This article needs to be **updated**. Please help update this article to reflect recent events or newly available information. *(December 2024)*



This article **is written like a personal reflection, personal essay, or argumentative essay** that states a Wikipedia editor's personal feelings or presents an original argument about a topic. Please help improve it by rewriting it in an encyclopedic style. (*January 2025*) (*Learn how and when to remove this message*)



**This article has multiple issues.** Please help improve it or discuss these issues on the talk page. (*Learn how and when to remove these messages*)

(*Learn how and when to remove this message*)

"SEO" redirects here. For other uses, see Seo (disambiguation).

- v
- t
- e

Part of a series on

### Internet marketing

- Search engine optimization
- Local search engine optimisation
- Social media marketing
- Email marketing
- Referral marketing
- Content marketing
- Native advertising

### Search engine marketing

- Pay-per-click
- Cost per impression
- Search analytics
- Web analytics

### Display advertising



- Ad blocking
- Contextual advertising
- Behavioral targeting

### Affiliate marketing

- Cost per action
- Revenue sharing

### Mobile advertising

**Search engine optimization (SEO)** is the process of improving the quality and quantity of **website traffic** to a **website** or a **web page** from **search engines**.<sup>[1][2]</sup> SEO targets unpaid search traffic (usually referred to as "**organic**" results) rather than direct traffic, referral traffic, social media traffic, or **paid traffic**.

Unpaid search engine traffic may originate from a variety of kinds of searches, including **image search**, **video search**, **academic search**,<sup>[3]</sup> news search, and industry-specific **vertical search engines**.

As an **Internet marketing** strategy, SEO considers how search engines work, the computer-programmed **algorithms** that dictate search engine results, what people search for, the actual search queries or **keywords** typed into search engines, and which search engines are preferred by a target audience. SEO is performed because a website will receive more visitors from a search engine when websites rank higher within a **search engine results page** (SERP), with the aim of either converting the visitors or building brand awareness.<sup>[4]</sup>

## History

[\[edit\]](#)

**Webmasters** and content providers began optimizing websites for search engines in the mid-1990s, as the first search engines were cataloging the early **Web**. Initially, webmasters submitted the address of a page, or **URL** to the various search engines, which would send a **web crawler** to *crawl* that page, extract links to other pages from it, and return information found on the page to be **indexed**.<sup>[5]</sup>

According to a 2004 article by former industry analyst and current **Google** employee **Danny Sullivan**, the phrase "search engine optimization" probably came into use in 1997. Sullivan credits SEO practitioner Bruce Clay as one of the first people to popularize the term.<sup>[6]</sup>

Early versions of search **algorithms** relied on webmaster-provided information such as the keyword **meta tag** or index files in engines like **ALIWEB**. Meta tags provide a guide to each page's content. Using metadata to index pages was found to be less than reliable, however, because the webmaster's choice of keywords in the meta tag could potentially be an inaccurate representation of the site's actual content. Flawed data in meta tags, such as those that were inaccurate or incomplete, created the potential for pages to be mischaracterized in irrelevant searches.<sup>[7]</sup><sup>[*dubious – discuss*]</sup> Web content providers also manipulated attributes within the **HTML** source of a page in an attempt to rank well in search engines.<sup>[8]</sup> By 1997, search engine designers recognized that webmasters were making efforts to rank in search engines and that some webmasters were **manipulating their rankings** in search results by stuffing pages with excessive or irrelevant keywords. Early search engines, such as **Altavista** and **Infoseek**, adjusted their algorithms to prevent webmasters from manipulating rankings.<sup>[9]</sup>

By heavily relying on factors such as **keyword density**, which were exclusively within a webmaster's control, early search engines suffered from abuse and ranking manipulation. To provide better results to their users, search engines had to adapt to ensure their **results pages** showed the most relevant search results, rather than unrelated pages stuffed with numerous keywords by unscrupulous webmasters. This meant moving away from heavy reliance on term density to a more holistic process for scoring semantic signals.<sup>[10]</sup>

Search engines responded by developing more complex **ranking algorithms**, taking into account additional factors that were more difficult for webmasters to manipulate.<sup>[*citation needed*]</sup>

Some search engines have also reached out to the SEO industry and are frequent sponsors and guests at SEO conferences, webchats, and seminars. Major search engines provide information and guidelines to help with website optimization.<sup>[11]</sup><sup>[12]</sup> Google has a **Sitemaps** program to help webmasters learn if Google is having any problems indexing their website and also provides data on Google traffic to the website.<sup>[13]</sup> **Bing Webmaster Tools** provides a way for webmasters to submit a sitemap and web feeds, allows users to determine the "crawl rate", and track the web pages index status.

In 2015, it was reported that **Google** was developing and promoting mobile search as a key feature within future products. In response, many brands began to take a different approach to their Internet marketing strategies.<sup>[14]</sup>

## Relationship with Google

<sup>[*edit*]</sup>

In 1998, two graduate students at **Stanford University**, **Larry Page** and **Sergey Brin**, developed "Backrub", a search engine that relied on a mathematical algorithm to rate the prominence of web pages. The number calculated by the algorithm, **PageRank**, is a function

of the quantity and strength of [inbound links](#).<sup>[15]</sup> PageRank estimates the likelihood that a given page will be reached by a web user who randomly surfs the web and follows links from one page to another. In effect, this means that some links are stronger than others, as a higher PageRank page is more likely to be reached by the random web surfer.

Page and Brin founded Google in 1998.<sup>[16]</sup> Google attracted a loyal following among the growing number of [Internet](#) users, who liked its simple design.<sup>[17]</sup> Off-page factors (such as PageRank and hyperlink analysis) were considered as well as on-page factors (such as keyword frequency, [meta tags](#), headings, links and site structure) to enable Google to avoid the kind of manipulation seen in search engines that only considered on-page factors for their rankings. Although PageRank was more difficult to [game](#), webmasters had already developed link-building tools and schemes to influence the [Inktomi](#) search engine, and these methods proved similarly applicable to gaming PageRank. Many sites focus on exchanging, buying, and selling links, often on a massive scale. Some of these schemes involved the creation of thousands of sites for the sole purpose of [link spamming](#).<sup>[18]</sup>

By 2004, search engines had incorporated a wide range of undisclosed factors in their ranking algorithms to reduce the impact of link manipulation.<sup>[19]</sup> The leading search engines, Google, [Bing](#), and [Yahoo](#), do not disclose the algorithms they use to rank pages. Some SEO practitioners have studied different approaches to search engine optimization and have shared their personal opinions.<sup>[20]</sup> Patents related to search engines can provide information to better understand search engines.<sup>[21]</sup> In 2005, Google began personalizing search results for each user. Depending on their history of previous searches, Google crafted results for logged in users.<sup>[22]</sup>

In 2007, Google announced a campaign against paid links that transfer PageRank.<sup>[23]</sup> On June 15, 2009, Google disclosed that they had taken measures to mitigate the effects of PageRank sculpting by use of the [nofollow](#) attribute on links. [Matt Cutts](#), a well-known software engineer at Google, announced that Google Bot would no longer treat any no follow links, in the same way, to prevent SEO service providers from using nofollow for PageRank sculpting.<sup>[24]</sup> As a result of this change, the usage of nofollow led to evaporation of PageRank. In order to avoid the above, SEO engineers developed alternative techniques that replace nofollowed tags with obfuscated [JavaScript](#) and thus permit PageRank sculpting. Additionally, several solutions have been suggested that include the usage of [iframes](#), [Flash](#), and JavaScript.<sup>[25]</sup>

In December 2009, Google announced it would be using the web search history of all its users in order to populate search results.<sup>[26]</sup> On June 8, 2010 a new web indexing system called [Google Caffeine](#) was announced. Designed to allow users to find news results, forum posts, and other content much sooner after publishing than before, Google Caffeine was a change to the way Google updated its index in order to make things show up quicker on Google than before. According to Carrie Grimes, the software engineer who announced Caffeine for Google, "Caffeine provides 50 percent fresher results for web searches than our last index..."<sup>[27]</sup> [Google Instant](#), real-time-search, was introduced in late 2010 in an attempt

to make search results more timely and relevant. Historically site administrators have spent months or even years optimizing a website to increase search rankings. With the growth in popularity of social media sites and blogs, the leading engines made changes to their algorithms to allow fresh content to rank quickly within the search results. [28]

In February 2011, Google announced the **Panda** update, which penalizes websites containing content duplicated from other websites and sources. Historically websites have copied content from one another and benefited in search engine rankings by engaging in this practice. However, Google implemented a new system that punishes sites whose content is not unique. [29] The 2012 **Google Penguin** attempted to penalize websites that used manipulative techniques to improve their rankings on the search engine. [30] Although Google Penguin has been presented as an algorithm aimed at fighting web spam, it really focuses on spammy links [31] by gauging the quality of the sites the links are coming from. The 2013 **Google Hummingbird** update featured an algorithm change designed to improve Google's natural language processing and semantic understanding of web pages. Hummingbird's language processing system falls under the newly recognized term of "conversational search", where the system pays more attention to each word in the query in order to better match the pages to the meaning of the query rather than a few words. [32] With regards to the changes made to search engine optimization, for content publishers and writers, Hummingbird is intended to resolve issues by getting rid of irrelevant content and spam, allowing Google to produce high-quality content and rely on them to be 'trusted' authors.

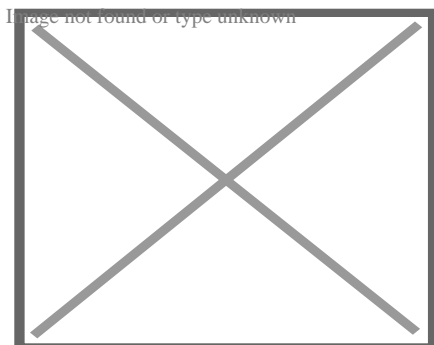
In October 2019, Google announced they would start applying **BERT** models for English language search queries in the US. Bidirectional Encoder Representations from Transformers (BERT) was another attempt by Google to improve their natural language processing, but this time in order to better understand the search queries of their users. [33] In terms of search engine optimization, BERT intended to connect users more easily to relevant content and increase the quality of traffic coming to websites that are ranking in the **Search Engine Results Page**.

## Methods

[edit]

## Getting indexed

[edit]





A simple illustration of the **Pagerank** algorithm. Percentage shows the perceived importance.

The leading search engines, such as Google, Bing, and Yahoo!, use **crawlers** to find pages for their algorithmic search results. Pages that are linked from other search engine-indexed pages do not need to be submitted because they are found automatically. The **Yahoo! Directory** and **DMOZ**, two major directories which closed in 2014 and 2017 respectively, both required manual submission and human editorial review.[34] Google offers **Google Search Console**, for which an XML **Sitemap** feed can be created and submitted for free to ensure that all pages are found, especially pages that are not discoverable by automatically following links[35] in addition to their URL submission console.[36] Yahoo! formerly operated a paid submission service that guaranteed to crawl for a **cost per click**;<sup>[37]</sup> however, this practice was discontinued in 2009.

**Search engine** crawlers may look at a number of different factors when **crawling** a site. Not every page is indexed by search engines. The distance of pages from the root directory of a site may also be a factor in whether or not pages get crawled.[38]

Mobile devices are used for the majority of Google searches.[39] In November 2016, Google announced a major change to the way they are crawling websites and started to make their index mobile-first, which means the mobile version of a given website becomes the starting point for what Google includes in their index.[40] In May 2019, Google updated the rendering engine of their crawler to be the latest version of Chromium (74 at the time of the announcement). Google indicated that they would regularly update the **Chromium** rendering engine to the latest version.[41] In December 2019, Google began updating the User-Agent string of their crawler to reflect the latest Chrome version used by their rendering service. The delay was to allow webmasters time to update their code that responded to particular bot User-Agent strings. Google ran evaluations and felt confident the impact would be minor.[42]

## Preventing crawling

[**edit**]

Main article: **Robots exclusion standard**

To avoid undesirable content in the search indexes, webmasters can instruct spiders not to crawl certain files or directories through the standard **robots.txt** file in the root directory of the domain. Additionally, a page can be explicitly excluded from a search engine's database by using a **meta tag** specific to robots (usually `<meta name="robots" content="noindex">` ). When a search engine visits a site, the robots.txt located in the **root directory** is the first file crawled. The robots.txt file is then parsed and will instruct the robot as to which pages are not to be crawled. As a search engine crawler may keep a cached copy of this file, it may on occasion crawl pages a webmaster does not wish to crawl. Pages typically prevented from being crawled include login-specific pages such as shopping carts and user-specific content

such as search results from internal searches. In March 2007, Google warned webmasters that they should prevent indexing of internal search results because those pages are considered search spam.[43]

In 2020, Google **sunsetting** the standard (and open-sourced their code) and now treats it as a hint rather than a directive. To adequately ensure that pages are not indexed, a page-level robot's meta tag should be included.[44]

## Increasing prominence

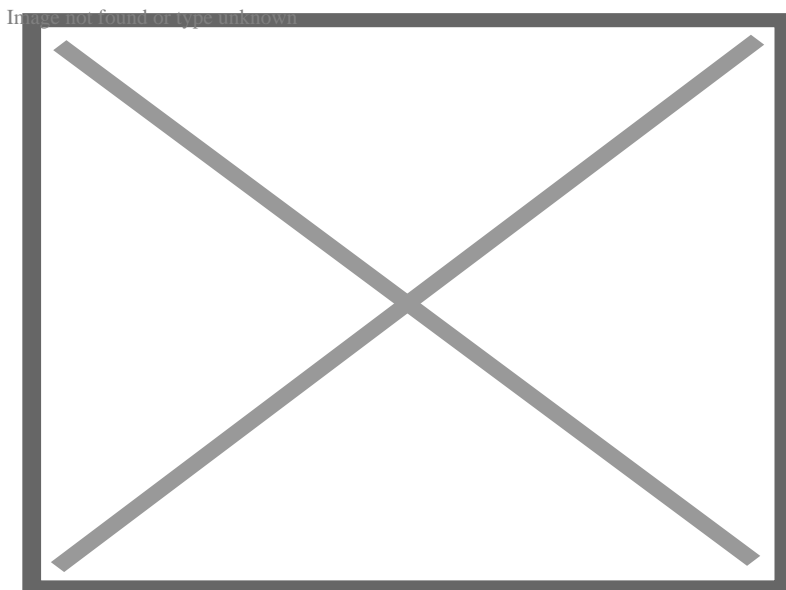
[**edit**]

A variety of methods can increase the prominence of a webpage within the search results. **Cross linking** between pages of the same website to provide more links to important pages may improve its visibility. Page design makes users trust a site and want to stay once they find it. When people bounce off a site, it counts against the site and affects its credibility.[45]

Writing content that includes frequently searched keyword phrases so as to be relevant to a wide variety of search queries will tend to increase traffic. Updating content so as to keep search engines crawling back frequently can give additional weight to a site. Adding relevant keywords to a web page's metadata, including the **title tag and meta description**, will tend to improve the relevancy of a site's search listings, thus increasing traffic. **URL canonicalization** of web pages accessible via multiple URLs, using the **canonical link element**[46] or via **301 redirects** can help make sure links to different versions of the URL all count towards the page's link popularity score. These are known as incoming links, which point to the URL and can count towards the page link's popularity score, impacting the credibility of a website.[45]

## White hat versus black hat techniques

[**edit**]



## Common white-hat methods of search engine optimization

SEO techniques can be classified into two broad categories: techniques that search engine companies recommend as part of good design ("white hat"), and those techniques of which search engines do not approve ("black hat"). Search engines attempt to minimize the effect of the latter, among them **spamdexing**. Industry commentators have classified these methods and the practitioners who employ them as either **white hat** SEO or **black hat** SEO.[47] White hats tend to produce results that last a long time, whereas black hats anticipate that their sites may eventually be banned either temporarily or permanently once the search engines discover what they are doing.[48]

An SEO technique is considered a white hat if it conforms to the search engines' guidelines and involves no deception. As the search engine guidelines[11][12][49] are not written as a series of rules or commandments, this is an important distinction to note. White hat SEO is not just about following guidelines but is about ensuring that the content a search engine indexes and subsequently ranks is the same content a user will see. White hat advice is generally summed up as creating content for users, not for search engines, and then making that content easily accessible to the online "spider" algorithms, rather than attempting to trick the algorithm from its intended purpose. White hat SEO is in many ways similar to web development that promotes accessibility,[50] although the two are not identical.

**Black hat SEO** attempts to improve rankings in ways that are disapproved of by the search engines or involve deception. One black hat technique uses hidden text, either as text colored similar to the background, in an invisible **div**, or positioned off-screen. Another method gives a different page depending on whether the page is being requested by a human visitor or a search engine, a technique known as **cloaking**. Another category sometimes used is **grey hat SEO**. This is in between the black hat and white hat approaches, where the methods employed avoid the site being penalized but do not act in producing the best content for users. Grey hat SEO is entirely focused on improving search engine rankings.

Search engines may penalize sites they discover using black or grey hat methods, either by reducing their rankings or eliminating their listings from their databases altogether. Such penalties can be applied either automatically by the search engines' algorithms or by a manual site review. One example was the February 2006 Google removal of both **BMW** Germany and **Ricoh** Germany for the use of deceptive practices.[51] Both companies subsequently apologized, fixed the offending pages, and were restored to Google's search engine results page.[52]

Companies that employ black hat techniques or other spammy tactics can get their client websites banned from the search results. In 2005, the *Wall Street Journal* reported on a company, **Traffic Power**, which allegedly used high-risk techniques and failed to disclose those risks to its clients.[53] *Wired* magazine reported that the same company sued blogger and SEO Aaron Wall for writing about the ban.[54] Google's **Matt Cutts** later confirmed that Google had banned Traffic Power and some of its clients.[55]

## As marketing strategy

[edit]

SEO is not an appropriate strategy for every website, and other Internet marketing strategies can be more effective, such as paid advertising through pay-per-click (PPC) campaigns, depending on the site operator's goals.<sup>[[editorializing](#)]</sup> Search engine marketing (SEM) is the practice of designing, running, and optimizing search engine ad campaigns. Its difference from SEO is most simply depicted as the difference between paid and unpaid priority ranking in search results. SEM focuses on prominence more so than relevance; website developers should regard SEM with the utmost importance with consideration to visibility as most navigate to the primary listings of their search.<sup>[56]</sup> A successful Internet marketing campaign may also depend upon building high-quality web pages to engage and persuade internet users, setting up analytics programs to enable site owners to measure results, and improving a site's conversion rate.<sup>[57][58]</sup> In November 2015, Google released a full 160-page version of its Search Quality Rating Guidelines to the public,<sup>[59]</sup> which revealed a shift in their focus towards "usefulness" and mobile local search. In recent years the mobile market has exploded, overtaking the use of desktops, as shown in by StatCounter in October 2016, where they analyzed 2.5 million websites and found that 51.3% of the pages were loaded by a mobile device.<sup>[60]</sup> Google has been one of the companies that are utilizing the popularity of mobile usage by encouraging websites to use their Google Search Console, the Mobile-Friendly Test, which allows companies to measure up their website to the search engine results and determine how user-friendly their websites are. The closer the keywords are together their ranking will improve based on key terms.<sup>[45]</sup>

SEO may generate an adequate return on investment. However, search engines are not paid for organic search traffic, their algorithms change, and there are no guarantees of continued referrals. Due to this lack of guarantee and uncertainty, a business that relies heavily on search engine traffic can suffer major losses if the search engines stop sending visitors.<sup>[61]</sup> Search engines can change their algorithms, impacting a website's search engine ranking, possibly resulting in a serious loss of traffic. According to Google's CEO, Eric Schmidt, in 2010, Google made over 500 algorithm changes – almost 1.5 per day.<sup>[62]</sup> It is considered a wise business practice for website operators to liberate themselves from dependence on search engine traffic.<sup>[63]</sup> In addition to accessibility in terms of web crawlers (addressed above), user web accessibility has become increasingly important for SEO.

## International markets and SEO

[edit]

Optimization techniques are highly tuned to the dominant search engines in the target market. The search engines' market shares vary from market to market, as does competition. In 2003, Danny Sullivan stated that Google represented about 75% of all searches.<sup>[64]</sup> In

markets outside the United States, Google's share is often larger, and data showed Google was the dominant search engine worldwide as of 2007.[65] As of 2006, Google had an 85–90% market share in Germany.[66] While there were hundreds of SEO firms in the US at that time, there were only about five in Germany.[66] As of March 2024, Google still had a significant market share of 89.85% in Germany.[67] As of June 2008, the market share of Google in the UK was close to 90% according to Hitwise.[68]<sup>[*obsolete source*]</sup> As of March 2024, Google's market share in the UK was 93.61%.[69]

Successful search engine optimization (SEO) for international markets requires more than just translating web pages. It may also involve registering a domain name with a **country-code top-level domain** (ccTLD) or a relevant **top-level domain** (TLD) for the target market, choosing web hosting with a local IP address or server, and using a **Content Delivery Network** (CDN) to improve website speed and performance globally. It is also important to understand the local culture so that the content feels relevant to the audience. This includes conducting keyword research for each market, using hreflang tags to target the right languages, and building local backlinks. However, the core SEO principles—such as creating high-quality content, improving user experience, and building links—remain the same, regardless of language or region.[66]

Regional search engines have a strong presence in specific markets:

- China: **Baidu** leads the market, controlling about 70 to 80% market share.[70]
- South Korea: Since the end of 2021, **Naver**, a domestic web portal, has gained prominence in the country.[71][72]
- Russia: **Yandex** is the leading search engine in Russia. As of December 2023, it accounted for at least 63.8% of the market share.[73]

## The Evolution of International SEO

[[edit](#)]

By the early 2000s, businesses recognized that the web and search engines could help them reach global audiences. As a result, the need for multilingual SEO emerged.[74] In the early years of international SEO development, simple translation was seen as sufficient. However, over time, it became clear that localization and transcreation—adapting content to local language, culture, and emotional resonance—were far more effective than basic translation.[75]

## Legal precedents

[[edit](#)]

On October 17, 2002, SearchKing filed suit in the [United States District Court](#), Western District of Oklahoma, against the search engine Google. SearchKing's claim was that Google's tactics to prevent spamdexing constituted a [tortious interference](#) with contractual relations. On May 27, 2003, the court granted Google's motion to dismiss the complaint because SearchKing "failed to state a claim upon which relief may be granted."[\[76\]](#)[\[77\]](#)

In March 2006, KinderStart filed a lawsuit against Google over search engine rankings. KinderStart's website was removed from Google's index prior to the lawsuit, and the amount of traffic to the site dropped by 70%. On March 16, 2007, the [United States District Court for the Northern District of California](#) ([San Jose](#) Division) dismissed KinderStart's complaint without leave to amend and partially granted Google's motion for [Rule 11](#) sanctions against KinderStart's attorney, requiring him to pay part of Google's legal expenses.[\[78\]](#)[\[79\]](#)

## See also

[\[edit\]](#)

- [Competitor backlinking](#)
- [List of search engines](#)
- [Search engine marketing](#)
- [Search neutrality](#), the opposite of search manipulation
- [User intent](#)
- [Website promotion](#)
- [Search engine results page](#)
- [Search engine scraping](#)

## References

[\[edit\]](#)

- ↑ "[SEO – search engine optimization](#)". *Webopedia*. December 19, 2001. *Archived from the original on May 9, 2019*. Retrieved May 9, 2019.
- ↑ Giomelakis, Dimitrios; Veglis, Andreas (April 2, 2016). "[Investigating Search Engine Optimization Factors in Media Websites: The case of Greece](#)". *Digital Journalism*. **4** (3): 379–400. doi:10.1080/21670811.2015.1046992. ISSN 2167-0811. S2CID 166902013. *Archived from the original on October 30, 2022*. Retrieved October 30, 2022.
- ↑ Beel, Jöran; Gipp, Bela; Wilde, Erik (2010). "[Academic Search Engine Optimization \(ASEO\): Optimizing Scholarly Literature for Google Scholar and Co](#)" (PDF). *Journal of Scholarly Publishing*. pp. 176–190. *Archived from the original* (PDF) on November 18, 2017. Retrieved April 18, 2010.
- ↑ Ortiz-Cordova, A. and Jansen, B. J. (2012) [Classifying Web Search Queries in Order to Identify High Revenue Generating Customers](#). *Archived* March 4, 2016, at the [Wayback Machine](#). *Journal of the American Society for Information Sciences and*



Technology. 63(7), 1426 – 1441.

5. ^ Brian Pinkerton. *"Finding What People Want: Experiences with the WebCrawler"* (PDF). The Second International WWW Conference Chicago, USA, October 17–20, 1994. *Archived* (PDF) from the original on May 8, 2007. Retrieved May 7, 2007.
6. ^ Danny Sullivan (June 14, 2004). *"Who Invented the Term "Search Engine Optimization"?"*. *Search Engine Watch*. Archived from *the original* on April 23, 2010. Retrieved May 14, 2007. See *Google groups thread Archived* June 17, 2013, at the *Wayback Machine*.
7. ^ "The Challenge is Open", *Brain vs Computer*, WORLD SCIENTIFIC, November 17, 2020, pp. 189–211, doi:10.1142/9789811225017\_0009, ISBN 978-981-12-2500-0, S2CID 243130517
8. ^ Pringle, G.; Allison, L.; Dowe, D. (April 1998). *"What is a tall poppy among web pages?"*. Monash University. Archived from *the original* on April 27, 2007. Retrieved May 8, 2007.
9. ^ Laurie J. Flynn (November 11, 1996). *"Desperately Seeking Surfers"*. *New York Times*. *Archived* from the original on October 30, 2007. Retrieved May 9, 2007.
10. ^ Jason Demers (January 20, 2016). *"Is Keyword Density Still Important for SEO"*. *Forbes*. *Archived* from the original on August 16, 2016. Retrieved August 15, 2016.
11. ^ **a b** *"Google's Guidelines on Site Design"*. *Archived* from the original on January 9, 2009. Retrieved April 18, 2007.
12. ^ **a b** *"Bing Webmaster Guidelines"*. *bing.com*. *Archived* from the original on September 9, 2014. Retrieved September 11, 2014.
13. ^ *"Sitemaps"*. *Archived* from the original on June 22, 2023. Retrieved July 4, 2012.
14. ^ *"By the Data: For Consumers, Mobile is the Internet" Google for Entrepreneurs Startup Grind September 20, 2015*. *Archived* from the original on January 6, 2016. Retrieved January 8, 2016.
15. ^ Brin, Sergey & Page, Larry (1998). *"The Anatomy of a Large-Scale Hypertextual Web Search Engine"*. Proceedings of the seventh international conference on World Wide Web. pp. 107–117. *Archived* from the original on October 10, 2006. Retrieved May 8, 2007.
16. ^ *"Co-founders of Google - Google's co-founders may not have the name recognition of say, Bill Gates, but give them time: Google hasn't been around nearly as long as Microsoft"*. *Entrepreneur*. October 15, 2008. *Archived* from the original on May 31, 2014. Retrieved May 30, 2014.
17. ^ Thompson, Bill (December 19, 2003). *"Is Google good for you?"*. *BBC News*. *Archived* from the original on January 25, 2009. Retrieved May 16, 2007.
18. ^ Zoltan Gyongyi & Hector Garcia-Molina (2005). *"Link Spam Alliances"* (PDF). Proceedings of the 31st VLDB Conference, Trondheim, Norway. *Archived* (PDF) from the original on June 12, 2007. Retrieved May 9, 2007.
19. ^ Hansell, Saul (June 3, 2007). *"Google Keeps Tweaking Its Search Engine"*. *New York Times*. *Archived* from the original on November 10, 2017. Retrieved June 6, 2007.
20. ^ Sullivan, Danny (September 29, 2005). *"Rundown On Search Ranking Factors"*. *Search Engine Watch*. *Archived* from the original on May 28, 2007. Retrieved May 8,

2007.

21. ^ Christine Churchill (November 23, 2005). *"Understanding Search Engine Patents"*. *Search Engine Watch*. Archived from *the original* on February 7, 2007. Retrieved May 8, 2007.
22. ^ *"Google Personalized Search Leaves Google Labs"*. *searchenginewatch.com*. Search Engine Watch. Archived from *the original* on January 25, 2009. Retrieved September 5, 2009.
23. ^ *"8 Things We Learned About Google PageRank"*. *www.searchenginejournal.com*. October 25, 2007. Archived from the original on August 19, 2009. Retrieved August 17, 2009.
24. ^ *"PageRank sculpting"*. Matt Cutts. Archived from the original on January 6, 2010. Retrieved January 12, 2010.
25. ^ *"Google Loses 'Backwards Compatibility' On Paid Link Blocking & PageRank Sculpting"*. *searchengineland.com*. June 3, 2009. Archived from the original on August 14, 2009. Retrieved August 17, 2009.
26. ^ *"Personalized Search for everyone"*. Archived from the original on December 8, 2009. Retrieved December 14, 2009.
27. ^ *"Our new search index: Caffeine"*. Google: Official Blog. Archived from the original on June 18, 2010. Retrieved May 10, 2014.
28. ^ *"Relevance Meets Real-Time Web"*. *Google Blog*. Archived from the original on April 7, 2019. Retrieved January 4, 2010.
29. ^ *"Google Search Quality Updates"*. *Google Blog*. Archived from the original on April 23, 2022. Retrieved March 21, 2012.
30. ^ *"What You Need to Know About Google's Penguin Update"*. *Inc.com*. June 20, 2012. Archived from the original on December 20, 2012. Retrieved December 6, 2012.
31. ^ *"Google Penguin looks mostly at your link source, says Google"*. Search Engine Land. October 10, 2016. Archived from the original on April 21, 2017. Retrieved April 20, 2017.
32. ^ *"FAQ: All About The New Google 'Hummingbird' Algorithm"*. *www.searchengineland.com*. September 26, 2013. Archived from the original on December 23, 2018. Retrieved March 17, 2018.
33. ^ *"Understanding searches better than ever before"*. Google. October 25, 2019. Archived from the original on January 27, 2021. Retrieved May 12, 2020.
34. ^ *"Submitting To Directories: Yahoo & The Open Directory"*. *Search Engine Watch*. March 12, 2007. Archived from *the original* on May 19, 2007. Retrieved May 15, 2007.
35. ^ *"What is a Sitemap file and why should I have one?"*. Archived from the original on July 1, 2007. Retrieved March 19, 2007.
36. ^ *"Search Console - Crawl URL"*. Archived from the original on August 14, 2022. Retrieved December 18, 2015.
37. ^ Sullivan, Danny (March 12, 2007). *"Submitting To Search Crawlers: Google, Yahoo, Ask & Microsoft's Live Search"*. *Search Engine Watch*. Archived from *the original* on May 10, 2007. Retrieved May 15, 2007.
38. ^ Cho, J.; Garcia-Molina, H.; Page, L. (1998). *"Efficient crawling through URL ordering"*. Seventh International World-Wide Web Conference. Brisbane, Australia: Stanford

InfoLab Publication Server. Archived from [the original](#) on July 14, 2019. Retrieved May 9, 2007.

39. ^ ["Mobile-first Index"](#). Archived from the original on February 22, 2019. Retrieved March 19, 2018.
40. ^ Phan, Doantam (November 4, 2016). ["Mobile-first Indexing"](#). Official Google Webmaster Central Blog. Archived from the original on February 22, 2019. Retrieved January 16, 2019.
41. ^ ["The new evergreen Googlebot"](#). Official Google Webmaster Central Blog. Archived from the original on November 6, 2020. Retrieved March 2, 2020.
42. ^ ["Updating the user agent of Googlebot"](#). Official Google Webmaster Central Blog. Archived from the original on March 2, 2020. Retrieved March 2, 2020.
43. ^ ["Newspapers Amok! New York Times Spamming Google? LA Times Hijacking Cars.com?"](#). Search Engine Land. May 8, 2007. Archived from the original on December 26, 2008. Retrieved May 9, 2007.
44. ^ Jill Kocher Brown (February 24, 2020). ["Google Downgrades Nofollow Directive. Now What?"](#). Practical Ecommerce. Archived from the original on January 25, 2021. Retrieved February 11, 2021.
45. ^ [a b c](#) Morey, Sean (2008). The Digital Writer. Fountainhead Press. pp. 171–187.
46. ^ ["Bing – Partnering to help solve duplicate content issues – Webmaster Blog – Bing Community"](#). www.bing.com. February 12, 2009. Archived from the original on June 7, 2014. Retrieved October 30, 2009.
47. ^ Andrew Goodman. ["Search Engine Showdown: Black hats vs. White hats at SES"](#). SearchEngineWatch. Archived from [the original](#) on February 22, 2007. Retrieved May 9, 2007.
48. ^ [Jill Whalen](#) (November 16, 2004). ["Black Hat/White Hat Search Engine Optimization"](#). searchengineguide.com. Archived from [the original](#) on November 17, 2004. Retrieved May 9, 2007.
49. ^ ["What's an SEO? Does Google recommend working with companies that offer to make my site Google-friendly?"](#). Archived from the original on April 16, 2006. Retrieved April 18, 2007.
50. ^ Andy Hagans (November 8, 2005). ["High Accessibility Is Effective Search Engine Optimization"](#). A List Apart. Archived from the original on May 4, 2007. Retrieved May 9, 2007.
51. ^ [Matt Cutts](#) (February 4, 2006). ["Ramping up on international webspam"](#). mattcutts.com/blog. Archived from the original on June 29, 2012. Retrieved May 9, 2007.
52. ^ [Matt Cutts](#) (February 7, 2006). ["Recent reinclusions"](#). mattcutts.com/blog. Archived from the original on May 22, 2007. Retrieved May 9, 2007.
53. ^ David Kesmodel (September 22, 2005). ["Sites Get Dropped by Search Engines After Trying to 'Optimize' Rankings"](#). Wall Street Journal. Archived from the original on August 4, 2020. Retrieved July 30, 2008.
54. ^ Adam L. Penenberg (September 8, 2005). ["Legal Showdown in Search Fracas"](#). Wired Magazine. Archived from the original on March 4, 2016. Retrieved August 11, 2016.



55. ^ [Matt Cutts](#) (February 2, 2006). ["Confirming a penalty"](#). [mattcutts.com/blog](#). Archived from the original on June 26, 2012. Retrieved May 9, 2007.
56. ^ Tapan, Panda (2013). "Search Engine Marketing: Does the Knowledge Discovery Process Help Online Retailers?". *IUP Journal of Knowledge Management*. **11** (3): 56–66. [ProQuest 1430517207](#).
57. ^ Melissa Burdon (March 13, 2007). ["The Battle Between Search Engine Optimization and Conversion: Who Wins?"](#). [Grok.com](#). Archived from the original on March 15, 2008. Retrieved April 10, 2017.
58. ^ ["SEO Tips and Marketing Strategies"](#). Archived from the original on October 30, 2022. Retrieved October 30, 2022.
59. ^ [""Search Quality Evaluator Guidelines" How Search Works November 12, 2015"](#) (PDF). Archived (PDF) from the original on March 29, 2019. Retrieved January 11, 2016.
60. ^ Titcomb, James (November 2016). ["Mobile web usage overtakes desktop for first time"](#). *The Telegraph*. Archived from the original on January 10, 2022. Retrieved March 17, 2018.
61. ^ Andy Greenberg (April 30, 2007). ["Condemned To Google Hell"](#). *Forbes*. Archived from the original on May 2, 2007. Retrieved May 9, 2007.
62. ^ Matt McGee (September 21, 2011). ["Schmidt's testimony reveals how Google tests algorithm changes"](#). Archived from the original on January 17, 2012. Retrieved January 4, 2012.
63. ^ [Jakob Nielsen](#) (January 9, 2006). ["Search Engines as Leeches on the Web"](#). [useit.com](#). Archived from the original on August 25, 2012. Retrieved May 14, 2007.
64. ^ Graham, Jefferson (August 26, 2003). ["The search engine that could"](#). *USA Today*. Archived from the original on May 17, 2007. Retrieved May 15, 2007.
65. ^ Greg Jarboe (February 22, 2007). ["Stats Show Google Dominates the International Search Landscape"](#). *Search Engine Watch*. Archived from the original on May 23, 2011. Retrieved May 15, 2007.
66. ^ [a b c](#) Mike Grehan (April 3, 2006). ["Search Engine Optimizing for Europe"](#). *Click*. Archived from the original on November 6, 2010. Retrieved May 14, 2007.
67. ^ ["Germany search engine market share 2024"](#). Statista. Retrieved January 6, 2025.
68. ^ Jack Schofield (June 10, 2008). ["Google UK closes in on 90% market share"](#). *Guardian*. London. Archived from the original on December 17, 2013. Retrieved June 10, 2008.
69. ^ ["UK search engines market share 2024"](#). Statista. Retrieved January 6, 2025.
70. ^ ["China search engines market share 2024"](#). Statista. Retrieved January 6, 2025.
71. ^ cycles, This text provides general information Statista assumes no liability for the information given being complete or correct Due to varying update; Text, Statistics Can Display More up-to-Date Data Than Referenced in the. ["Topic: Search engines in South Korea"](#). Statista. Retrieved January 6, 2025.
72. ^ ["South Korea: main service used to search for information 2024"](#). Statista. Retrieved January 6, 2025.
73. ^ ["Most popular search engines in Russia 2023"](#). Statista. Retrieved January 6, 2025.

74. <sup>^</sup> Arora, Sanjog; Hemrajani, Naveen (September 2023). *"A REVIEW ON: MULTILINGUAL SEARCH TECHNIQUE"*. *International Journal of Applied Engineering & Technology*. **5** (3): 760–770 – via ResearchGate.
75. <sup>^</sup> *"SEO Starter Guide: The Basics | Google Search Central | Documentation"*. Google for Developers. Retrieved January 13, 2025.
76. <sup>^</sup> *"Search King, Inc. v. Google Technology, Inc., CIV-02-1457-M"* (PDF). docstoc.com. May 27, 2003. *Archived* from the original on May 27, 2008. Retrieved May 23, 2008.
77. <sup>^</sup> Stefanie Olsen (May 30, 2003). *"Judge dismisses suit against Google"*. CNET. *Archived* from the original on December 1, 2010. Retrieved May 10, 2007.
78. <sup>^</sup> *"Technology & Marketing Law Blog: KinderStart v. Google Dismissed—With Sanctions Against KinderStart's Counsel"*. blog.ericgoldman.org. March 20, 2007. *Archived* from the original on May 11, 2008. Retrieved June 23, 2008.
79. <sup>^</sup> *"Technology & Marketing Law Blog: Google Sued Over Rankings—KinderStart.com v. Google"*. blog.ericgoldman.org. *Archived* from the original on June 22, 2008. Retrieved June 23, 2008.

## External links

[[edit](#)]

Listen to this article (22 minutes)

## Spoken Wikipedia icon

Image not found or type unknown

**This audio file** was created from a revision of this article dated 20 May 2008, and does not reflect subsequent edits.

([Audio help](#) · [More spoken articles](#))

- [Webmaster Guidelines](#) from Google
- [Google Search Quality Evaluators Guidelines](#) (PDF)
- [Webmaster resources](#) from Yahoo!
- [Webmaster Guidelines](#) from Microsoft Bing
- [The Dirty Little Secrets of Search](#) in *The New York Times* (February 12, 2011)

- **v**
- **t**
- **e**

[Search engine optimization](#)

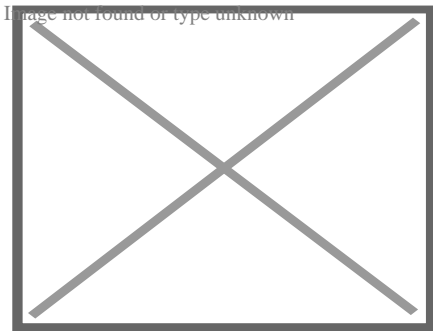
<b>Exclusion standards</b>	<ul style="list-style-type: none"><li>○ Robots exclusion standard</li><li>○ Meta element</li><li>○ nofollow</li></ul>
<b>Marketing topics</b>	<ul style="list-style-type: none"><li>○ Online advertising</li><li>○ Email marketing</li><li>○ Display advertising</li><li>○ Web analytics</li></ul>
<b>Search marketing</b>	<ul style="list-style-type: none"><li>○ Search engine marketing</li><li>○ Social media optimization</li><li>○ Online identity management</li><li>○ Paid inclusion</li><li>○ Pay per click</li><li>○ Google bomb</li></ul>
<b>Search engine spam</b>	<ul style="list-style-type: none"><li>○ Spamdexing</li><li>○ Web scraping</li><li>○ Scraper site</li><li>○ Link farm</li><li>○ Link building</li></ul>
<b>Linking</b>	<ul style="list-style-type: none"><li>○ Backlink</li><li>○ Link building</li><li>○ Link exchange</li><li>○ Organic linking</li></ul>
<b>People</b>	<ul style="list-style-type: none"><li>○ Danny Sullivan</li><li>○ Matt Cutts</li><li>○ Barry Schwartz</li></ul>
<b>Other</b>	<ul style="list-style-type: none"><li>○ Geotargeting</li><li>○ Human search engine</li><li>○ Stop words</li><li>○ Content farm</li></ul>



- Germany

- Poland

This article is about the internet bot. For the search engine, see [WebCrawler](#). "Web spider" redirects here; not to be confused with [Spider web](#). "Spiderbot" redirects here. For the video game, see [Arac \(video game\)](#).



A **Web crawler**, sometimes called a **spider** or **spiderbot** and often shortened to **crawler**, is an **Internet bot** that systematically browses the **World Wide Web** and that is typically operated by search engines for the purpose of **Web indexing** (*web spidering*).<sup>[1]</sup>

Web **search engines** and some other **websites** use Web crawling or spidering **software** to update their **web content** or indices of other sites' web content. Web crawlers copy pages for processing by a search engine, which **indexes** the downloaded pages so that users can search more efficiently.

Crawlers consume resources on visited systems and often visit sites unprompted. Issues of schedule, load, and "politeness" come into play when large collections of pages are accessed. Mechanisms exist for public sites not wishing to be crawled to make this known to the crawling agent. For example, including a **robots.txt** file can request **bots** to index only parts of a website, or nothing at all.

The number of Internet pages is extremely large; even the largest crawlers fall short of making a complete index. For this reason, search engines struggled to give relevant search results in the early years of the World Wide Web, before 2000. Today, relevant results are

given almost instantly.

Crawlers can validate [hyperlinks](#) and [HTML](#) code. They can also be used for [web scraping](#) and [data-driven programming](#).

## Nomenclature

[\[edit\]](#)

A web crawler is also known as a *spider*,<sup>[2]</sup> an *ant*, an *automatic indexer*,<sup>[3]</sup> or (in the [FOAF](#) software context) a *Web scutter*.<sup>[4]</sup>

## Overview

[\[edit\]](#)

A Web crawler starts with a list of [URLs](#) to visit. Those first URLs are called the *seeds*. As the crawler visits these URLs, by communicating with [web servers](#) that respond to those URLs, it identifies all the [hyperlinks](#) in the retrieved web pages and adds them to the list of URLs to visit, called the *crawl frontier*. URLs from the frontier are [recursively](#) visited according to a set of policies. If the crawler is performing archiving of [websites](#) (or [web archiving](#)), it copies and saves the information as it goes. The archives are usually stored in such a way they can be viewed, read and navigated as if they were on the live web, but are preserved as 'snapshots'.<sup>[5]</sup>

The archive is known as the *repository* and is designed to store and manage the collection of [web pages](#). The [repository](#) only stores [HTML](#) pages and these pages are stored as distinct files. A repository is similar to any other system that stores data, like a modern-day database. The only difference is that a repository does not need all the functionality offered by a database system. The repository stores the most recent version of the web page retrieved by the crawler.<sup>[[citation needed](#)]</sup>

The large volume implies the crawler can only download a limited number of the Web pages within a given time, so it needs to prioritize its downloads. The high rate of change can imply the pages might have already been updated or even deleted.

The number of possible URLs crawled being generated by server-side software has also made it difficult for web crawlers to avoid retrieving [duplicate content](#). Endless combinations of [HTTP](#) GET (URL-based) parameters exist, of which only a small selection will actually return unique content. For example, a simple online photo gallery may offer three options to users, as specified through HTTP GET parameters in the URL. If there exist four ways to sort images, three choices of [thumbnail](#) size, two file formats, and an option to disable user-provided content, then the same set of content can be accessed with 48 different URLs, all of which may be linked on the site. This [mathematical combination](#) creates a problem for

crawlers, as they must sort through endless combinations of relatively minor scripted changes in order to retrieve unique content.

As Edwards *et al.* noted, "Given that the **bandwidth** for conducting crawls is neither infinite nor free, it is becoming essential to crawl the Web in not only a scalable, but efficient way, if some reasonable measure of quality or freshness is to be maintained." [6] A crawler must carefully choose at each step which pages to visit next.

## Crawling policy

[edit]

The behavior of a Web crawler is the outcome of a combination of policies: [7]

- a *selection policy* which states the pages to download,
- a *re-visit policy* which states when to check for changes to the pages,
- a *politeness policy* that states how to avoid overloading **websites**.
- a *parallelization policy* that states how to coordinate distributed web crawlers.

## Selection policy

[edit]

Given the current size of the Web, even large search engines cover only a portion of the publicly available part. A 2009 study showed even large-scale **search engines** index no more than 40–70% of the indexable Web; [8] a previous study by **Steve Lawrence** and **Lee Giles** showed that no **search engine indexed** more than 16% of the Web in 1999. [9] As a crawler always downloads just a fraction of the **Web pages**, it is highly desirable for the downloaded fraction to contain the most relevant pages and not just a random sample of the Web.

This requires a metric of importance for prioritizing Web pages. The importance of a page is a function of its **intrinsic** quality, its popularity in terms of links or visits, and even of its URL (the latter is the case of **vertical search engines** restricted to a single **top-level domain**, or search engines restricted to a fixed Web site). Designing a good selection policy has an added difficulty: it must work with partial information, as the complete set of Web pages is not known during crawling.

Junghoo Cho *et al.* made the first study on policies for crawling scheduling. Their data set was a 180,000-pages crawl from the stanford.edu domain, in which a crawling simulation was done with different strategies. [10] The ordering metrics tested were **breadth-first**, **backlink** count and partial **PageRank** calculations. One of the conclusions was that if the crawler wants to download pages with high Pagerank early during the crawling process, then the partial Pagerank strategy is the better, followed by breadth-first and backlink-count. However, these

results are for just a single domain. Cho also wrote his PhD dissertation at Stanford on web crawling.[11]

Najork and Wiener performed an actual crawl on 328 million pages, using breadth-first ordering.[12] They found that a breadth-first crawl captures pages with high Pagerank early in the crawl (but they did not compare this strategy against other strategies). The explanation given by the authors for this result is that "the most important pages have many links to them from numerous hosts, and those links will be found early, regardless of on which host or page the crawl originates."

Abiteboul designed a crawling strategy based on an **algorithm** called OPIC (On-line Page Importance Computation).[13] In OPIC, each page is given an initial sum of "cash" that is distributed equally among the pages it points to. It is similar to a PageRank computation, but it is faster and is only done in one step. An OPIC-driven crawler downloads first the pages in the crawling frontier with higher amounts of "cash". Experiments were carried in a 100,000-pages synthetic graph with a power-law distribution of in-links. However, there was no comparison with other strategies nor experiments in the real Web.

Boldi *et al.* used simulation on subsets of the Web of 40 million pages from the .it domain and 100 million pages from the WebBase crawl, testing breadth-first against depth-first, random ordering and an omniscient strategy. The comparison was based on how well PageRank computed on a partial crawl approximates the true PageRank value. Some visits that accumulate PageRank very quickly (most notably, breadth-first and the omniscient visit) provide very poor progressive approximations.[14][15]

Baeza-Yates *et al.* used simulation on two subsets of the Web of 3 million pages from the .gr and .cl domain, testing several crawling strategies.[16] They showed that both the OPIC strategy and a strategy that uses the length of the per-site queues are better than **breadth-first** crawling, and that it is also very effective to use a previous crawl, when it is available, to guide the current one.

Daneshpajouh *et al.* designed a community based algorithm for discovering good seeds.[17] Their method crawls web pages with high PageRank from different communities in less iteration in comparison with crawl starting from random seeds. One can extract good seed from a previously-crawled-Web graph using this new method. Using these seeds, a new crawl can be very effective.

## Restricting followed links

[edit]

A crawler may only want to seek out HTML pages and avoid all other **MIME types**. In order to request only HTML resources, a crawler may make an HTTP HEAD request to determine a

Web resource's MIME type before requesting the entire resource with a GET request. To avoid making numerous HEAD requests, a crawler may examine the URL and only request a resource if the URL ends with certain characters such as .html, .htm, .asp, .aspx, .php, .jsp, .jspx or a slash. This strategy may cause numerous HTML Web resources to be unintentionally skipped.

Some crawlers may also avoid requesting any resources that have a "?" in them (are dynamically produced) in order to avoid **spider traps** that may cause the crawler to download an infinite number of URLs from a Web site. This strategy is unreliable if the site uses **URL rewriting** to simplify its URLs.

## URL normalization

[[edit](#)]

Main article: [URL normalization](#)

Crawlers usually perform some type of **URL normalization** in order to avoid crawling the same resource more than once. The term *URL normalization*, also called *URL canonicalization*, refers to the process of modifying and standardizing a URL in a consistent manner. There are several types of normalization that may be performed including conversion of URLs to lowercase, removal of "." and ".." segments, and adding trailing slashes to the non-empty path component.<sup>[18]</sup>

## Path-ascending crawling

[[edit](#)]

Some crawlers intend to download/upload as many resources as possible from a particular web site. So *path-ascending crawler* was introduced that would ascend to every path in each URL that it intends to crawl.<sup>[19]</sup> For example, when given a seed URL of `http://llama.org/hamster/monkey/page.html`, it will attempt to crawl `/hamster/monkey/`, `/hamster/`, and `/`. Cothey found that a path-ascending crawler was very effective in finding isolated resources, or resources for which no inbound link would have been found in regular crawling.

## Focused crawling

[[edit](#)]

Main article: [Focused crawler](#)

The importance of a page for a crawler can also be expressed as a function of the similarity of a page to a given query. Web crawlers that attempt to download pages that are similar to each other are called **focused crawler** or **topical crawlers**. The concepts of topical and focused crawling were first introduced by Filippo Menczer[20][21] and by Soumen Chakrabarti *et al.*[22]

The main problem in focused crawling is that in the context of a Web crawler, we would like to be able to predict the similarity of the text of a given page to the query before actually downloading the page. A possible predictor is the anchor text of links; this was the approach taken by Pinkerton[23] in the first web crawler of the early days of the Web. Diligenti *et al.*[24] propose using the complete content of the pages already visited to infer the similarity between the driving query and the pages that have not been visited yet. The performance of a focused crawling depends mostly on the richness of links in the specific topic being searched, and a focused crawling usually relies on a general Web search engine for providing starting points.

## Academic focused crawler

[edit]

An example of the **focused crawlers** are academic crawlers, which crawls free-access academic related documents, such as the *citeseerxbot*, which is the crawler of **CiteSeer<sup>X</sup>** search engine. Other academic search engines are **Google Scholar** and **Microsoft Academic Search** etc. Because most academic papers are published in **PDF** formats, such kind of crawler is particularly interested in crawling **PDF**, **PostScript** files, **Microsoft Word** including their **zipped** formats. Because of this, general open-source crawlers, such as **Heritrix**, must be customized to filter out other **MIME types**, or a **middleware** is used to extract these documents out and import them to the focused crawl database and repository.[25] Identifying whether these documents are academic or not is challenging and can add a significant overhead to the crawling process, so this is performed as a post crawling process using **machine learning** or **regular expression** algorithms. These academic documents are usually obtained from home pages of faculties and students or from publication page of research institutes. Because academic documents make up only a small fraction of all web pages, a good seed selection is important in boosting the efficiencies of these web crawlers.[26] Other academic crawlers may download plain text and **HTML** files, that contains **metadata** of academic papers, such as titles, papers, and abstracts. This increases the overall number of papers, but a significant fraction may not provide free PDF downloads.

## Semantic focused crawler



[edit]

Another type of focused crawlers is semantic focused crawler, which makes use of **domain ontologies** to represent topical maps and link Web pages with relevant ontological concepts for the selection and categorization purposes.[27] In addition, ontologies can be automatically updated in the crawling process. Dong et al.[28] introduced such an ontology-learning-based crawler using a **support-vector machine** to update the content of ontological concepts when crawling Web pages.

## Re-visit policy

[edit]

The Web has a very dynamic nature, and crawling a fraction of the Web can take weeks or months. By the time a Web crawler has finished its crawl, many events could have happened, including creations, updates, and deletions.

From the search engine's point of view, there is a cost associated with not detecting an event, and thus having an outdated copy of a resource. The most-used cost functions are freshness and age.[29]

**Freshness:** This is a binary measure that indicates whether the local copy is accurate or not. The freshness of a page  $p$  in the repository at time  $t$  is defined as:

$$F_p(t) = \begin{cases} 1 & \text{if } p \text{ is equal to the local copy at time } t \\ 0 & \text{otherwise} \end{cases}$$

Image not found or type unknown

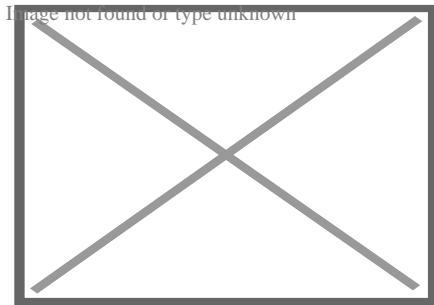
**Age:** This is a measure that indicates how outdated the local copy is. The age of a page  $p$  in the repository, at time  $t$  is defined as:

$$A_p(t) = \begin{cases} 0 & \text{if } p \text{ is not modified at time } t \\ t - \text{modification time} & \text{otherwise} \end{cases}$$

Image not found or type unknown

**Coffman et al.** worked with a definition of the objective of a Web crawler that is equivalent to freshness, but use a different wording: they propose that a crawler must minimize the fraction of time pages remain outdated. They also noted that the problem of Web crawling can be modeled as a multiple-queue, single-server polling system, on which the Web crawler is the server and the Web sites are the queues. Page modifications are the arrival of the customers, and switch-over times are the interval between page accesses to a single Web site. Under this model, mean waiting time for a customer in the polling system is equivalent to the average age for the Web crawler.[30]

The objective of the crawler is to keep the average freshness of pages in its collection as high as possible, or to keep the average age of pages as low as possible. These objectives are not equivalent: in the first case, the crawler is just concerned with how many pages are outdated, while in the second case, the crawler is concerned with how old the local copies of pages are.



Evolution of Freshness and Age in a web crawler

Two simple re-visiting policies were studied by Cho and Garcia-Molina: [31]

- Uniform policy: This involves re-visiting all pages in the collection with the same frequency, regardless of their rates of change.
- Proportional policy: This involves re-visiting more often the pages that change more frequently. The visiting frequency is directly proportional to the (estimated) change frequency.

In both cases, the repeated crawling order of pages can be done either in a random or a fixed order.

Cho and Garcia-Molina proved the surprising result that, in terms of average freshness, the uniform policy outperforms the proportional policy in both a simulated Web and a real Web crawl. Intuitively, the reasoning is that, as web crawlers have a limit to how many pages they can crawl in a given time frame, (1) they will allocate too many new crawls to rapidly changing pages at the expense of less frequently updating pages, and (2) the freshness of rapidly changing pages lasts for shorter period than that of less frequently changing pages. In other words, a proportional policy allocates more resources to crawling frequently updating pages, but experiences less overall freshness time from them.

To improve freshness, the crawler should penalize the elements that change too often. [32] The optimal re-visiting policy is neither the uniform policy nor the proportional policy. The optimal method for keeping average freshness high includes ignoring the pages that change too often, and the optimal for keeping average age low is to use access frequencies that monotonically (and sub-linearly) increase with the rate of change of each page. In both cases, the optimal is closer to the uniform policy than to the proportional policy: as Coffman *et al.* note, "in order to minimize the expected obsolescence time, the accesses to any particular page should be kept as evenly spaced as possible". [30] Explicit formulas for the re-visit policy are not attainable in general, but they are obtained numerically, as they depend on the

distribution of page changes. Cho and Garcia-Molina show that the exponential distribution is a good fit for describing page changes,[32] while Ipeirotis *et al.* show how to use statistical tools to discover parameters that affect this distribution.[33] The re-visiting policies considered here regard all pages as homogeneous in terms of quality ("all pages on the Web are worth the same"), something that is not a realistic scenario, so further information about the Web page quality should be included to achieve a better crawling policy.

## Politeness policy

[edit]

Crawlers can retrieve data much quicker and in greater depth than human searchers, so they can have a crippling impact on the performance of a site. If a single crawler is performing multiple requests per second and/or downloading large files, a server can have a hard time keeping up with requests from multiple crawlers.

As noted by Koster, the use of Web crawlers is useful for a number of tasks, but comes with a price for the general community.[34] The costs of using Web crawlers include:

- network resources, as crawlers require considerable bandwidth and operate with a high degree of parallelism during a long period of time;
- server overload, especially if the frequency of accesses to a given server is too high;
- poorly written crawlers, which can crash servers or routers, or which download pages they cannot handle; and
- personal crawlers that, if deployed by too many users, can disrupt networks and Web servers.

A partial solution to these problems is the robots exclusion protocol, also known as the robots.txt protocol that is a standard for administrators to indicate which parts of their Web servers should not be accessed by crawlers.[35] This standard does not include a suggestion for the interval of visits to the same server, even though this interval is the most effective way of avoiding server overload. Recently commercial search engines like Google, Ask Jeeves, MSN and Yahoo! Search are able to use an extra "Crawl-delay:" parameter in the robots.txt file to indicate the number of seconds to delay between requests.

The first proposed interval between successive pageloads was 60 seconds.[36] However, if pages were downloaded at this rate from a website with more than 100,000 pages over a perfect connection with zero latency and infinite bandwidth, it would take more than 2 months to download only that entire Web site; also, only a fraction of the resources from that Web server would be used.

Cho uses 10 seconds as an interval for accesses,[31] and the WIRE crawler uses 15 seconds as the default.[37] The MercatorWeb crawler follows an adaptive politeness policy: if it took  $t$  seconds to download a document from a given server, the crawler waits for  $10t$

seconds before downloading the next page.[38] Dill *et al.* use 1 second.[39]

For those using Web crawlers for research purposes, a more detailed cost-benefit analysis is needed and ethical considerations should be taken into account when deciding where to crawl and how fast to crawl.[40]

Anecdotal evidence from access logs shows that access intervals from known crawlers vary between 20 seconds and 3–4 minutes. It is worth noticing that even when being very polite, and taking all the safeguards to avoid overloading Web servers, some complaints from Web server administrators are received. Sergey Brin and Larry Page noted in 1998, "... running a crawler which connects to more than half a million servers ... generates a fair amount of e-mail and phone calls. Because of the vast number of people coming on line, there are always those who do not know what a crawler is, because this is the first one they have seen." [41]

## Parallelization policy

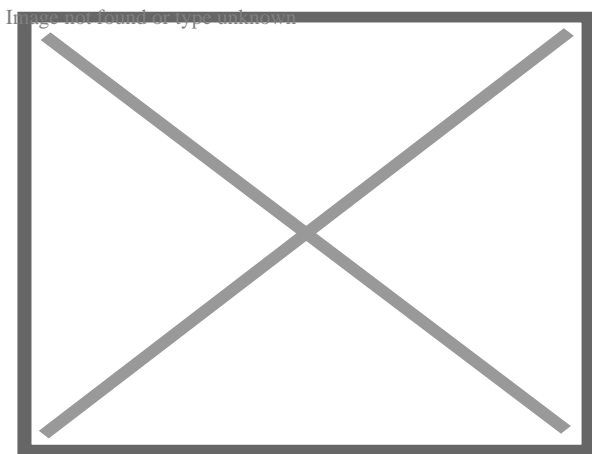
[edit]

Main article: [Distributed web crawling](#)

A **parallel** crawler is a crawler that runs multiple processes in parallel. The goal is to maximize the download rate while minimizing the overhead from parallelization and to avoid repeated downloads of the same page. To avoid downloading the same page more than once, the crawling system requires a policy for assigning the new URLs discovered during the crawling process, as the same URL can be found by two different crawling processes.

## Architectures

[edit]



High-level architecture of a standard Web crawler

A crawler must not only have a good crawling strategy, as noted in the previous sections, but it should also have a highly optimized architecture.

Shkapenyuk and Suel noted that:[42]

While it is fairly easy to build a slow crawler that downloads a few pages per second for a short period of time, building a high-performance system that can download hundreds of millions of pages over several weeks presents a number of challenges in system design, I/O and network efficiency, and robustness and manageability.

Web crawlers are a central part of search engines, and details on their algorithms and architecture are kept as business secrets. When crawler designs are published, there is often an important lack of detail that prevents others from reproducing the work. There are also emerging concerns about "[search engine spamming](#)", which prevent major search engines from publishing their ranking algorithms.

## Security

[\[edit\]](#)

While most of the website owners are keen to have their pages indexed as broadly as possible to have strong presence in [search engines](#), web crawling can also have [unintended consequences](#) and lead to a [compromise](#) or [data breach](#) if a search engine indexes resources that should not be publicly available, or pages revealing potentially vulnerable versions of software.

Main article: [Google hacking](#)

Apart from standard [web application security](#) recommendations website owners can reduce their exposure to opportunistic hacking by only allowing search engines to index the public parts of their websites (with [robots.txt](#)) and explicitly blocking them from indexing transactional parts (login pages, private pages, etc.).

## Crawler identification

[\[edit\]](#)

Web crawlers typically identify themselves to a Web server by using the [User-agent](#) field of an [HTTP](#) request. Web site administrators typically examine their [Web servers'](#) log and use the user agent field to determine which crawlers have visited the web server and how often. The user agent field may include a [URL](#) where the Web site administrator may find out more information about the crawler. Examining Web server log is tedious task, and therefore some administrators use tools to identify, track and verify Web crawlers. [Spambots](#) and other malicious Web crawlers are unlikely to place identifying information in the user agent field, or they may mask their identity as a browser or other well-known crawler.

Web site administrators prefer Web crawlers to identify themselves so that they can contact the owner if needed. In some cases, crawlers may be accidentally trapped in a **crawler trap** or they may be overloading a Web server with requests, and the owner needs to stop the crawler. Identification is also useful for administrators that are interested in knowing when they may expect their Web pages to be indexed by a particular **search engine**.

## Crawling the deep web

[**edit**]

A vast amount of web pages lie in the **deep or invisible web**.<sup>[43]</sup> These pages are typically only accessible by submitting queries to a database, and regular crawlers are unable to find these pages if there are no links that point to them. Google's **Sitemaps** protocol and **mod oai**<sup>[44]</sup> are intended to allow discovery of these **deep-Web** resources.

Deep web crawling also multiplies the number of web links to be crawled. Some crawlers only take some of the URLs in `<a href="URL">` form. In some cases, such as the **Googlebot**, Web crawling is done on all text contained inside the hypertext content, tags, or text.

Strategic approaches may be taken to target deep Web content. With a technique called **screen scraping**, specialized software may be customized to automatically and repeatedly query a given Web form with the intention of aggregating the resulting data. Such software can be used to span multiple Web forms across multiple Websites. Data extracted from the results of one Web form submission can be taken and applied as input to another Web form thus establishing continuity across the Deep Web in a way not possible with traditional web crawlers.<sup>[45]</sup>

Pages built on **AJAX** are among those causing problems to web crawlers. **Google** has proposed a format of AJAX calls that their bot can recognize and index.<sup>[46]</sup>

## Visual vs programmatic crawlers

[**edit**]

There are a number of "visual web scraper/crawler" products available on the web which will crawl pages and structure data into columns and rows based on the users requirements. One of the main difference between a classic and a visual crawler is the level of programming ability required to set up a crawler. The latest generation of "visual scrapers" remove the majority of the programming skill needed to be able to program and start a crawl to scrape web data.

The visual scraping/crawling method relies on the user "teaching" a piece of crawler technology, which then follows patterns in semi-structured data sources. The dominant method for teaching a visual crawler is by highlighting data in a browser and training columns



and rows. While the technology is not new, for example it was the basis of Needlebase which has been bought by Google (as part of a larger acquisition of ITA Labs<sup>[47]</sup>), there is continued growth and investment in this area by investors and end-users.<sup>[citation needed]</sup>

## List of web crawlers

[edit]

Further information: [List of search engine software](#)

The following is a list of published crawler architectures for general-purpose crawlers (excluding focused web crawlers), with a brief description that includes the names given to the different components and outstanding features:

## Historical web crawlers

[edit]

- **WolfBot** was a massively multi threaded crawler built in 2001 by Mani Singh a Civil Engineering graduate from the University of California at Davis.
- **World Wide Web Worm** was a crawler used to build a simple index of document titles and URLs. The index could be searched by using the **grep Unix** command.
- Yahoo! Slurp was the name of the **Yahoo!** Search crawler until Yahoo! contracted with **Microsoft** to use **Bingbot** instead.

## In-house web crawlers

[edit]

- Applebot is **Apple's** web crawler. It supports **Siri** and other products.<sup>[48]</sup>
- **Bingbot** is the name of Microsoft's **Bing** webcrawler. It replaced **Msnbot**.
- Baiduspider is **Baidu's** web crawler.
- DuckDuckBot is **DuckDuckGo's** web crawler.
- **Googlebot** is described in some detail, but the reference is only about an early version of its architecture, which was written in C++ and **Python**. The crawler was integrated with the indexing process, because text parsing was done for full-text indexing and also for URL extraction. There is a URL server that sends lists of URLs to be fetched by several crawling processes. During parsing, the URLs found were passed to a URL server that checked if the URL have been previously seen. If not, the URL was added to the queue of the URL server.
- **WebCrawler** was used to build the first publicly available full-text index of a subset of the Web. It was based on **lib-WWW** to download pages, and another program to parse and order URLs for breadth-first exploration of the Web graph. It also included a real-time

crawler that followed links based on the similarity of the anchor text with the provided query.

- **WebFountain** is a distributed, modular crawler similar to Mercator but written in C++.
- **Xenon** is a web crawler used by government tax authorities to detect fraud. [\[49\]](#)[\[50\]](#)

## Commercial web crawlers

[\[edit\]](#)

The following web crawlers are available, for a price::

- **Diffbot** - programmatic general web crawler, available as an **API**
- **SortSite** - crawler for analyzing websites, available for **Windows** and **Mac OS**
- Swiftbot - **Swifttype**'s web crawler, available as **software as a service**
- Aleph Search - web crawler allowing massive collection with high scalability

## Open-source crawlers

[\[edit\]](#)

- **Apache Nutch** is a highly extensible and scalable web crawler written in Java and released under an **Apache License**. It is based on **Apache Hadoop** and can be used with **Apache Solr** or **Elasticsearch**.
- **Grub** was an open source distributed search crawler that **Wikia Search** used to crawl the web.
- **Heritrix** is the **Internet Archive**'s archival-quality crawler, designed for archiving periodic snapshots of a large portion of the Web. It was written in **Java**.
- **ht://Dig** includes a Web crawler in its indexing engine.
- **HTTrack** uses a Web crawler to create a mirror of a web site for off-line viewing. It is written in **C** and released under the GPL.
- Norconex Web Crawler is a highly extensible Web Crawler written in **Java** and released under an **Apache License**. It can be used with many repositories such as **Apache Solr**, **Elasticsearch**, **Microsoft Azure Cognitive Search**, **Amazon CloudSearch** and more.
- **mnoGoSearch** is a crawler, indexer and a search engine written in C and licensed under the GPL (\*NIX machines only)
- **Open Search Server** is a search engine and web crawler software release under the GPL.
- **Scrapy**, an open source webcrawler framework, written in python (licensed under **BSD**).
- **Seeks**, a free distributed search engine (licensed under **AGPL**).
- **StormCrawler**, a collection of resources for building low-latency, scalable web crawlers on **Apache Storm** (Apache License).
- **tkWWW Robot**, a crawler based on the **tkWWW** web browser (licensed under GPL).

- **GNU Wget** is a **command-line**-operated crawler written in **C** and released under the **GPL**. It is typically used to mirror Web and FTP sites.
- **YaCy**, a free distributed search engine, built on principles of peer-to-peer networks (licensed under GPL).

## See also

[[edit](#)]

- [Automatic indexing](#)
- [Gnutella crawler](#)
- [Web archiving](#)
- [Webgraph](#)
- [Website mirroring software](#)
- [Search Engine Scraping](#)
- [Web scraping](#)

## References

[[edit](#)]

1. <sup>^</sup> *"Web Crawlers: Browsing the Web"*. Archived from *the original* on 6 December 2021.
2. <sup>^</sup> Spetka, Scott. *"The TkWWW Robot: Beyond Browsing"*. NCSA. Archived from *the original* on 3 September 2004. Retrieved 21 November 2010.
3. <sup>^</sup> Kobayashi, M. & Takeda, K. (2000). "Information retrieval on the web". *ACM Computing Surveys*. **32** (2): 144–173. *CiteSeerX* 10.1.1.126.6094. doi: 10.1145/358923.358934. S2CID 3710903.
4. <sup>^</sup> See definition of scutter on FOAF Project's wiki Archived 13 December 2009 at the Wayback Machine
5. <sup>^</sup> Masanès, Julien (15 February 2007). *Web Archiving*. Springer. p. 1. ISBN 978-3-54046332-0. Retrieved 24 April 2014.
6. <sup>^</sup> Edwards, J.; McCurley, K. S.; and Tomlin, J. A. (2001). "An adaptive model for optimizing performance of an incremental web crawler". *Proceedings of the 10th international conference on World Wide Web*. pp. 106–113. *CiteSeerX* 10.1.1.1018.1506 . doi:10.1145/371920.371960. ISBN 978-1581133486. S2CID 10316730. Archived from *the original* on 25 June 2014. Retrieved 25 January 2007.cite book: CS1 maint: multiple names: authors list ([link](#))
7. <sup>^</sup> Castillo, Carlos (2004). *Effective Web Crawling* (PhD thesis). University of Chile. Retrieved 3 August 2010.
8. <sup>^</sup> Gulls, A.; A. Signori (2005). "The indexable web is more than 11.5 billion pages". *Special interest tracks and posters of the 14th international conference on World Wide Web*. ACM Press. pp. 902–903. doi:10.1145/1062745.1062789.

9. ^ Lawrence, Steve; C. Lee Giles (8 July 1999). "Accessibility of information on the web". *Nature*. **400** (6740): 107–9. *Bibcode*:1999Natur.400..107L. doi:10.1038/21987. PMID 10428673. S2CID 4347646.
10. ^ Cho, J.; Garcia-Molina, H.; Page, L. (April 1998). "Efficient Crawling Through URL Ordering". Seventh International World-Wide Web Conference. Brisbane, Australia. doi:10.1142/3725. ISBN 978-981-02-3400-3. Retrieved 23 March 2009.
11. ^ Cho, Junghoo, "Crawling the Web: Discovery and Maintenance of a Large-Scale Web Data", PhD dissertation, Department of Computer Science, Stanford University, November 2001.
12. ^ Najork, Marc and Janet L. Wiener. "Breadth-first crawling yields high-quality pages". Archived 24 December 2017 at the Wayback Machine In: *Proceedings of the Tenth Conference on World Wide Web*, pages 114–118, Hong Kong, May 2001. Elsevier Science.
13. ^ Abiteboul, Serge; Mihai Preda; Gregory Cobena (2003). "Adaptive on-line page importance computation". *Proceedings of the 12th international conference on World Wide Web*. Budapest, Hungary: ACM. pp. 280–290. doi:10.1145/775152.775192. ISBN 1-58113-680-3. Retrieved 22 March 2009.
14. ^ Boldi, Paolo; Bruno Codenotti; Massimo Santini; Sebastiano Vigna (2004). "UbiCrawler: a scalable fully distributed Web crawler" (PDF). *Software: Practice and Experience*. **34** (8): 711–726. CiteSeerX 10.1.1.2.5538. doi:10.1002/spe.587. S2CID 325714. Archived from the original (PDF) on 20 March 2009. Retrieved 23 March 2009.
15. ^ Boldi, Paolo; Massimo Santini; Sebastiano Vigna (2004). "Do Your Worst to Make the Best: Paradoxical Effects in PageRank Incremental Computations" (PDF). *Algorithms and Models for the Web-Graph. Lecture Notes in Computer Science*. Vol. 3243. pp. 168–180. doi:10.1007/978-3-540-30216-2\_14. ISBN 978-3-540-23427-2. Archived from the original (PDF) on 1 October 2005. Retrieved 23 March 2009.
16. ^ Baeza-Yates, R.; Castillo, C.; Marin, M. and Rodriguez, A. (2005). "Crawling a Country: Better Strategies than Breadth-First for Web Page Ordering." In: *Proceedings of the Industrial and Practical Experience track of the 14th conference on World Wide Web*, pages 864–872, Chiba, Japan. ACM Press.
17. ^ Shervin Daneshpajouh, Mojtaba Mohammadi Nasiri, Mohammad Ghodsi, A Fast Community Based Algorithm for Generating Crawler Seeds Set. In: *Proceedings of 4th International Conference on Web Information Systems and Technologies (Webist-2008)*, Funchal, Portugal, May 2008.
18. ^ Pant, Gautam; Srinivasan, Padmini; Menczer, Filippo (2004). "Crawling the Web" (PDF). In Levene, Mark; Poulovassilis, Alexandra (eds.). *Web Dynamics: Adapting to Change in Content, Size, Topology and Use*. Springer. pp. 153–178. ISBN 978-3-540-40676-1. Archived from the original (PDF) on 20 March 2009. Retrieved 9 May 2006.
19. ^ Cothey, Viv (2004). "Web-crawling reliability" (PDF). *Journal of the American Society for Information Science and Technology*. **55** (14): 1228–1238. CiteSeerX 10.1.1.117.185. doi:10.1002/asi.20078.
20. ^ Menczer, F. (1997). ARACHNID: Adaptive Retrieval Agents Choosing Heuristic Neighborhoods for Information Discovery Archived 21 December 2012 at the Wayback

- Machine. In D. Fisher, ed., Machine Learning: Proceedings of the 14th International Conference (ICML97). Morgan Kaufmann
21. ^ Menczer, F. and Belew, R.K. (1998). *Adaptive Information Agents in Distributed Textual Environments Archived* 21 December 2012 at the *Wayback Machine*. In K. Sycara and M. Wooldridge (eds.) Proc. 2nd Intl. Conf. on Autonomous Agents (Agents '98). ACM Press
  22. ^ Chakrabarti, Soumen; Van Den Berg, Martin; Dom, Byron (1999). *"Focused crawling: A new approach to topic-specific Web resource discovery"* (PDF). *Computer Networks*. **31** (11–16): 1623–1640. doi:10.1016/s1389-1286(99)00052-3. Archived from *the original* (PDF) on 17 March 2004.
  23. ^ Pinkerton, B. (1994). *Finding what people want: Experiences with the WebCrawler*. In Proceedings of the First World Wide Web Conference, Geneva, Switzerland.
  24. ^ Diligenti, M., Coetzee, F., Lawrence, S., Giles, C. L., and Gori, M. (2000). *Focused crawling using context graphs*. In Proceedings of 26th International Conference on Very Large Databases (VLDB), pages 527-534, Cairo, Egypt.
  25. ^ Wu, Jian; Teregowda, Pradeep; Khabsa, Madian; Carman, Stephen; Jordan, Douglas; San Pedro Wandelper, Jose; Lu, Xin; Mitra, Prasenjit; Giles, C. Lee (2012). "Web crawler middleware for search engine digital libraries". *Proceedings of the twelfth international workshop on Web information and data management - WIDM '12*. p. 57. doi:10.1145/2389936.2389949. ISBN 9781450317207. S2CID 18513666.
  26. ^ Wu, Jian; Teregowda, Pradeep; Ramírez, Juan Pablo Fernández; Mitra, Prasenjit; Zheng, Shuyi; Giles, C. Lee (2012). "The evolution of a crawling strategy for an academic document search engine". *Proceedings of the 3rd Annual ACM Web Science Conference on - Web Sci '12*. pp. 340–343. doi:10.1145/2380718.2380762. ISBN 9781450312288. S2CID 16718130.
  27. ^ Dong, Hai; Hussain, Farookh Khadeer; Chang, Elizabeth (2009). *"State of the Art in Semantic Focused Crawlers"*. *Computational Science and Its Applications – ICCSA 2009. Lecture Notes in Computer Science*. Vol. 5593. pp. 910–924. doi:10.1007/978-3-642-02457-3\_74. hdl:20.500.11937/48288. ISBN 978-3-642-02456-6.
  28. ^ Dong, Hai; Hussain, Farookh Khadeer (2013). *"SOF: A semi-supervised ontology-learning-based focused crawler"*. *Concurrency and Computation: Practice and Experience*. **25** (12): 1755–1770. doi:10.1002/cpe.2980. S2CID 205690364.
  29. ^ Junghoo Cho; Hector Garcia-Molina (2000). *"Synchronizing a database to improve freshness"* (PDF). *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*. Dallas, Texas, United States: ACM. pp. 117–128. doi:10.1145/342009.335391. ISBN 1-58113-217-4. Retrieved 23 March 2009.
  30. ^ a b E. G. Coffman Jr; Zhen Liu; Richard R. Weber (1998). "Optimal robot scheduling for Web search engines". *Journal of Scheduling*. **1** (1): 15–29. CiteSeerX 10.1.1.36.6087. doi:10.1002/(SICI)1099-1425(199806)1:1<15::AID-JOS3>3.0.CO;2-K.
  31. ^ a b Cho, Junghoo; Garcia-Molina, Hector (2003). "Effective page refresh policies for Web crawlers". *ACM Transactions on Database Systems*. **28** (4): 390–426. doi:10.1145/958942.958945. S2CID 147958.



32. ^ **a b** Junghoo Cho; Hector Garcia-Molina (2003). "Estimating frequency of change". *ACM Transactions on Internet Technology*. **3** (3): 256–290. *CiteSeerX* 10.1.1.59.5877. doi:10.1145/857166.857170. S2CID 9362566.
33. ^ Ipeirotis, P., Ntoulas, A., Cho, J., Gravano, L. (2005) **Modeling and managing content changes in text databases Archived** 5 September 2005 at the **Wayback Machine**. In Proceedings of the 21st IEEE International Conference on Data Engineering, pages 606-617, April 2005, Tokyo.
34. ^ Koster, M. (1995). Robots in the web: threat or treat? *ConneXions*, 9(4).
35. ^ Koster, M. (1996). **A standard for robot exclusion Archived** 7 November 2007 at the **Wayback Machine**.
36. ^ Koster, M. (1993). **Guidelines for robots writers Archived** 22 April 2005 at the **Wayback Machine**.
37. ^ Baeza-Yates, R. and Castillo, C. (2002). **Balancing volume, quality and freshness in Web crawling**. In *Soft Computing Systems – Design, Management and Applications*, pages 565–572, Santiago, Chile. IOS Press Amsterdam.
38. ^ Heydon, Allan; Najork, Marc (26 June 1999). "**Mercator: A Scalable, Extensible Web Crawler**" (PDF). Archived from *the original* (PDF) on 19 February 2006. Retrieved 22 March 2009. *cite journal*: Cite journal requires |journal= (help)
39. ^ Dill, S.; Kumar, R.; Mccurley, K. S.; Rajagopalan, S.; Sivakumar, D.; Tomkins, A. (2002). "**Self-similarity in the web**" (PDF). *ACM Transactions on Internet Technology*. **2** (3): 205–223. doi:10.1145/572326.572328. S2CID 6416041.
40. ^ M. Thelwall; D. Stuart (2006). "**Web crawling ethics revisited: Cost, privacy and denial of service**". *Journal of the American Society for Information Science and Technology*. **57** (13): 1771–1779. doi:10.1002/asi.20388.
41. ^ Brin, Sergey; Page, Lawrence (1998). "**The anatomy of a large-scale hypertextual Web search engine**". *Computer Networks and ISDN Systems*. **30** (1–7): 107–117. doi:10.1016/s0169-7552(98)00110-x. S2CID 7587743.
42. ^ Shkapenyuk, V. and Suel, T. (2002). **Design and implementation of a high performance distributed web crawler**. In Proceedings of the 18th International Conference on Data Engineering (ICDE), pages 357-368, San Jose, California. IEEE CS Press.
43. ^ Shestakov, Denis (2008). **Search Interfaces on the Web: Querying and Characterizing Archived** 6 July 2014 at the **Wayback Machine**. TUCS Doctoral Dissertations 104, University of Turku
44. ^ Michael L Nelson; Herbert Van de Sompel; Xiaoming Liu; Terry L Harrison; Nathan McFarland (24 March 2005). "mod\_oai: An Apache Module for Metadata Harvesting": cs/0503069. *arXiv:cs/0503069*. Bibcode:2005cs.....3069N. *cite journal*: Cite journal requires |journal= (help)
45. ^ Shestakov, Denis; Bhowmick, Sourav S.; Lim, Ee-Peng (2005). "**DEQUE: Querying the Deep Web**" (PDF). *Data & Knowledge Engineering*. **52** (3): 273–311. doi:10.1016/s0169-023x(04)00107-7.
46. ^ "**AJAX crawling: Guide for webmasters and developers**". Retrieved 17 March 2013.



47. ^ ITA Labs "ITA Labs Acquisition" Archived 18 March 2014 at the Wayback Machine 20 April 2011 1:28 AM
48. ^ "About Applebot". Apple Inc. Retrieved 18 October 2021.
49. ^ Norton, Quinn (25 January 2007). "Tax takers send in the spiders". Business. Wired. Archived from the original on 22 December 2016. Retrieved 13 October 2017.
50. ^ "Xenon web crawling initiative: privacy impact assessment (PIA) summary". Ottawa: Government of Canada. 11 April 2017. Archived from the original on 25 September 2017. Retrieved 13 October 2017.

## Further reading

[edit]

- Cho, Junghoo, "Web Crawling Project", UCLA Computer Science Department.
- A History of Search Engines, from Wiley
- WIVET is a benchmarking project by OWASP, which aims to measure if a web crawler can identify all the hyperlinks in a target website.
- Shestakov, Denis, "Current Challenges in Web Crawling" and "Intelligent Web Crawling", slides for tutorials given at ICWE'13 and WI-IAT'13.
- v
- t
- e

## Internet search

### Types

- Web search engine (List)
- Metasearch engine
- Multimedia search
- Collaborative search engine
- Cross-language search
- Local search
- Vertical search
- Social search
- Image search
- Audio search
- Video search engine
- Enterprise search
- Semantic search
- Natural language search engine
- Voice search

## Tools

- Cross-language information retrieval
- Search by sound
- Search engine marketing
- Search engine optimization
- Evaluation measures
- Search oriented architecture
- Selection-based search
- Document retrieval
- Text mining
- Web crawler
- Multisearch
- Federated search
- Search aggregator
- Index/Web indexing
- Focused crawler
- Spider trap
- Robots exclusion standard
- Distributed web crawling
- Web archiving
- Website mirroring software
- Web query
- Web query classification

## Protocols and standards

- Z39.50
- Search/Retrieve Web Service
- Search/Retrieve via URL
- OpenSearch
- Representational State Transfer
- Wide area information server

## See also

- Search engine
- Desktop search
- Online search

- **v**
- **t**
- **e**

Web crawlers

Internet bots designed for Web crawling and Web indexing

### Active

- 80legs
- bingbot
- Crawljax
- Fetcher
- Googlebot
- Heritrix
- HTTrack
- PowerMapper
- Wget

### Discontinued

- FAST Crawler
- msnbot
- RBSE
- TkWWW robot
- Twiceler

### Types

- Distributed web crawler
- Focused crawler

Authority control databases: National  [Edit this at Wikidata](#)

Image not found or type unknown

Check our other pages :

- [SEO keywords](#)
- [SEO agency Sydney](#)
- [digital agency Sydney](#)
- [SEO check](#)
- [Sydney SEO consultant](#)
- [SEO agency Sydney](#)
- [search engine optimisation consultants](#)

## Frequently Asked Questions

## **What is a content agency in Sydney?**

A content agency in Sydney focuses on creating high-quality, SEO-optimized content that resonates with your target audience. Their services typically include blog writing, website copy, video production, and other forms of media designed to attract traffic and improve search rankings.

## **Why should I consider SEO packages in Australia?**

SEO packages in Australia typically bundle essential optimization services such as keyword research, technical audits, content creation, and link building at a set price. They are designed to simplify the process, provide consistent results, and help businesses of all sizes improve their online visibility.

## **What is involved in SEO consulting?**

SEO consulting involves analyzing a website's current performance, identifying areas for improvement, and recommending strategies to boost search rankings. Consultants provide insights on keyword selection, on-page and technical optimization, content development, and link-building tactics.

## **What are the benefits of working with an SEO consultant in Sydney?**

An SEO consultant in Sydney can provide tailored advice and strategies that align with your business's goals and local market conditions. They bring expertise in keyword selection, content optimization, technical SEO, and performance monitoring, helping you achieve better search rankings and more organic traffic.

## **What role do SEO consultants play in a digital marketing strategy?**

SEO consultants are responsible for improving your website's visibility and performance in search engines. By analyzing data, refining keyword strategies, and optimizing site elements, they enhance your overall digital marketing efforts, leading to more traffic, better user engagement, and higher conversions.

### SEO in marketing

SEO Sydney

Phone : 1300 684 339

City : Sydney

State : NSW

Zip : 2000

[Google Business Profile](#)

[Google Business Website](#)

Company Website : <https://sydney.website/seo-sydney/>

## USEFUL LINKS

[SEO Website](#)

[SEO Services Sydney](#)

[Local SEO Sydney](#)

[SEO Ranking](#)

[SEO optimisation](#)

## LATEST BLOGPOSTS

[SEO community](#)

[SEO Buzz](#)

[WordPress SEO](#)

[SEO Audit](#)

[Sitemap](#)

[Privacy Policy](#)

[About Us](#)

[SEO Castle Hill](#) | [SEO Fairfield](#) | [SEO Hornsby](#) | [SEO Liverpool](#) | [SEO North Sydney](#) | [SEO Norwest](#) | [SEO Parramatta](#) | [SEO Penrith](#) | [SEO Strathfield](#) | [SEO Wetherill Park](#)

Follow us