Encyclopedia of Information Systems and Technology

Phillip A. Laplante, Vladimir Stantchev, Jeff Voas, Anne-Marie Smith, Qiuming Zhu, Morris Chang, Irena Bojanova, David Alan Grier, Jia Zhang, Thomas Costello, San Murugesan, Jim Goldman, Joanna DeFranco

Search Engines

Publication details
Randolph Hock
Published online on: 29 Dec 2015

How to cite:- Randolph Hock. 29 Dec 2015, Search Engines from: Encyclopedia of Information Systems and Technology CRC Press
Accessed on: 19 Aug 2023

PLEASE SCROLL DOWN FOR DOCUMENT

Full terms and conditions of use: https://www.routledgehandbooks.com/legal-notices/terms
This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
Search Engines

Randolph Hock
Online Strategies, Vienna, Virginia, U.S.A.

Abstract
This entry provides an overview of Web search engines, looking at the definition, components, leading engines, searching capabilities, and types of engines. It examines the components that make up a search engine and briefly discusses the process involved in identifying content for the engines’ databases and the indexing of that content. Typical search options are reviewed and the major Web search engines are identified and described. Also identified and described are various specialty search engines, such as those for special content such as video and images, and engines that take significantly different approaches to the search problem, such as visualization engines and metasearch engines.

INTRODUCTION
Web search engines, for the public at large, have come to be perhaps the most frequently used computer services for locating information. To some degree the same is true for many researchers, information professionals, and others. To most effectively and efficiently utilize these services, some understanding of the structure, make-up, content, features, and variety and breadth of these services is essential. This entry addresses those various aspects including just what is meant by “search engines,” the components of a search engine, and typical search features, and it provides a profile of the major general Web search engines and a look at specialty search engines, visualization engines, and metasearch engines.

WHAT IS MEANT BY “SEARCH ENGINES?”
The term “search engines” can have a variety of meanings, in the broadest sense referring to any computer program that facilitates the searching of a database. In the context of library and information science, however, the term has come to primarily refer to “Web search engines,” that is, those services on the Web that allow searching of a large database of Web pages and other Web content by word, phrase, and other criteria. (For this discussion, hereafter, “search engines” will be taken to refer to “Web search engines.”) A certain level of ambiguity becomes apparent, however, when it is realized that what is often referred to as a “search engine” is often a reference to the overall service that is provided, beyond just a search of Web sites. (“Google” is thought of not as just the searching part of the Google enterprise, but the many added features and content as well.) It is often impossible and unproductive to discuss the narrower “searching” part without discussing the broader range of services. That ambiguity in terminology is a result and artifact of the history of search engines but recognition of the ambiguity is necessary for an understanding of the current nature of such services.

Search engines vary in a number of ways and most could be considered to fall into one of the following four categories: General Web Search Engines (which have the purpose of searching a large portion of all pages that exist on the Web), Specialty Search Engines (which focus on searching a specific kind of document, file type, or source from a particular subject or geographic region), Visualization Search Engines (which furnish diagrams, images, or other “visuals” to show relationships among the items in a particular set of retrieved items), and Metasearch Engines (which gather together the search results on a specific topic from multiple search engines).

COMPONENTS OF A SEARCH ENGINE
General Web search engines and specialty search engines can be considered to have four major components that correspond to the steps required to create the service: 1) the identification and gathering of the material (Web pages, etc.) to be included in the engine’s database; 2) an indexing program and the corresponding generated indexes; 3) the searching and ranking algorithms; and 4) the user interface.

Identifying Material to Be Included
Search engines identify those Web pages (and other items) to be included in the service’s database by two means: “crawling” and “submissions” of pages. The first, “crawling” consists of having programs (“crawlers”
or “spiders”) that on an ongoing basis scan the Internet to identify new sites or sites that have changed, gather information from those sites, and feed that information to the search engine’s indexing mechanism. The crawlers start by examining pages that the service already knows about and looking there for “new” links (links that the service does not already know about). When such links are identified, the pages to which the links lead are likewise examined for “new” links, and so on. More popular Web sites (such as those that have lots of links to them) may be crawled more thoroughly and more frequently than less popular sites.

The second way search engines identify new items to be added to the database is by having Web site owners (or others) “submit” sites or pages. Most engines provide a form by which this can be done. Search services maintain their own policies as to whether submitted (or for that matter, pages identified by crawling) will indeed be added to the database, particularly looking to exclude unacceptable content (spam, sexually explicit material, etc.)

**Search Engine’s Index and Indexing Program**

After a new or changed page is identified by the search engine’s crawler, the page will typically be indexed under virtually every word on the page (up to some usually undisclosed limit). In addition to text words, other parts or characteristics of the page may also be indexed, including the URL (Uniform Resource Locator, the “Web address”), parts of the URL, links, metadata found in the “head” of the document, the URLs of links on the page, image filenames, words in linked text, etc. By identifying and indexing these pieces of data (pieces or characteristics of the Web page or other type of indexed document, such as an Excel file), they become searchable “fields,” thereby allowing users to use those fields to increase the quality of their search. The search system may also “derive” additional fields, such as language, by analysis of the document.

**The Search Engine’s Retrieval and Ranking Algorithms**

By narrow definition, the actual search “engine” is the search service’s retrieval program, that is, the program that identifies (retrieves) those pages in the database that match the criteria indicated by a user’s query. That identification function is necessarily supplemented by another important and more challenging program that is used to determine the order in which the retrieved records should be displayed, based on measures that try to identify which retrieved records (pages, etc.) are likely to have the highest relevance in respect to the user’s query.

This “relevance-ranking” algorithm usually takes many factors into account.

Exactly what factors go into the relevance ranking process varies, but they include: use of keywords in titles, text, headings, etc.; popularity of the sites (how many and which sites link to the site); words used in anchors (clickable text); internal links (how many and what kind of links within the larger site point to the page); quality of links leading out to other pages (whether they point to high quality pages) and so on.[1]

The success or the failure of the relevance ranking algorithm is critical to the user’s perception of the search engine, the user’s continued use of that system, and the commercial success of the engine.

**The Interface Presented to the User for Gathering Queries**

This interface that the user typically sees includes the home page of the search service and other pages (such as an advanced search page) that present search options to the users and accept the users’ search queries, as well as the search results page. The search service can choose to have their page focus almost exclusively on “search” (as with Google) or be a more general, wide-reaching “portal” page, providing much more than just searching capabilities. (The “portal” dilemma for search services will be discussed in more detail later.)

Regardless of what other services and information are provided on the service’s homepage, the “searching” part usually consists of a single search box plus links to an advanced search page and to other searchable databases that are made available by the service (images, video, news, etc.). Usually there are also links to “help” screens, etc. While the simplicity of a single search box appeals to the less experienced user, it also usually provides substantial, but not obvious, capabilities for extensive searching sophistication, such as the potential for using Boolean logic and “prefixes” (e.g., “title:”) to perform field searching and other functions. The advanced search page much more explicitly lays out the possibilities to the user, providing a menu-based approach to utilization of features.

**The Portal Dilemma**

From the early days of search engines, search engine providers have wrestled with the decision as to whether to make their home page one that focuses almost exclusively on “search” or one that provides a variety of added services such as news, weather, etc., the latter approach often referred to as a “portal.” From its beginning, before it was even a “search engine” and was just a directory, Yahoo! preferred the portal approach. AltaVista, a leading search engine in the 1990s, went back
and forth between the two extremes, a situation which may have contributed to its demise. Google was, from the beginning, almost purely a “search engine” and the simplicity of its interface was undoubtedly one factor in its rapid rise in popularity. Search services tend to “cover their bets” however, by providing alternatives. Yahoo! provides a Google-like option at search.yahoo.com and Google provides a personalizable Yahoo-style page with its iGoogle portal page.

**Searching Options Typically Provided**

All leading search engines provide a range of user-accessible options that permit the user to modify their search queries in ways that can improve both the precision and the recall of their search results. Which specific options are provided varies from engine to engine, but there are several that are fairly typical (and some that are unique to a particular engine). The most typical options include Boolean operations, phrase searching, language specification, and specifying that only those pages are retrieved for which the search term appears in a particular part (field) of the record such as the title, URL, or links. Since engines now cover other document types beyond just pages written in Hyper-Text Markup Language (HTML), with several engines, users can also narrow their search to a specific file format (Web pages, Adobe Acrobat files, Excel files, etc.). Most engines also provide an option to filter “adult content” material.

**Boolean logic**

In the context of Web searching, “Boolean logic” refers to the process of identifying those items found in the database that contain a particular combination of search terms. It is used to indicate that a particular group of terms must all be present (the Boolean “AND”), that any of a particular group of terms is acceptable (the Boolean “OR”), or that if a particular term is present, the item is rejected (the Boolean “NOT”). (See the entry, Boolean Algebras [Elis Classic], p. 660.)

Engines usually provide two different ways to qualify a query with Boolean operations: 1) the option of applying a syntax directly to what is entered in the search box and 2) menu options on an advanced search page. Using the menus can be thought of as “simplified Boolean” and, depending upon the structure of the advanced search page, may or may not provide the precision achievable by the use of syntax in the main search box. (For example, the ability to apply “OR’s” to more than one of the concepts included in the query may be done in the main search box but may not be allowed for on the advanced search page.)

The exact syntax used varies with the search engine. All major engines currently automatically apply an “AND” between your terms, so when the following is entered:

```
prague economics tourism
```

what will be retrieved is what more traditionally would have been expressed as: prague AND economics AND tourism.

Very precise search requirements can be expressed using combinations of the operators along with parentheses to indicate the order of operations. For example:

```
(grain OR corn OR wheat) (production OR harvest)
```

Oklahoma 1997

At various times, search engines have allowed the use of symbols (+, &, ~, etc.) instead of words (AND, OR, NOT) and indeed, for the “NOT” most search engines currently suggest the use of a minus sign in front of the term. Some search engines require the use of parentheses around “nested” (OR’ed) terms, some do not.

For details on Boolean syntax for any search engine, the help pages for that engine should be consulted. There are also Web sites, such as Search Engine Showdown from Greg Notess (http://www.searchengineshowdown.com) that summarize the syntax (and other features) for all major engines.

The alternative to using syntax to apply Boolean is the use of menus on an advanced search page. There, for example, you may find a pull-down menu, where, if you choose the “all the words” option, you are requesting the Boolean AND. If you choose the “any of the words” option from such a menu, you are specifying an OR. There is usually also a box for excluding terms (NOT).

**Phrase searching**

Phrase searching is an option that is available in virtually every search engine, and almost always uses the same syntax, the use of quotation marks around the phrase. For example, searching on “Red River” (with the quotation marks) will assure that you get only those pages that contain the word “red” immediately in front of the word “river.” Of all search engine techniques, this is widely regarded as one of the most useful and easiest for achieving higher precision in a Web search. It is also useful for such things as identifying quotations and identifying plagiarism.

**Title searching**

Title searching, that is, limiting your retrieval to only those items (pages) that have a particular term or combination of terms in their title, is one example of “field searching,” as referred to earlier. It is also another example of a technique that can yield very high precision in a search. Most search engines use the “intitle:” prefix and/or the “allintitle:” prefix for the syntax for title searching (“allintitle:” allows specifying that more
than one term be included in the title, not necessarily in any particular order).

**URL, site, and domain searching**

Search engines typically index Web pages (and other document types) by both the overall URL and by the segments of the URL. This facilitates the finding of any document that comes from a particular domain or part of a domain (also a specific site or part of a site). Doing a search in which results are limited to a specific site allows one, in effect, to perform a search of that site. Even for sites that have a “site search” box on their home page, more complete results can often be found by using this technique than by using the site’s own search feature. “inurl:,” “allinurl:,” and “site:” are the prefixes commonly used.

The term, “Domain searching” is sometimes used to refer to the above process and the use of the term, “Domain,” points out that this approach can be used to limit retrieval to sites having a particular top-level domain, such as: gov, edu, uk, ca, or fr. This could be used, for example, to identify only Canadian sites that mention tariffs, or to only get educational sites that mention biodiversity.

**Link searching**

There are two varieties of “link” searching. In the more common variety, one can search for all pages that have a hypertext link to a particular URL, and in the other variety, one can search for words contained in the linked text on the page. In the former, you can check, for example, which Web pages have linked to your organization’s URL. In the second variety, you can see which Web pages have the name of your organization as linked text. Either variety can be very informative in terms of who is interested in either your organization or your Web site. Also, if you are looking for information on an organization, it can sometimes be useful to know who is linking to that organization’s site.

This searching option is available in some search engines on their advanced page and/or on the main page with the use of prefixes. (usually “link:”). Engines may allow you to find links to an overall site, or to a specific page within a site.

**Language searching**

Although all of the major engines allow limiting retrieval to pages written in a given language, they differ in terms of which languages can be specified. The 40 or so most common languages are specifiable in most of the major engines. Though some engines provide a pre-fix option for searching for languages, more typically one would go to the engine’s advanced search page to narrow to a language.

**Date searching**

Searching by the date of Web pages is an obviously desirable option, and most major engines provide such an option. Unfortunately, because of lack of clear or reliable information on a page regarding when the page itself was initially created, the date on which the content of the page was created, or even when the content on the page was significantly modified, it is often impossible for a search engine to assign a truly “reliable” date to a Web page. As a “workaround,” engines may take the date when the page was last modified or may assign a date based on when the page was last crawled by the engine. For searching Web pages, users should be aware of this approximation and its effect on precision when using the date searching option that is offered by most search engines (usually on their advanced search page). (On the other hand, for some of the other databases an engine may provide, such as news, the date searching may be very precise.)

**Searching by file type**

For most of the 1990s, most search engines only indexed and allowed searching of regular HTML pages. In the crawling process (or for submitted pages) when the engine’s indexing program encountered a link that led to another type of document, such as an Adobe Acrobat (pdf), or Excel (xls) file, the link was ignored. Starting with Adobe Acrobat files, other file types were fairly rapidly added to the corpus of “indexable” pages. This not only increased the breadth of resources available to the searcher, but also provided the capability for the searcher to limit retrieval by type of file. Limiting to Adobe Acrobat files provides documents more suited to printing. Narrowing to PowerPoint files can provide convenient summaries of a topic. Limiting to Excel files can often enable a greater focus on statistics.

**Search Results Pages**

As well as providing enhanced searching capabilities, search engines also enhance the content of results pages, beyond presenting just a listing of the Web page results that match the user’s query. At the same time they search their Web database, they may automatically search the other databases they have, such as news, images, and video, and on search pages may automatically provide links to the matching items from those additional databases. Some search engines may search additional “reference” resources, such as dictionaries, encyclopedias, maps, etc., and likewise display matching content from those sources.
As well as displaying such supplemental content on
results pages, search engines may also provide suggestions
for ways in which the user might further qualify
search criteria. This is done by suggesting related, nar-
rower, or broader topics. Some engines also provide
links to narrow the search by file type, language, or
type of site (weblog, forum, commercial, or noncom-
mmercial, etc.)

Specific options may also be offered on results pages
for each retrieved item. Some engines keep a copy of
each page they have indexed and provide a link to that
“cached” page. This is particularly useful if, in the time
since the page was indexed, the page was removed, is
not available because of a server problem, or has
changed in a way such that the term the user searched
for is no longer on the page.

With records for pages that are not in the language of
the search engine interface, there may be an option to
translate the record (for example, if the user is using an
English language version of Google and a page is in French
or if the user is using the French version and the page is in
English). Click on the “translate” link to receive a machine
translation of the page. As with other machine translations,
what you get may not be a “good” translation, but it may be
an “adequate” translation, adequate in that it will give you
a good idea of what the page is talking about. Also keep in
mind that only “words” are translated. The translation pro-
cram cannot translate words you see on a page that are
actually “images” rather than “text.”

One feature offered on search results pages by all of
the major engines is a spell-checker. If you misspelled a
word, or the search engine thinks you might have, it
generously asks something like “Did you mean?” and
gives you a likely alternative. If it was indeed a mistake,
just click on the suggested alternative to correct the
problem.

Search results pages will usually display links labeled
as “Sponsor Results,” “Sponsored Links,” etc.—These
are “ads” for Web sites and are there because the Web
site has paid to appear on the search engine’s results
pages. Major engines keep these sponsor sites clearly
identifiable by, for example, putting them in a blue
background, or to the side of the page. Searchers should
remain aware that it is the presence of these ads that
makes the existence of search engines possible.

The Search Engine Leaders—Post-2000

Popularity of various search engines can change fairly
quickly. In the early and mid-1990s, a list of the most
popular engines included, among others, AltaVista, Hot-
bot, Excite, InfoSeek, and Lycos (Yahoo! was still
primarily a directory, and though it had a search engine
function, for that function it made use of, at various
times, AltaVista’s and Google’s databases.)

By the latter part of the 2000s the following were the
leaders: Google, Yahoo!, MSN/Windows/Live Search,
AOL, and Ask. (in that order). Those five search engines
represented 94 percent of all U.S. searches.[3] (Brief pro-
files of the engines just mentioned are given below.)

Google

Google, which emerged as a company in 1998, grew
very rapidly, its growth attributed largely to the simplic-
ity of its interface, the lack of advertisements on the
home page, and the quality of its relevance ranking (that
fact significantly affected by Google’s patented Page
Rank program).[3] Google rather quickly went beyond
“search” and began providing additional features and
content, some of the enhancements emerging from
within the Google organization and some (such as its
e-mail service, Gmail) being patterned after such serv-
ices already offered by its competitors. By the late
2000s, Google claimed more of the search market than
all of its competitors combined and was offering a broad
range of search services and a number of services not
directly related to search.

For its Web search offerings, Google provides all of
the typical search options (Boolean, field searching,
etc.) plus some unique searching features, the latter
including numeric range searching (e.g., China history
1850 to1890), and synonym searching (e.g., ~cars). As
well as the searching of Web pages, Google also offers
searches of databases of images, maps, news, products,
video, groups, books (Google Book Search), journal
articles (Google Scholar), and blogs. Some of these
search offerings are very similar to corresponding serv-
ces offered by Google’s competitors, but some, such as
“Google Book Search,” were original and regarded by
many as “ground-breaking” and even in some cases,
controversial. (Google Books Search is a major book
digitalization project, in cooperation with major publish-
ers and libraries.) The search features provided with
each of these databases is typically tailored to the spe-
cific nature of that kind of content.

Many of Google’s Web search features are features
that were already found on other search engines, but for
which Google provided significant enhancements. One
example is Google Language Tools. Many search
engines have provided a translation option that allows
retrieved items from a number of non-English languages
to be translated, using programs such as SYSTRAN’s
Babel Fish. In 2007, Google enhanced its own transla-
tion feature by allowing the user not just to translate a
specific result, but to input a search in the user’s own
language, then have Google automatically translate the
search terms, perform the search, and then deliver
results in both languages. Translations are done using
Google’s own statistical translation technology.
As it grew, Google rather rapidly redefined itself to be much more than a “search engine,” adding services that went beyond “search” and even beyond usual Web site content. Some services had a direct relationship to “search,” such as Google News Alerts, Google’s financial portal (“Google Finance”), the Google Toolbar for Web browsers, a desktop search tool for searching the content of one’s own computer, and Google’s own Web browser (“Chrome”). Some of the services Google began to offer included types of things that already existed as “portal” features in other search services. These offerings included a customizable portal page (iGoogle) with Google’s own calendar and notebook and links to a variety of other content such as newsfeeds. Among other services are Gmail (a Web-based e-mail service), Google Earth (imagery and related geospatial content for the entire Earth, as well as the Moon and the sky) and Google Talk (an instant messaging service). One of the manifestations of “Web 2.0” is the availability of user-accessible software that is resident on the Web, rather than on the user’s own computer. (The term, “Web 2.0,” refers not to an actual “version” of the Web, but to the fact that the nature of the Web, by the middle of the first decade of the 20th century, had changed from being primarily a place to go to find information to being a place that was much more personal, interactive, and collaborative, with the Web as a “platform” where programs are provided, used, and shared.) Google has moved very much in the Web 2.0 direction, providing Picasa (a photo-sharing and editing service), SketchUp (a computer-aided design, CAD, program), Google Docs (a collaborative spreadsheet, word-processor, and presentations program), and Sites (for creating Web sites). Google also offers “mobile” services (including mobile search, maps, text messaging, Gmail, etc.), an enterprise version of Google’s search engine, and a custom search engine that allows a user to have a search box (on their own Web site or as a page on Google) that delivers a search of only the user’s own selection of Web sites.

Yahoo!

Yahoo! was among the earliest Web sites that had the purpose of leading users to specific content on the Web. In the beginning, Yahoo! was exclusively a “Web directory,” a categorized list of selected Web sites. By 2000, however, it had begun a transformation to a portal site, having, in addition to the directory, over three dozen links to news, services, and other resources provided by Yahoo and its affiliates, including pages for shopping, auctions, phone numbers, a calendar, and more. From its earliest days, the Yahoo! homepage contained a search box, but results for that search came from a search of the directory, and later a search of Web databases from other search providers.

Yahoo!’s directory function became less and less central and in 2004 Yahoo! created its own database of Web pages. Though emphasis on “search” continued to increase and the emphasis on the directory declined significantly, Yahoo!’s main image continued to be that of a portal, with the emphasis on the wide range of other services provided by Yahoo! and its partners, including Yahoo!’s highly popular e-mail service and its sections on autos, finance, games, groups, health, job listings, maps, real estate, travel, and over 50 other content areas.

In the area of Web search, Yahoo! currently provides typical Web search features such as Boolean and field searching, though a continued absence of a link on its main page to its advanced search page, reinforces the impression of Yahoo!’s preference for a portal focus over search focus. It’s personalized portal page, My Yahoo!, is judged by some to be the most popular portal on the Web.[4]

In addition to Web search, Yahoo! offers searching of the following databases: news, images, video, maps, local (businesses), shopping, audio, jobs, Creative Commons, people (phone numbers and addresses), and travel reservations search.

MSN/Live search

Microsoft has made several attempts since the mid-1990s to produce a Web search engine that is competitive with Google and Yahoo!. The attempts, made available primarily through Microsoft’s MSN portal, have gone by a variety of names, including Microsoft Search, MSN Search, Windows Live, and, in 2008, Live Search (live.com). Search features have varied considerably and have at times been less robust than those of its competitors. Live Search presented some innovative features such as a design that allowed continuous scrolling through search results, but it, like some other features in the MSN search products, was short-lived. The 2008 version provided the typical Boolean and field searching options, plus some additional options such as “prefer,” by which the user can adjust the ranking weight for search terms, and “feed:” and “hasfeed:” which identify Web sites that contain RSS links on the user’s chosen topic. In addition to the search for Web pages, Live Search also offers searches for images, video, news, maps, health information, local (businesses), products, and travel.

AOL

AOL Search is the search engine found on AOL’s main portal page and is also available at search.aol.com. The search is provided in conjunction with Google and Web search results come from the Google database (but are typically fewer in number than when the search is done on Google itself). AOL Search also provides options for
searching images (using Google), video ("Powered by TRUVEO"), news, shopping, jobs, maps, movies, music, personals, travel, and yellow pages.

**Ask**

Ask, which was formerly AskJeeves, underwent a number of significant changes as it changed from the "question and answer" format of the original AskJeeves. Ask created a substantial Web database with fairly typical search functionality, though missing some features such as an OR Boolean function. In 2008, the company underwent a reorganization which produced some doubts among those who watch search engines as to Ask’s commitment to “search.” As well as its Web search, with Ask you can also search databases of images, news, maps, businesses, shopping, TV listings, events, videos, recipes, and blogs. Results pages for Web searches automatically incorporate results from multiple databases and provide a “binoculars” icon for previewing results without leaving the results page.

**Other general search engines**

There are a number of other general Web search engines, including GigaBlast, Exalead, and others. Exalead (http://www.exalead.com/search), from France, incorporates a number of features unavailable in other current search engines, including truncation ("words starting with"), phonetic spelling, approximate spelling, and NEAR. These are important to note because they are reflective of a level of sophistication of search techniques a bit closer to those found in commercial search services such as Lexis/Nexis, Factiva, and DIALOG, but not found in Web search engines.

**Specialty Search Engines**

Over the years, a variety of search engines have appeared that could be classified as “specialty” search engines. Among these there have been attempts to create search engines that focus on a particular topic or geographic location. In most cases, an examination of these showed that what was provided was more of a “directory” of selected sites than a broad ranging crawler-based search of Web pages for the specific topic or locality. On the other hand, there have been many successful attempts to produce search engines that provide searching for a particular format or type of document, such as images, video, blogs, forums, etc.

**News**

Searching of news databases is available from all the general Web search engines. There are numerous other Web sites that specialize in searching news content. Each of these have varying degrees of searchability, and from the research perspective it is important to note that the coverage can vary significantly, especially in regard to the number of news sources included, the time span for the content of the database, and the languages covered. Among the better-known news search engines are: NewsNow, Silobreaker, NewsExplorer, RocketNews, Topix.net, World Press Review Online, and NewsTin.

**Images**

The most commonly encountered image search engines are those that are included as databases provided by the general Web search engines, including Google, Yahoo!, Live Search, AOL, and Ask. As well as subject searching, most of these engines allow for Boolean, and narrowing by size, coloration, site, and adult-content filtering. On Google’s advanced image search page you can also narrow to news or photo content, or those that appear to include faces. Flickr (flicker.com), an image sharing Web site, has also gained extensive popularity as an image search engine. The extensive tagging of photos by Flickr users makes millions of images searchable. PicSearch provides an extensive collection of images from the Web and in addition to the above search criteria also allows narrowing to animated images. There are also image search engines such as Corbis, Fotssearch, and Stock. XCHNG which enable users (for a fee) to have use of photos from commercial photographers and photo archives.

**Video**

As with image searching, searching for video is available from major search engines, including Google, Yahoo!, Live Search, and AOL. Extensive searching of videos produced by individuals, as well as commercial video, is available from YouTube, the leading video-sharing site. Depending upon the search engine, options are provided for searching by Boolean, language, duration, domain/site/source, format, popularity, aspect ratio, and resolution, plus filtering for adult-content. Some video search engines specialize in video from TV, including news programs, interviews, etc. These include Blinkx (free) and TVEyes (fee-based). Both of these utilize voice-recognition technologies to create searchable transcripts for their video content.

**Forums**

Content found in forums (discussion groups, groups, newsgroups, etc.) can be utilized for a number of applications, ranging from hobbies to tracking terrorist activities, and there are search engines that specialize in
finding this category of document. Among the search engines that provide such access to forums from multiple sources across the Web are BoardReader and OMGILLI. (There are a number of other places where groups can be searched, such as Google, Yahoo!, Topica, Delphi Forums, but those sites focus on searching only the content that is hosted on their own Web sites.)

Other specialty search engines

There are still other categories of specialty search engines, including those for searching blogs and RSS feeds (examples: Technorati, IceRocket, Bloglines, and Google Blog Search), for searching podcasts (examples, Podcastdirectory.com, and Podcast.com), and for searching for information on people (examples, plip Search, Infobel, Yahoo People Search, Intelius, and PeopleFinders).

Visualization Engines

Visualization search engines are Web sites (or programs) that provide a very different “look” (literally) at search results. Instead of the traditional linear, textual list of retrieved items, results are shown on a map that spatially shows conceptual connections. Most current visualization engines utilize not a database of their own, but borrow one from other engines (such as Google, or Yahoo!) or other sites such as Amazon.com. Visualization has been, and continues to be, an area of extensive research and there are several sites that demonstrate various visualization approaches. The type of conceptual and visual mapping done by these sites can be especially useful for quickly exploring the concept possibilities, directions, and terminology for a particular search. It presents a “connect the dots” approach, enabling understanding relationships among the concepts found in various search results—rather than just browsing lists of results. Among the leaders in this area are Kartoo, TouchGraph, and Grokker, and Quintura.

Metasearch Engines

The term “metasearch engine” (or “metasearch site”) usually refers to Web sites that search multiple search engines in a single search. The degree of overlap (or lack thereof) between search engine results is something that professional searchers frequently consider and allow for as they search and searching more than one engine is a widely encouraged technique. Metasearch engines have been available since the 1990s and include sites such as Dogpile, Clusty, Ixquick, Mamma, Search.com, and many others. Each of these may provide additional benefits beyond just a compilation of results from more than one engine, for example, the “clustering” (categorization) of retrieved results, a feature that may not be provided by the target engines themselves. However, users should be aware of several shortcomings that may be encountered with these tools: 1) Most of the current metasearch engines do not cover the largest major engines, particularly Google and Yahoo!, which tend to block queries from metasearch engines; 2) Metasearch engines typically only return the first 10–20 results from any of the “target” engines; 3) Metasearch engine results often discard useful and search-relevant information found on the actual search engine’s results pages; 4) Metasearch sites, even if they do cover the largest engines, may be required by those engines to show paid listings first; and 5) Metasearch engines typically do not allow application of many of the search features available in the target engines themselves.

Metasearch engines should be distinguished from “comparison search” sites, such as Zuula.com and Twin-engine (twingine.no) which provide more of a side-by-side comparison of actual results from the target engines.

CONCLUSION

Web search engines have evolved significantly since they were first introduced in the early 1990s. The basic concept has remained the same, but the quality of results, the size of their databases, and the types of material that they include have increased dramatically. The total number of general Web search engines “in the race” has decreased and at present is dominated by one service, Google. Where the field of players has expanded is in the area of specialty search engines which focus on a specific type of Web “document.” What has evolved even more dramatically is the “mission” of search services, which particularly in the case of Google, has gone far beyond “search.” With advancing technologies, increasing interactivity of the Web, and a more and more Internet-centered society, users can expect continued, fast-paced innovation.

REFERENCES

webtrends.about.com/od/webportals/a/topten_portals.htm (accessed April 2009).

BIBLIOGRAPHY

