DOWNLOAD PDF WEB BASED SEARCH ENGINE

Chapter 1 : List of search engines - Wikipedia

The results of a search for the term "lunar eclipse" in a web-based image search engine A web search engine is a software system that is designed to search for information on the World Wide Web. The search results are generally presented in a line of results, often referred to as search engine results pages (SERPs).

SEM - search engine marketing Without sophisticated search engines, it would be virtually impossible to locate anything on the Web without knowing a specific URL. Search engines are the key to finding specific information on the vast expanse of the World Wide Web. Without sophisticated search engines, it would be virtually impossible to locate anything on the Web without knowing a specific URL. But do you know how search engines work? And do you know what makes some search engines more effective than others? When people use the term search engine in relation to the Web, they are usually referring to the actual search forms that searches through databases of HTML documents, initially gathered by a robot. There are basically three types of search engines: Those that are powered by robots called crawlers; ants or spiders and those that are powered by human submissions; and those that are a hybrid of the two. The crawler returns all that information back to a central depository, where the data is indexed. The crawler will periodically return to the sites to check for any information that has changed. The frequency with which this happens is determined by the administrators of the search engine. Human-powered search engines rely on humans to submit information that is subsequently indexed and catalogued. Only information that is submitted is put into the index. These indices are giant databases of information that is collected and stored and subsequently searched. This explains why sometimes a search on a commercial search engine, such as Yahoo! It will remain that way until the index is updated. So why will the same search on different search engines produce different results? Part of the answer to that question is because not all indices are going to be exactly the same. It depends on what the spiders find or what the humans submitted. But more important, not every search engine uses the same algorithm to search through the indices. The algorithm is what the search engines use to determine the relevance of the information in the index to what the user is searching for. One of the elements that a search engine algorithm scans for is the frequency and location of keywords on a Web page. Those with higher frequency are typically considered more relevant. But search engine technology is becoming sophisticated in its attempt to discourage what is known as keyword stuffing , or spamdexing. Another common element that algorithms analyze is the way that pages link to other pages in the Web. By analyzing how pages link to each other, an engine can both determine what a page is about if the keywords of the linked pages are similar to the keywords on the original page and whether that page is considered "important" and deserving of a boost in ranking. Just as the technology is becoming increasingly sophisticated to ignore keyword stuffing, it is also becoming more savvy to Web masters who build artificial links into their sites in order to build an artificial ranking. The first tool for searching the Internet, created in , was called "Archie". It downloaded directory listings of all files located on public anonymous FTP servers; creating a searchable database of filenames. A year later "Gopher" was created. It indexed plain text documents. The first actual Web search engine was developed by Matthew Gray in and was called "Wandex".

DOWNLOAD PDF WEB BASED SEARCH ENGINE

Chapter 2 : 5 Web Search Evaluator Jobs That Pay $$15 Per Hour

A search engine is a web-based tool that enables users to locate information on the World Wide Web. Popular examples of search engines are Google, Yahoo!, and MSN Search. Search engines utilize automated software applications (referred to as robots, bots, or spiders) that travel along the Web, following links from page to page, site to site.

Twitter Advertisement Wherever you go, your face exposes you. Facial recognition in combination with surveillance cameras is a powerful tool that can track your every step. Search engines are becoming ever smarter in managing massive amounts of data. Face search and facial recognition are just a few of many tools that target individuals. All public data combined, they can quickly unravel what an individual has been up to. Here are three face search engines that may give you a thrill. Rather than a keyword, you can use an image to search for similar images. Click the camera icon to search by image. You can either paste the image URL or upload an image and Google will find similar images. Moreover, you can make Google search for faces only by adding a small bit of code. This will further improve the results of your face-related search. Unfortunately, the feature is limited to look-alike celebrities. For demonstration purposes, I used my own headshot. While PicTriev correctly identified me as overwhelmingly female, the number one match was Jason Clarke. The age estimation of 30, however, is very flattering. It works much better if you search for a celebrity image. PicTriev also lets you compare the similarity of two faces or estimate whether photos of two faces are the same person. Click the meter icon in the top right, upload two photos, select similarity or identity, and let PicTriev do its calculations. Before you add photos, be sure to follow the instructions on formatting for best results. The demos using celebrity faces like Angelina Jolie or Zac Efron look promising. PimEyes will find the original photos, as well as other shots of Aniston. Or does the algorithm take image resolution, size, brightness, and other digital alternations into account? The GIF below illustrates the process. And FindFace promises to find anybody on VK. You can fine-tune the initial result by gender, age, location, and relationship status. In my case, it did find several surprising look-a-likes, but nothing too crazy. He took photos of strangers on the subway, found them on VK. The project highlights how invasive a simple photo can be. Betaface â€" Facial Recognition Demo Betaface offers facial recognition How Facial Recognition is Invading your Privacy How Facial Recognition is Invading your Privacy Is facial recognition -- a staple of science fiction for the past 50 years -- really a means of oppression, part of a surveillance state and a form of control? Or is it more useful This tool is useful for uploading and comparing photos in bulk. Both of these slow down the processing, but will increase the quality of your matches. This tool is powered by Microsoft. Interestingly, the site highlighted a small section of the photos to underscore its decision. What Does Your Face Reveal? Face recognition and search tools have a range of useful applications. Not only can they help the police identify suspects from security camera footage. They can also help professional photographers or media companies index visual material and build large and easy to search archives. But there are a few risks associated with them. Read More and keys. Not too long ago, the Facezam viral marketing scam highlighted what face recognition could do to your privacy. Essentially, FindFace for Facebook. Today, the impact and dangers of online privacy breaches are major. These few resources explain the pitfalls clearly and concisely. How do you keep your face private?

Chapter 3 : local area network - Web based file search in the lan? - Server Fault

This is a list of search engines, including web search engines, selection-based search engines, metasearch engines, desktop search tools, and web portals and vertical market websites that have a search facility for online databases.

These data types present new challenges to big data science. Here, we present GeNemo, a web-based search engine for functional genomic data. The user can input any complete or partial functional genomic dataset, for example, a binding intensity file bigWig or a peak file. GeNemo reports any genomic regions, ranging from hundred bases to hundred thousand bases, from any of the online ENCODE datasets that share similar functional binding, modification, accessibility patterns. This is enabled by a Markov Chain Monte Carlo-based maximization process, executed on up to 24 parallel computing threads. GeNemo is available at www. The immediate outputs are DNA sequences, which does not become biologically meaningful until being further processed. After processing, the data are typically stored as genome-wide intensities e. These processed data provide functional information of the genome. The formats of these processed data are very different from those storing DNA sequences 2 , 3. Thus, functional genomic data bring new computational challenges. A pressing challenge is to effectively search functional genomic data from online data repositories. There are at least two conceivable means to search these data. It would be straightforward to use Google for such a task. To better appreciate the difference of the two types of searches, let us compare the functional genomic data files with video files. Again, it should be noted that the data formats of concern here are not DNA sequences, but rather genome-wide intensities. There is yet no software for executing the second type of searches online. Here we present GeNemo. Providing GeNemo with a bigWig or a peak file, users can search online for functional genomic data that share similar patterns at any genomic regions. Alternatively, the user can designate any online bigWig or peak file as the input to initiate the search, by providing the URL of the input file to GeNemo. The search results are reported to the user in two steps. The initial return is a synopsis of all found datasets and the corresponding genomic regions that partially matched with the input data. If the user clicks on an item in the synopsis, GeNemo will retrieve the specific regions of the found dataset and display it side-by-side with the user input data. Although the actual data are stored on remote servers, the current release of GeNemo offers nearly instant data retrieval and display to users. This allows the user to discover, for example, within certain genomic regions, the pattern of binding of a protein is similar to that of an epigenetic modification. This search is achieved by comparing the input data to every indexed target dataset, using parallel computing. Between the input data and every target dataset, the search is based on a maximization process, which maximizes a local similarity score Rt over the start location i1 and the end location i2 of a genomic region, namely: The local similarity score is defined as: We will describe the initialization, the auxiliary chain and the acceptance-rejection rule. We devised the auxiliary chain as follows. Such a scenario never happens in practice because there are no functional signals in telemeric regions due to their degenerative sequence. Our actual chain is generated by superimposing the following acceptance-rejection rule onto the auxiliary chain. It is easy to see that our actual chain is ergodic and reversible. Parallel computing To minimize response time, the current release of GeNemo creates up to 24 parallel computing threads for every search. The target datasets are separated into up to non-overlapping subsets, and each subset is handled by a computing thread. We tested a series input files using 1, 10 and 20 computing threads, and found that 20 threads typically accelerated the search by 7â€”9 folds Supplementary Table S4. These datasets are located on remote data servers and are not managed by GeNemo developers. We developed a program to index these remote datasets and create a metadata on the GeNemo server. Anticipating future data releases, we will use this program to index additional datasets and thus expand the pool of target files. GeNemo has a simple user interface Figure 1A. The minimum required input from the user is a bigWig file or a peak file in BED format and choice of the species. GeNemo will automatically detect the file type. With the input files, the user can click the search button to initiate the search. View large Download slide GeNemo screens. A Input screen, references and the data file is needed to search; B Results screen, the coordinates and tracks matching are shown; C Visualization screen, the input track, genes, matching tracks shown as results and other annotations

are shown. The more sophisticated users can take advantage of a few optional inputs. The user can supply an email address. In this case, GeNemo will send an email with a link to the search results. For example, the user can constraint the search space to certain cell types or certain kinds of experiments. GeNemo returns the search results typically in minutes. In addition, a web link to the results is sent to user, if an email address was provided. GeNemo reports the search results in two steps. The initial return is a synopsis of all found datasets Figure 1B. Each entry reports a genomic region where a found dataset exhibited similar patterns to the input. This is an intuitive design imagine clicking on a Google found entry to browse the found website. This action will invoke a genome browser-like display 3. The user input data will be displayed in parallel to the found data, centered at the genomic region where a pattern match was found Figure 1C. Even though the found data are stored on remote data servers that are not managed by GeNemo developers, we optimized the data retrieval strategy such that the user would not experience noticeable waiting time for data visualization. GeNemo instantly displays the input and the found data, even when the user navigates to other genomic regions. Simulation analysis We carried out three simulation studies to test algorithm performance. In the first study, we simulated the cases where all the genomic regions with matched signals were completely known gold standard was available. We generated three datasets. Each dataset was composed of one input file and one target file. Both simulated files covered the entire length of the human genome. Each matched region contained a random number of signal segments between 1 and 10 , and each signal segment had a random length between and bp. These matched regions were inserted to the simulated genome at random positions. The input and the target files had the same matched regions at the same locations. For the rest of the genome, we inserted additional signal segments at random locations to the target files, and kept the input file free of any additional randomly inserted signals. These three pairs of input and target files were subjected to the search. Consistent to the simplicity of the simulated data, all the matched regions in datasets 1 and 2 were found; the algorithm did not report any additional matched regions Supplementary Table S1. In dataset 3, the algorithm only reported of the matched regions. This was expected because our program was set to only output the top matches. The second simulation was carried out in a similar fashion, except that additional signal segments were randomly added to both the input and the target files Supplementary Table S2. This is to better mimic the noises in actual experiments and the errors in data processing. If we consider two random typos that happened to match each other, from the perspective of a text search, it would be correct to find a match between the two typos. We recognize that this choice would lead to underestimation of the precision of the algorithm. Still, the algorithm found the majority of the matched regions with high precision Supplementary Table S2. These regions were regarded as positive regions. This produced a synthetic dataset Dataset 7, Supplementary Table S3. We obtained top GeNemo returned regions. Data applications We will present a data example. We retrieved a public dataset from the GeNemo indexed files. GeNemo returned similarity regions Supplementary Figure S1. Some of these regions were precisely decorated with H3K4me3, an epigenetic mark associated with transcriptional activation CH12 tracks, Figure 2B or bivalent domains 9. Although E2F4 was generally considered a transcriptional repressor, these data suggest that E2F4 may also contribute to transcriptional activation or bivalent regulation in blood cancers. Consistent to this idea, overexpression of E2F4 and its transcriptional cofactors led to both transcriptional repression and activation in lymphoblastoid cell lines A Result in chromosome B Result in chromosome X. This makes us to posit that the identified regions are regulatory sequences used for transcriptional control in normal tissues; some blood cancers potentially used these regulatory sequences for transcriptional activation, which were attached with E2F4. This example illustrates that GeNemo may be used as a hypothesis generating tool. However, we recognize the gap between any hypotheses generated by association and functional validation. However, except for medical records, many types of biomedical data cannot be searched as text. Functional genomic data are a point in case. Despite their increasing importance to biomedical research, to our knowledge, there is yet no online search engine for them. We note that the search methods for genomic sequences text based or string based are very different from, and probably irrelevant to the searches for functional genomic data. The latter represents the extent of molecular activities at every genomic location. Therefore, it requires new computational engineering efforts that are customized to this data type. GeNemo is the first online search

engine for functional genomic data. Many aspects of the design and the implementation of this search engine were made to optimize the speed.

Chapter 4 : 3 Fascinating Search Engines That Search for Faces

Many web sites found through Internet search engines contain licensed, proprietary information and require you to logon with a user account. You must already be a member or pay for a subscription in order to access the material from these web sites.

Introduction by Maryam Allahyar spring The Internet is a great resource that is available to anyone who has access to it. The World Wide Web contains information from all over the world and is useful for all ages and for all purposes - from very complex things to very simple things, such as door-to- door directions and maps. The information contained on the Web can also be useful for academic research. Although other sources such as PsychLit are very useful too, it is important to realize that many sources are not updated frequently enough. ERIC, on the other hand, is updated on monthly basis. Overall, the traditional print resources available in the library are great, but Web resources and ERIC should not be neglected. Lynx is a non-graphical browser that gives a computer user only text on the computer screen during research on the Web. Because it is only text, the advantage of using Lynx is speed: So if the purpose of your research is to get information in text style, Lynx is a good way to go. Of course, a disadvantage of using it is the inability to view graphics or video - or hearing sound clips. To access Lynx on many servers, type "lynx" at the computer prompt and use the menu instructions. By following simple commands which appear on the bottom of the screen, what you can do by using a graphical browser like Netscape Navigator can also be done with Lynx. Lynx also maintains the location that you visit during the current session. Creating a "book mark" file that is a list of specified URLs is also available. You can print of save documents to a file as well. All these can be done in a relatively short time compared to a graphical browser. If you do not have a lot of time, or do not need graphical information, Lynx is the best way to do research on the Web. The biggest advantage in using the World Wide Web as a source for research is that it lets us look at specific topics from an interdisciplinary perspective. Due to the large volume of published literature in the library, researchers have had a tendency to stay within their own fields when they search for references. By doing so, they may have been restricting themselves to their own fields and may have had little idea of the kind of studies in other disciplines that may be helpful. This also happens within a single discipline such as Psychology. The developmental psychologists have their own journals while the social psychologists have others, with little chance for the various disciplines to integrate. In looking for information on the Web, searches are often more general in nature, which may bring us information that otherwise may not have caught our attention. Another advantage in using the Web for academic research is the ability to gain access to the most current information. Since studies can take months or years to get published, data can be outdated by the time it reaches the shelves of our libraries. Direct access to current information increases the effectiveness of scientists in their search for information in their areas of interest. This is where the third advantage of using the Web comes in: Although we have been and are still somewhat restricted by only having faculty members as our direct source of guidance in our research, this may be slowly changing as we gain the ability to contact many researchers via the Internet. The only disadvantage that I have experienced using the Web for academic research is the overwhelming amount of information that is available. One can easily get lost in the seemingly infinite amount of titles, abstracts and texts. I have found it helpful to keep a piece of paper in front of me that states exactly what I am looking for - this helps to help keep me focused. Experiences on the Internet by Patricia A. One of the best features of the Internet is its convenience. I am a "commuter student" because I live 75 miles away from the campus. But when I use the Internet, the only "traffic" problem I encounter is the occasional busy signal when I try to log on. With the Internet, I can do my research from the comfort and convenience of my own home. I can get up on a Saturday or Sunday morning, make myself a cup of coffee, go into the study, and begin doing my research. In addition, using the Internet saves me valuable time and allows me to manage my resources better: Not driving to the campus saves me about three hours round trip driving time. Therefore I can make less trips to campus and when I do drive down, I already know exactly which journal articles or books that I need as well as whether or not our library even has them! The only negative features I have found using the Internet for academic

research are the abundance of information available and the nebulous origin of some of the information. Occasionally when doing a search, even on a very specific item or topic, I will get hundreds and even thousands of hits! There is so much information available that it can be overwhelming and I sometimes get "lost" in all of it - and even forget why I went there in the first place! This problem can usually be solved by narrowing the topic down as much as possible and by selecting the appropriate Search Engine. I have found that certain Search Engines are more useful for some subjects while other search engines are more appropriate for other subjects. For instance the WebCrawler seems to be a good Search Engine to use for general or non-academic information, while Alta Vista usually provides the best information for topics related to Psychology. At times it is also difficult to authenticate the validity of the information found on the Web. Millions of files available on the Web make information available for research purposes, and much of this information may not readily be available through other means. However, with the tremendous amount of information available on the Web, the task of finding relevant and reliable information can be difficult and sometimes very time-consuming. Since the amount of information on the Web can be overwhelming, a successful researcher must often use efficient and creative search strategies. Search Engines, such as Webcrawler, provide a valuable tools with which a researcher is often able to locate specific information on a subject of interest. Creative use of search strategies can allow a researcher to find new and interesting sources of information that can lead to productive results. Much of the information obtained may not be directly useful for research purposes, so the researcher must carefully sift through the various files carefully in order to find useful information resources. Because of the questionable reliability of some of the information available on the Web, a researcher must be cautious about the information obtained. It is important to followed up the information gathered through the Web by using traditional sources of research information, such as published articles in recognized journals. Search Engines allow you to search millions of sites on the World Wide Web. Since each Search Engine uses a different method for locating information, it is better not to limit yourself to just one. Some Search Engines e. Other Search Engines are based only on headers and page titles, and some e. Do not neglect smaller Search Engines because they may be very useful and may specialize in the area of your topic. Be sure to read the description for each Search Engine before you use it. Try to begin each search with a "fast search" if possible since a comprehensive search may take too long. If you conceptualize some key words, phrases or concepts before beginning your search, you may be able to avoid retrieving information that is not relevant to your topic. Try using synonyms or related words if you are not finding the information that you are looking for. AltaVista contains a very large Web search database. Since it is so comprehensive, you may have to spend more time looking, but you may find every important reference relevant to your topic. Type in a primary word or words. Type in both primary key words and concepts for your topic. The ability to search for concepts is an added feature that many Search Engines like AltaVista does not have. Yahoo and Magellan are classified as catalogues rather than as Search Engines. Catalogues are ideal for broad category search of established sites. Yahoo provides an hierarchical subject index and allows you to begin with a general topic, then become more specific. You can also search the Yahoo index instead of searching the entire Web. If you do not find any information on your subject, then you can switch to a Search Engine like AltaVista. FirstSearch is another way to use the Internet for research. As a research tool, FirstSearch is a starting point of gaining information on specific topics in the published journal literature. Of course, in order to get the entire copyrighted article, you have to find the journal in the library or request a copy of the article through Interlibrary Loan. But this preliminary work can be done from home and when you arrive at the library, you have a list the list of articles that you will need for your research. Until recently most of the available time an efficient student had was allocated in trips to the nearby library. Various search methods have been at our disposal for years, even decades, e. Most students who are comfortable with using computers have already discovered the many resources found online. Since this self-selection process leaves behind students who are not at ease with computers, a huge disparity is occurring in our universities. And students who attend technologically-advanced schools like CSUN can benefit from all these advantages, although perhaps only those students probably grew up with access to computers take advantage of them. That could mean that a student who did not have this sort of opportunity might not have the same educational experience

in college. So better-prepared students coming to the university have an additional advantage over students who may not have had the benefits of computer literacy in high school. How do we break this cycle so that all students can benefit from this technology? My personal experience with computers even a few years ago was quite limited. I recall not knowing how to do even the most basic task without assistance. Looking back to what forced me to delve into the world of technology, I must say it was my own curiosity. I suggest to the academic professionals interested in having their students move into the 21st century prepared: Another possibility would be to require that so many hours be spent in the computer lab, per class credit. One professor I know requires his students to all log into the University server at the same time - instead of class meetings, he holds class discussions in this manner. Listing the many benefits gained from online research shows how logical it can be to use these services: And in searching for information, one topic may lead to another and expose the student to a much more interdisciplinary approach to a topic. Unfortunately, logic here may not prevail. Fear of the unknown is most likely the factor which keeps the majority of students away from the wonderful benefits of online research. The bottom line here is providing students with the tools for success after graduation, and that responsibility lies with the instructor. If the instructor is not comfortable with the electronic medium, perhaps growing along with the students is a possibility!

Chapter 5 : Yahoo Search - Web Search

Comparing Web search engine performance in searching consumer health information: evaluation and recommendations. G Wu and J Li Shiffman Medical Library, Wayne State University, Detroit, Michigan , USA.

To make it easier, these 25 search engines can do the work for you. From searching the PDR to finding journal articles, you are sure to find helpful sites to bookmark on your computer from the list below. Gathering information from many of the top medical professional sites such as PubMed, NIH, and Merck, this search engine provides information from peer level sources. They also offer a "reference desk of hard-to-find medical resources. MedNets offers a search specifically for medical professionals in addition to one for the general public. The professional version can be customized by specialty. Sponsored by the University of Iowa, this site allows you to search for a disease or general health topic alphabetically to get links to a variety of online articles and photos about each disease. You can also view photos from classic medical books via this site. Many of the full articles are available for a fee, while others are free of charge. In addition to the powerful search engine, they also have Subject Guides under the "eResources" section that offers links to topics ranging from Alternative Medicine to Grants and Funding to Writing and Publishing. Access one of the most well-known and frequently used resources for FDA-approved prescription drugs. You must register to use this service, but it is free of charge to all U. The site also offers free download to your PDA. Search this world wide registry of "federally and privately supported clinical trials. There is also a link for professionals who want to register their trials with this site. Out of the U. They also provide links to a handful of resource brochures and training site tutorials. Search for diagnoses and treatments while staying on top of health and wellness with this medical site. Healthline searches the best of the health sites available on the Internet, reducing your search time. Browse by topic or use their keyword search. Also visit the top 10 diagnostic tests or browse their dictionaries. This database provides access to most of the major news and research publications in the life sciences. Almost half of the full-text articles available are free of charge. You must register to have access to all the features at this site. Sponsored by the U. National Library of Medicine and the National Institutes of Health, this database provides access to citations going back for the past 40 years. You must register free with NCBI before having access to this powerful search engine. Specifically for healthcare professionals, this website will keep you on top of the latest in your field. They feature four interactive journals for primary care, managed care, emergency medicine, and pediatrics. Additionally, you can search across all the available databases which gather information from journal articles, books, online books, and more. A part of WebMD, this site is geared to the medical professional. Describing itself as an "open access comprehensive medical textbook," eMedicine offers over 6, clinical articles written by contributing physicians. Geared toward medical professionals and those in the biotechnology field, this search engine finds information from journals, organizations, and databases. Use their tools, directories, dictionaries, and read the blog for even more information. Search for a variety of information with this medical search engine. Available are specific searches for medications, information in specific journals, medical definitions, medical books, articles, and web searches and much more. There is also an updated medical feed right on the home page to keep you abreast of medical news while you perform your searches. Not only can you search for specific topics on any imaginable health topic, but once your results pop up in the window, you can click on different tabs to find conference information, news, and images that relate to your query without re-typing the keyword. Browse this guide by specific antibiotics, diagnosis, pathogens, management, and vaccines. Look for free, updated CME programs that are also available. Updated daily, this online resource monitors generic prescription drugs and posts updates with new generic drug approvals, application approvals, discontinuations, patents, and exclusivity information. Use one of five different search types to find the medicine you want to learn about. Working in any field in the health profession, having access to information for hospitals is always handy. Keep this search nearby for the next time you need to contact a hospital anywhere in America. Search for hospitals by area code, zip code, or by city and state. Specifically geared to searches for genes and proteins, this search engine relies on text mining PubMed articles to find any source with a specific gene or

protein mentioned in it. For any researcher or physician working in genetics, this search engine will keep up with all that is happening in the field for you. Check out their page of awards, reviews, and comments. Monitoring medical journals, this site offers both journal searches as well as short, daily email updates. Choose between Primary Care Physician, Cardiology, Gastroenterology, and other specialties to specialize the content according to the type of medicine you practice. For other healthcare professionals, they also offer subscriptions for non-physicians as well. This site reviews over journals and provides a search by specialties and subspecialties. In addition, they offer conference and job listings. You must register, but it is free of charge. Medical professionals and laypersons alike may sign up for newsletters within their specialty. For both professionals and non-professionals, Medscape offers searches in a number of databases. They also offer specialized sections for non-physician professionals such as pharmacists, med students, and nurses, as well as a specialty section with information that is specialty-specific. Registration is required, but is free of charge. Did you enjoy this article? Leave a Reply Mail will not be published required Website.

Chapter 6 : How a Search Engine Works

Another way to define Search Engines is special web sites designed to serve as searchable "indices" for the Search engines typically employ computer programs that continually collect information about web pages and organize this information in a searchable database.

Best for college level research. When you need to find credible information quickly. Best for personal information needs including shopping and entertainment. When you have time to more carefully evaluate information found on the open web. All material in database is evaluated for accuracy and credibility by subject experts and publishers. Reviewed and updated regularly. Lack of control allows anybody to publish their opinions and ideas on the Internet. Not evaluated for the most part. Need to more carefully evaluate web sites for bias, accuracy, and completeness. Many sites are not updated regularly and can become outdated. To access the Reynolds Libraries databases from off-campus, you will need to logon with your My Reynolds username and password. Most information found through a search engine is free. Library databases cannot be accessed through search engines or the open web. Many web sites found through Internet search engines contain licensed, proprietary information and require you to logon with a user account. You must already be a member or pay for a subscription in order to access the material from these web sites. Usability The organization and various search capabilities of library databases allow users to search for and retrieve focused and relevant results. Less ability to search for and retrieve precise results using search engines like Google. Most material remains in database for a significant length of time and can be easily retrieved again. Web site content can often change. Web pages and sites may disappear for a number of reasons. May not be able to retrieve the same content at a later time. Citing Many databases include a citation tool that will automatically generate an APA or MLA style reference for the article you select. Most web sites found on the open web do not provide a citation tool or an already formatted APA or MLA style reference for the web pages on their site. You will need to start your citation from scratch using APA or MLA style manuals or handouts from your instructor or the library.

Chapter 7 : 25 Search Engines Every Medical Professional Should Bookmark | www.nxgvision.com

I want a web-based search engine that crawls the LAN with an ultra-simple webinterface like Google. The crawlers should be able to index the most common file types doc, docx, pdf, xls, txt and the like.

Since users may employ special operators in their query, including Boolean, adjacency, or proximity operators, the system needs to parse the query first into query terms and operators. These operators may occur in the form of reserved punctuation e. In the case of an NLP system, the query processor will recognize the operators implicitly in the language used no matter how the operators might be expressed e. At this point, a search engine may take the list of query terms and search them against the inverted file. In fact, this is the point at which the majority of publicly available search engines perform the search. Steps 3 and 4: Stop list and stemming. Some search engines will go further and stop-list and stem the query, similar to the processes described above in the Document Processor section. How each particular search engine creates a query representation depends on how the system does its matching. If a statistically based matcher is used, then the query must match the statistical representations of the documents in the system. Good statistical queries should contain many synonyms and other terms in order to create a full representation. At this point, a search engine may take the query representation and perform the search against the inverted file. More advanced search engines may take two further steps. Since users of search engines usually include only a single statement of their information needs in a query, it becomes highly probable that the information they need may be expressed using synonyms, rather than the exact query terms, in the documents which the search engine searches against. Therefore, more sophisticated systems may expand the query into all possible synonymous terms and perhaps even broader and narrower terms. This process approaches what search intermediaries did for end users in the earlier days of commercial search systems. Back then, intermediaries might have used the same controlled vocabulary or thesaurus used by the indexers who assigned subject descriptors to documents. Today, resources such as WordNet are generally available, or specialized expansion facilities may take the initial query and enlarge it by adding associated vocabulary. Query term weighting assuming more than one query term. The final step in query processing involves computing weights for the terms in the query. Sometimes the user controls this step by indicating either how much to weight each term or simply which term or concept in the query matters most and must appear in each retrieved document to ensure relevance. Leaving the weighting up to the user is not common, because research has shown that users are not particularly good at determining the relative importance of terms in their queries. Second, most users seek information about an unfamiliar subject, so they may not know the correct terminology. Few search engines implement system-based query weighting, but some do an implicit weighting by treating the first term s in a query as having higher significance. After this final step, the expanded, weighted query is searched against the inverted file of documents. Since making the distinctions between these models goes far beyond the goals of this article, we will only make some broad generalizations in the following description of the search and matching function. Those interested in further detail should turn to R. Searching the inverted file for documents meeting the query requirements, referred to simply as "matching," is typically a standard binary search, no matter whether the search ends after the first two, five, or all seven steps of query processing. While the computational processing required for simple, unweighted, non-Boolean query matching is far simpler than when the model is an NLP-based query within a weighted, Boolean model, it also follows that the simpler the document representation, the query representation, and the matching algorithm, the less relevant the results, except for very simple queries, such as one-word, non-ambiguous queries seeking the most generally known information. After computing the similarity of each document in the subset of documents, the system presents an ordered list to the user. The sophistication of the ordering of the documents again depends on the model the system uses, as well as the richness of the document and query weighting mechanisms. More sophisticated systems will go even further at this stage and allow the user to provide some relevance feedback or to modify their query based on the results they have seen. If either of these are available, the system will then adjust its query representation to reflect this value-added feedback and re-run the search with the improved query to

produce either a new set of documents or a simple re-ranking of documents from the initial search. What Document Features Make a Good Match to a Query We have discussed how search engines work, but what features of a query make for good matches? While most often true, several situations can undermine this premise. Think of words like "pool" or "fire. Many search engines give preference to words found in the title or lead paragraph or in the metadata of a document. Terms occurring in the title of a document or page that match a query term are therefore frequently weighted more heavily than terms occurring in the body of the document. Similarly, query terms occurring in section headings or the first paragraph of a document may be more likely to be relevant. Web-based search engines have introduced one dramatically different feature for weighting and ranking pages. Link analysis works somewhat like bibliographic citation practices, such as those used by Science Citation Index. Link analysis is based on how well-connected each page is, as defined by Hubs and Authorities, where Hub documents link to large numbers of other pages out-links , and Authority documents are those referred to by many other pages, or have a high number of "in-links" J. Google and several other search engines add popularity to link analysis to help determine the relevance or value of pages. Popularity utilizes data on the frequency with which a page is chosen by all users as a means of predicting relevance. While popularity is a good indicator at times, it assumes that the underlying information need remains the same. Some search engines assume that the more recent the information is, the more likely that it will be useful or relevant to the user. The engines therefore present results beginning with the most recent to the less current. While length per se does not necessarily predict relevance, it is a factor when used to compute the relative merit of similar pages. So, in a choice between two documents both containing the same query terms, the document that contains a proportionately higher occurrence of the term relative to the length of the document is assumed more likely to be relevant. When the terms in a query occur near to each other within a document, it is more likely that the document is relevant to the query than if the terms occur at greater distance. While some search engines do not recognize phrases per se in queries, some search engines clearly rank documents in results higher if the query terms occur adjacent to one another or in closer proximity, as compared to documents in which the terms occur at a distance. While this may be useful, if the search engine assumes that you are searching for a name instead of the same word as a normal everyday term, then the search results may be peculiarly skewed. Imagine getting information on "Madonna," the rock star, when you were looking for pictures of madonnas for an art history class. Summary The above explanation lays out the range of processing that might occur in a search engine, along with the many options that a search engine provider decides on. Up till now, search engine providers have mainly opted for less, versus more, complex processing of documents and queries. The typical search results therefore leave a lot of work to be done by the searcher, who must wend their way through the results, clicking on and exploring a number of documents before finding exactly what they seek. The typical evolution of products and services suggests that this status-quo will not continue. Searchers should keep watching for the best and pursuing it.

Chapter 8 : What is a search engine? | DesignHammer Website Design and Development in North Carolina

A search engine will allow you to search for information found on the web using simple keywords, but they lack the advanced search capabilities provide by most databases.

Have you ever considered working as a home based web search evaluator? If so, this is a great way to put a little extra money in your pocket. What is a Web Search Evaluator? So I know many are wondering what is a web search evaluator. The role of a web search evaluator is to ensure that search engines are always updated. For instance, if someone decides to search for a particular keyword or phrase, your job is to make sure the web page results remain accurate. Working as a web search evaluator is one of the best paying part-time jobs you can do from home. You would typically need a computer with high-speed Internet and excellent web search skills. Please keep in mind that it is best to have a noise-free environment to be effective at this job, in case you have small kids or pets. Below is a list of 5 companies that hire web search evaluators to work from home. Who Hires Web Search Evaluators? It has been around for over 10 years and only provides jobs to natives residing in the U. You must be good at searching the web and commit to working a minimum of 20 hours per week. You are required to pass a two-part evaluation exam to be considered for a position. Positions are open to those who live in the U. You will need to be very good at searching the web to be an ideal candidate for a position in this company. Even so, there are also some technical requirements for the job which are specified on their site. The schedule is quite flexible but you have to work a minimum of 20 hours per week. They hire part-time telecommuters to evaluate websites for Google. To find their current job openings, visit Career Builder, Indeed, or Craigslist. The position is part-time and allows you to work from anywhere for at least hours per week. Looking for more scam-free work at home jobs? I hope this post will give hope to those looking for a non-phone job with great pay. Consider all of these jobs as a good way to earn a side income, since the work is mostly part-time. If you have any prior experience with any of these companies, feel free to share your thoughts in my comment section below. You May Also Like:

Chapter 9 : AWS | Amazon CloudSearch - Search Service in the Cloud

With over Million search queries operated per day by Yandex, it is the fourth largest Web Search engine in the World and the leading search engine in Russia. Although it does feature a English localized version, www.nxgvision.com is by far the most popular search engine.

Searching [15] Web search engines get their information by web crawling from site to site. The "spider" checks for the standard filename robots. Due to infinite websites, spider traps, spam, and other exigencies of the real web, crawlers instead apply a crawl policy to determine when the crawling of a site should be deemed sufficient. Some sites are crawled exhaustively, while others are crawled only partially". The associations are made in a public database, made available for web search queries. A query from a user can be a single word. The index helps find information relating to the query as quickly as possible. Between visits by the spider, the cached version of page some or all the content needed to render it stored in the search engine working memory is quickly sent to an inquirer. If a visit is overdue, the search engine can just act as a web proxy instead. In this case the page may differ from the search terms indexed. High-level architecture of a standard Web crawler Typically when a user enters a query into a search engine it is a few keywords. The real processing load is in generating the web pages that are the search results list: Every page in the entire list must be weighted according to information in the indexes. These are only part of the processing each search results web page requires, and further pages next to the top require more of this post processing. Beyond simple keyword lookups, search engines offer their own GUI- or command-driven operators and search parameters to refine the search results. These provide the necessary controls for the user engaged in the feedback loop users create by filtering and weighting while refining the search results, given the initial pages of the first search results. For example, from the Google. Boolean operators are for literal searches that allow the user to refine and extend the terms of the search. The engine looks for the words or phrases exactly as entered. Some search engines provide an advanced feature called proximity search , which allows users to define the distance between keywords. As well, natural language queries allow the user to type a question in the same form one would ask it to a human. While there may be millions of web pages that include a particular word or phrase, some pages may be more relevant, popular, or authoritative than others. Most search engines employ methods to rank the results to provide the "best" results first. How a search engine decides which pages are the best matches, and what order the results should be shown in, varies widely from one engine to another. There are two main types of search engine that have evolved: The other is a system that generates an " inverted index " by analyzing texts it locates. This first form relies much more heavily on the computer itself to do the bulk of the work. Most Web search engines are commercial ventures supported by advertising revenue and thus some of them allow advertisers to have their listings ranked higher in search results for a fee. Search engines that do not accept money for their search results make money by running search related ads alongside the regular search engine results. The search engines make money every time someone clicks on one of these ads. In Russia, Yandex commands a marketshare of Taiwan are the most popular avenues for internet search in Japan and Taiwan, respectively. Biases can also be a result of social processes, as search engine algorithms are frequently designed to exclude non-normative viewpoints in favor of more "popular" results. Several scholars have studied the cultural changes triggered by search engines, [33] and the representation of certain controversial topics in their results, such as terrorism in Ireland [34] and conspiracy theories. This leads to an effect that has been called a filter bubble. The term describes a phenomenon in which websites use algorithms to selectively guess what information a user would like to see, based on information about the user such as location, past click behaviour and search history. This puts the user in a state of intellectual isolation without contrary information. According to Eli Pariser , who coined the term, users get less exposure to conflicting viewpoints and are isolated intellectually in their own informational bubble. Pariser related an example in which one user searched Google for "BP" and got investment news about British Petroleum while another searcher got information about the Deepwater Horizon oil spill and that the two search results pages were "strikingly different". More than usual safe search filters, these Islamic web portals categorizing websites into

being either " halal " or " haram ", based on modern, expert, interpretation of the "Law of Islam". ImHalal came online in September Halalgoogling came online in July These use haram filters on the collections from Google and Bing and others. SeekFind filters sites that attack or degrade their faith. While search engine submission is sometimes presented as a way to promote a website, it generally is not necessary because the major search engines use web crawlers, that will eventually find most web sites on the Internet without assistance. They can either submit one web page at a time, or they can submit the entire site using a sitemap , but it is normally only necessary to submit the home page of a web site as search engines are able to crawl a well designed website. There are two remaining reasons to submit a web site or web page to a search engine: Some search engine submission software not only submits websites to multiple search engines, but also add links to websites from their own pages. However, John Mueller of Google has stated that this "can lead to a tremendous number of unnatural links for your site" with a negative impact on site ranking.