A Proposed System for Search Engine Using Web Annotation

Mrs.Sonal Bawankule[1]
Ms.Aditi Bhaje[2], Ms.Arshiya Sheikh[3], Ms.Prachita Darode[4],Ms. Radhika Sangole[5]
[1]Assistant Professor, *Computer Science and Engineering, Priyadarshini JL College of Engineering, Nagpur, Maharashtra, India*
[2,3,4,5] *Student, Computer Science and Engineering, Priyadarshini JL College of Engineering, Nagpur Maharashtra, India*

ABSTRACT

The Web has now become a great source of information with more than millions of Webpages. This data may be in many arrangements like documents, images, videos as well as text too. But the web pages contain a vast amount of data. So with this vast data, it is common problem for everyone to get the exact and appropriate information that one wants. We usually do searching with the help of available platforms like Google, Yahoo, or many more platforms or automatically in the forms of web resources.

When the user will enter some keywords in the search engine using web Annotation, then the system will automatically call for the relevant web pages containing that keyword using the web annotation. In the web annotation, a user can have the right ability relevant edit, remove or add a piece of information from a web resource. Without the need of modifying the resource itself, the data will be modified. Web users can easily find the contents and documents through this technique. A web annotation is the type of annotation that will interrelate with the web pages. It creates user- friendliness. Annotation is associated with web pages, images, videos, etc.

Keywords: *Search Engine, Web Annotation, Web Pages, Keyword.*

1. Introduction

There are millions of pages on the web which are all set to present a diversity of information on interesting and amusing topics. Whenever you need them the Search Engines acts as the messengers of the same data at your ejection. The technical definition as quoted by Wikipedia: 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." If we put it in simple terms: Search engines are a web-based tool that enables users to find information on the World Wide Web. A Web annotation means an online annotation that interacts with a web page. In the case of a search engine using web annotation, when a user puts in some keywords in the search engine, the system will automatically fetch the web pages that contain the keyword using web annotation.

Web annotation helps to create user-friendliness. Web annotation is connected with web resources such as web pages or images. Without changing the resource itself with a Web annotation system, a user can add, modify or remove information from a Web resource. Web annotations are visible to the users. We here present a system where the user will enter search keywords of content or image. The system will match the keyword with annotation if the system finds matching content or image. The system will exhibit those content and image to the user. We here used an effective algorithm to generate query result page or search result records based on a user's query. Users can search for any details faster. The system is user-friendly, any user can use the system.

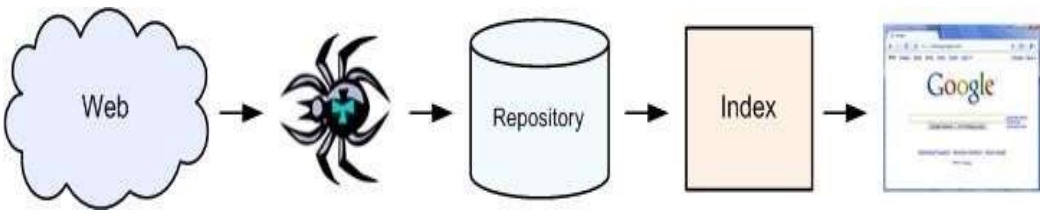

Figure 1 - The Web is crawled and placed into a local repository where it is indexed and retrieved when using a search engine.

2. LITERATURE SURVEY

W. Liu, X. Meng, and W. Meng.[2]

It is a big challenge for designing a system for extracting structured data from several web pages due to the complicated structures of such pages. A number of research and solutions are proposed for such problems, but each of them has certain restrains because these web pages are programming- language-dependent.

This approach primarily studies the content of the web pages and then according to its sufficient and useful content, it ranks the pages and shows the highest ranked web page with sufficient data as per the user's search keyword.

R. Khalil and N.A.K. Muhammad [3]

'The most tips for attaining better ranking in search end result through SEO' proposed a method that makes consumer seek records pretty proficient. This method offers affiliation between searches, files, and user queries .In the deliberate approach, results are higher than in previous approaches. The research paper describes the new set of rules for calculating web page rank in step with numerous parameters. The paper affords a modified web page rating algorithm. The new set of rules computes page rank based totally on inbound go to links on pages. This PR set of rules is called VOL which gives improved consequences from the original one.

R. Seema and G. Upasana [4]

' A evaluate Paper on web web page rating Algorithms' proposed content-based Hidden web rating set of rules (CHWRA). It includes four attributes. This technique tries to cowl all of the capabilities which affect the website fame directly or in a roundabout way. This method generates an ordered Hidden net searched end result set. The CHWRA set of rules gave the favored end result.

J. Ayush and D. Meenu [5]

"The Role of Backlinks in Search Engine Ranking" their research explains that Search Engines are designed to efficiently crawl and index web pages for better search results. There may be a huge contribution of hyperlink constructing to the popularity of a website. The quantity of inbound links performs a tremendous function within the website ranking at the device. Back links are hyperlinks which might be received through a website from another internet site. For a good SEO ranking, a link needs to be linked to a website that also has a high ranking. The backlinks produced must be relevant to the niche website. If the backlink on a keyword increases then the ranking of keywords on search engines will also increase

In recent years Web information abstraction and annotation is an active area. Due to some administered training and learning processes, systems can gain high extraction accuracy. The respondent of our survey were mainly Indians who claim that Google is the most sought-after search engine platform. Every search engine has its pros and cons. The survey can be of benefit both to designers of search engines to help them optimize their engines and to the users to decide which to go for which need. The paper summarizes that there is a lot of work done in the search engine field in past but there is still a lot of scope for research to happen.

3. EXISTING SYSTEM

Each search engines use different complex mathematical algorithms for generating seek outcomes. one-of-a-kind search engines like Google and yahoo understand distinctive elements of an internet web page consisting of page title, content material, meta description, after which give you their effects to rank on. Following are the exclusive current serfs.

ARCHIE: Archie did now not index the contents of the websites as the amount of information turned into so limited that it is able to be searched manually.

GOPHER: Searched most effective Titles and names of files stored of their index.
EXCITE: searched to enhance the relevancy of searches at the internet, by using the usage of statistical evaluation of phrase relationships.

YAHOO: At the beginning of Yahoo, the hunt become primarily based on the front-quit results that came from internet crawlers. Inside the 12 months 2003, Yahoo released its self-crawling seek Engine. Yahoo's largest contribution changed into the listing services, which created a large listing to find the quest consequences.

WEB CRAWLER: It turned into the first seek engine that got here up with a complete-textual content seek. This seek engine allowed its users to search for any word on any website.

INFOSEEK: Infoseek was used as a carrier that was paid, helped the website proprietors to post the internet page in correct time.

ALTAVISTA: It accredited humans to add or delete their key phrases within 24 hours. It have become the first searchable complete-text database on the web.

ASK: It become a natural language web search engine aiming to rank hyperlinks by their popularity.

Disadvantages:

1. In the sooner days of engines like google, the best of results turned into now not very constant with the question of the users. It changed into more of a directory, as a result the search consequences have been of a constrained fee.

2. There are hundreds of thousands and thousands of pages prepared to idiot you with faulty records formed via the whim and fancy of the writer. in this state of affairs, search engines like google and yahoo are the saviors, who save you these fraudulent websites/content material from accomplishing out to you.

4. PROPOSED SYSTEM

There are over 200 parameters that are used by Search Engines to determine the relevancy and popularity of the search results via the different search engines. This includes the title, keyword density, meta description, relevance of the content, backlinks profile, etc. The key features of the proposed system are as follows:

Accurate results: This application can provide accurate results as per the user's needs with great ease.

Describe: This application also helps in describing the information precisely and the user can get the correct information.

The query asked: The results will be given based on the exact query that has been asked by the users with great ease.

Organized way: The queries can be asked in a more organized way using this application.

Website owners: The website owners will be able to get various suggestions since the user will search using different keywords.

Easy access: Users can access this application anytime and anywhere in the world.

Saves time: This application can help in saving time for searching for the required information.

5. ALGORITHM

Crawling –

- To discover new documents over the internet search engines organize their crawlers.
- Crawlers' crawling frequency depends on the popularity of the website how frequently they found new content on your website. Less popular websites get a low frequency of crawling.

- But you can still submit your site map to Web Search Console. If you don't know what a sitemap is. A sitemap is a list of links to your all web pages. The search console will also let you know the other crawling errors on your website.

- The most common crawling errors are:
  Accessibility issues, Structural issues, too many redirections of **URL**, and Server issues.

Indexing –

- Search engines store the content they found in crawling in their huge data centres for later results retrieval. When you enter a keyword or phrase, the search engine looks at its pre-stored data and provides quick results to the users.
- If you want to block a few pages from the search results you can block crawlers by restricting them by uploading a **robots.txt** file with Disallow command.
- **Example: -** This is the **Robots.txt** of very popular news websites. They are restricting the crawler to enter a few directories.
- User-agent: *

  Disallow: /images/ndtvvideo/*

  Disallow: /news/redirect/*

Disallow: /news/feeds/*

- Create your own and save this text file to the root directory of your website.

Ranking–

- Search engines look to satisfy their users by providing quality and accurate results to their searchers. If results are not satisfying to their users they might opt for other search engines.
- Search engines use a ranking algorithm to assign a score to web pages and documents according to the user's query. So, more the score high in the ranking.
- The algorithm is the set of rules to determine the quality of web documents and web pages and provides the user with accurate information. These algorithms keep changing from time to time by search engines to improve the search results.

6. CONCLUSION

In this paper, we have proposed a framework for a personalized web search for creating an Enhanced User Profile using domain knowledge and browsing history. This Enhanced user profile is then implemented for suggesting relevant web pages to the user. Our work is significant as it improves the overall search efficiency, catering to the personal interest of the user. The biologist is searching for the "virus" that is a microorganism and the programmer is searching for the malicious software. For this type of query, several documents on distinct topics are returned by generic search engines. Hence it becomes tedious for the user to get the relevant content. Moreover, it is also time- consuming. Personalized web search is considered as a promising solution to handle these problems since different search results can be provided depending upon the choice and information needs of users. It exploits user information and searches context to learning in which sense a query refers.

7. REFERENCE

1. https://en.wikipedia.org/wiki/Search_engine
2. W.Liu, X.Meng, and W.Meng, ViDE: A Vision-Based Approach for Deep Web Data Extraction, IEEE Trans. Knowledge and Data Eng.,vol.22,no.3,pp.447-460,Mar.2010. https://www.ijert.org/survey-on-data-annotation-for-search-results-from-web-databases/
3. R. Khalil and N.AK. Muhammad, "The Foremost Guidelines for Achieving Higher Ranking in search Result through Search Engine Optimization",IJAST,vol.52,(2013).
4. R. Seema and G. Upasana, " A Review Paper on Web Page Ranking Algorithms", IJECS, vol. 3, Issue 8 (2014), pp, 7946-7949.

5. J. Ayush and D. Meenu, "The Role of Backlinks in Search Engine Ranking", IJARCSSE, vol. 3, Issue 4, (2013), online. https://www.researchgate.net/publication/344927616The_Application_of_Search_Engine_Optimization_in_Internet_Marketinghttps://nevonprojects.com/search-engine-using-web-annotation/
6. www.alexa.com
7. www.searchenginejournal.com
8. https://blog.expertrec.com/search-engine-using-web-annotation/#:~:text=A%20Web%20annotation%20is%20basically,the%20keywords%20using%20web%20annotation
9. https://projectideas.co.in/search-engine-using-web-annotation-dotnet/
10. https://developers.google.com/custom-search/docs/annotations
11. www.ukessays.com
12. https://studentprojectguide.com/otherprojects/search-engine-using-web-annotation/
13. https://www.studocu.com/in/document/sant-gadge-baba-amravati-university/project-report/search-engine-using-web-annotation/20588013
14. https://repository.aust.edu.ng/xmlui/bitstream/handle/123456789/402/Sam%20Manthalu.pdf?sequence=1&isAllowed=y
15. https://chrome.google.com/webstore/detail/annotate-web-annotations/gdojjgflncpbcfmenbkndfhoamlhajmf
16. https://partheniumprojects.com/search-engine-web-annotation/