

# Search Engine : Problems and Solutions

Mohit kushwaha

**Abstract:-** This report, presents the tools used in search engine and the working of a search engine. The search engine is used surf both content and picture search. The search engine comprises of various tools such as web crawler, parser, indexer, front end, backend and ranking algorithm. Search engines are important as they provide every detail related to different topics through billion of web pages. They are used to filter data on internet and then convert it into result that are relevant and easily accessible. The Original Page Rank calculation has been chosen and completely implemented. The Search engine will be furnished with custom design alternative( one can modify it regarding format ,add various surfaces and so many things).The valuable data, for example, passwords, email ids and personal informations are made secured by using appropriate protocols which is covering a decent degree of security and privacy.

## Software Specifications:

OS used : Windows.  
 IN Front End : javascript, php ,html  
 In Back End : MySQL

**Keywords:-** Search Engine, Algorithm, Tools, Web Pages, Data, Privacy, Security.

## I. INTRODUCTION

Web crawler innovation. An internet searcher regularly comprises of four parts for example search interface, crawler (otherwise called a bug or bot),indexer, and information base. The crawler navigates an archive assortment, deconstructs record text, and allots proxies for capacity in the internet searcher list.

Presently a days each client who utilizes Internet need to look for anything, as Educational schools, about Information Technology, books, news and so on, utilizing Web searchers. In todays world each and every individual is dependent on the internet for acquiring knowledge and content.

Internet searcher innovation was conceived nearly simultaneously as the WWW(World Wide Web) and has enhanced significantly over the previous years and has become a necessity of everyone's Web surfing experience, particularly after the incredible achievement of Google that was introduced in 1998.. At the Global level apparently Search Engines have been concentrated quite well, and numerous and enhanced. This report will include the issues that are present in search engines.

Each Search engine includes tools such as:

- Web crawler
- Parser
- Ranking system
- Frontend user interface
- Database

## II. SOME POPULAR SEARCH ENGINES WHICH ARE PRESENT NOWDAYS AND THEIR FEATURES

### a) Google

It is the most popular and acquires the 86% of search market share.

### b) Bing

Image search result and GUI is even better than google.

### c) Yandex

It is used by more than 45% of Russian Internet users (Extentions).

### d) DuckDuckGo

It doesn't collect or store any personal data.

### e) Baidu

It is controlling the 76.05 percent of China's market share.

## III. PROBLEMS IN NOWDAYS SEARCH ENGINES

- The most of the search engines which are present today they are mostly focused on text
- They does not produce better results about images, sounds and other searches.
- None of the search engine has multimedia searching modules, and none of them has produces better result in searching of images , sounds and video at one place.
- There is also a big consent about the privacy of the users and how they uses the data of the users.
- They are either specialized only in a particular area, or not adopting good ranking algorithms, or have not been maintained for quite a while.

There are some open source search engines the problem with them is that they are not as much popular and they are not being maintained and updated from time to time.

#### IV. TOOLS REQUIRED TO BUILD SEARCH ENGINE

##### A. Web Crawler

Web crawler is also known as spider or spiderbot. Crawler is a bot that browse the internet and gathers the data and information. It gets the data of all html pages as well as other all formats of documents which may include word, pdf, powerpoint, excel, slides and also images, sounds, videos. It collects the data of the title ,description and also some important content which they contain.

##### B. Parser or syntax analyzer

A parser breakdown the components of document or data or query into the smaller sizes to form indexable token. In tokenization we can also use stemming which means we change the derived words into their basic form but there is problem in this that the result which will be produce may be less relevant. It is also helpful in finding and detecting the language. Basically it takes input and build a data structure.

##### C. Search Engine Ranking System

Page rank algorithm provides a stability to the search engine. Google first algorithm has been used.

##### D. Front-end Development

In front end HTML, javascript, css php will make it more customizable and attractive.

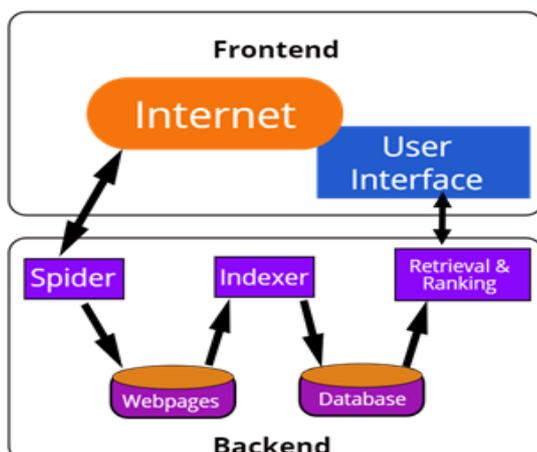
##### E. Back end development

As front end comprises of various tools backend technology will contains databases like mysql that will enhance the quality of internet surfing .

#### V. ARCHITECTURE DIAGRAM FOR PROPOSED METHOD

Adaptability and versatility are the two most significant plan objectives for search engine.

As we all know every search engine involves a structure or blueprint that helps in determining suitable aspects for search engine . in this report we present the basic design of our search engine that demonstrate the front end and backend.



#### VI. RESULT

Search engine is a software that uses tools such as crawlers, indexer, parser, html, javascript, css, mysql tools for searching.

After detailed analysis we came to the conclusion that None of the search engine has multimedia searching modules, and none of them has produces better result in searching of images , sounds and video at one place.

So we need to build a search engine which has most of the required features and overcome the problems mentioned above.

Each country must have their own private search engines because it will produce the better local results and also solves the issue of privacy and data security of country.It will also break the monopoly of one company .

#### REFERENCES

- [1]. Wikipedia
- [2]. [https://en.wikipedia.org/wiki/Search\\_engine](https://en.wikipedia.org/wiki/Search_engine)
- [3]. Github
- [4]. News