Search Engine and Portal for B.E. Projects

Gauri Agarwal
Department of Information Technology
Shah and Anchor Kutchhi Engineering College

Harshili Desai
Department of Information Technology
Shah and Anchor Kutchhi Engineering College

Aparna Mane
Department of Information Technology
Shah and Anchor Kutchhi Engineering College

Shikha Mishra
Department of Information Technology
Shah and Anchor Kutchhi Engineering College

ABSTRACT

The website will contain 2 sections, a search engine that will solve the problems of student searching for different projects and a portal that will be used to maintain all project related activities. The search engine will basically have information of different final year projects, along with its-features, working, application, screenshots of working projects and black book.

It will have 3 major Users: Students who want to provide the details. Students who want to fetch the relevant information. Users who want to verify whether the project is done previously or not.

It will provide a query box, for users to enter the project information to be searched. Its working will be somewhat similar to browser's search engine optimization i.e. the results obtained after searching will be content based which will help users to obtain and compare the features of existing project rather than name of project. The portal will be used to keep a track of various activities pertaining to final year project. It will be used by the project co-ordinator, guides and students.

Keywords
Meta semantic search engine, semantic search, link structure, page ranking, damping value.

1. INTRODUCTION

A **web search engine** is a software system that is designed to search for information on the World Wide Web. Web search engines get their information by web crawling from site to site. Web search engines are the central starting point for their exploration of Web content. Search engines lead us to new websites we have never heard of, help us reencounter familiar websites and offer us a wide variety of content from the many sources of the Web, which we would not be able to discover with other tools. Most users use search engines every day, and the amount of queries entered into general-purpose Web search engines such as Google worldwide exceeds 100 billion queries per month.

Domain specific search engines are becoming increasingly popular because they provide increased accuracy and which are not possible in general. The search engine to be developed will focus only on different categories of project developed along with its basic information, which will make the search more efficient and search results more precise. Because a user can specify a search domain (e.g. image processing or web programming), a search query may bring up Web pages both within and from the desired domain. Since provide to domain in which the information has to searched it becomes quite to look in to that domain for retrieving information.

The search results obtained will be then provided to **Page Ranking Algorithm** to display the contents according to most significant project, for the user entered query to least. Since the displayed obtained under the user specified domain and arranged as per the document relevance, the probability to satisfy user command increases, which is one of the major objective of developing a search engine. Page ranking algorithm will be 50% based on user's feedback and 50% by calculating the page rank value of each page. Apart from page ranking algorithm search engine optimization techniques will also be applied. [9]

A **web crawler** (also known as a web spider or web robot) is a program or automated script which browses the World Wide Web in a methodical, automated manner. This process is called Web crawling or spidering. Many legitimate sites, in particular search engines, use spidering as a means of providing up-to-date data. Web crawlers are mainly used to create a copy of all the visited pages for later processing by a search engine that will index the downloaded pages to provide fast searches.

2. RELATED WORKS
2.1 Meta Semantics Search engine ranking algorithm:-

Existing search engines are vulnerable to provide relevant answers to users query due to their dependency on simple data available in web pages. On other hand, semantic search engines provide efficient and relevant results. A Meta-Search engine is a search tool that forwards users query to several existing search engines and provides combined results by using their own page ranking algorithm.

Seman Telli is a meta-semantic search engine that fetches results from different semantic search engines such as Hakia, DuckDuckGo, SenseBot.

It sends the user query to these 3 search engines. These engines display results according to their page ranking algorithm and these results are provided to the Seman Telli search engine which then uses its own page ranking algorithm to provide results to the user. [2]
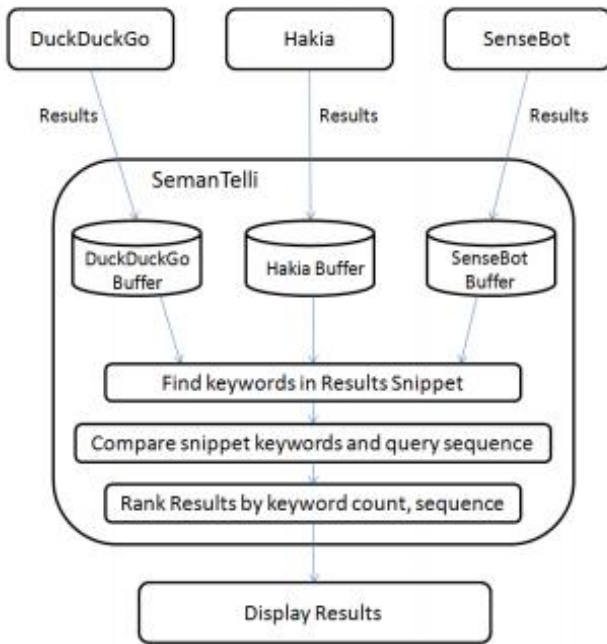
Fig. 5. Implementation details

Algorithm:
Input: Obtained results
Output: Ranked results

Step1:
Apply Conflation algorithm on query.
a) Removal of stop words (most common occurring words).
b) Suffix Stripping (remove the suffixes of words and retrieve stems (keyword). E.g. moving ->move).
c) Retrieve the keywords.

Step2:
Apply similar conflation algorithm as in step 1 to all obtained result snippets.

Step3:
Count the number of keywords from snippet and query. Consider, NKs = no of keywords in snippet. NKQ = no of keywords in query.

Step4:
For each keyword KS in snippet count (KS) = 0 for each keyword KQ in query If KS and KQ are same count (KS) ++

Step5:
Arrange results on the basis of decreasing order of keyword count.

Step6:
If keyword count (count (KS)) of more than one snippets is identical, then a) Check the sequence of keywords in snippet with sequence in query. b) Specific snippet having similar sequence as in query will get more priority.

Step7:
If sequence of keywords among more than one snippet is identical, then arrange the selected results considering their ranking from respective search engine and considered initial weight (W_i) .

Step8:
Return the ranked results. [2]

2.2 IP address Tracking algorithm:-
Normally search engines provide results based on relevancy of searched query and the contents of web page. The proposed ranking algorithm provides results by tracking IP address.

Page rank is an algorithm in which a numerical weight is assign to a web page. It uses incoming link information to assign a rank to all pages on the web. Number of incoming links from quality sites improve the ranking of websites. It depend on quality and quantity of links .pages with higher rank are shown on the top in the search engine result list. [3]

Algorithm
Step 1:
Start
Step 2:
Initially the ranking of all websites are set to 0.
Step 3:
When User click on any of the website its IP Address, Date of visit, time and No. of clicks are stored in the database and a particular rank is given to the website.
Step 4:
If user click on the same website again within 24 hours of time than algorithm check again the IP address, time, Date and number of clicks which are already stored in the database. If the IP address of user match than it increase only no. of clicks in the database and do not increase ranking of the website. If user click on the same website after 24 hours of time than its ranking increases.
Step 5:
If user click on different website, all process of ranking algorithm repeat and a different rank is given to the website.
Step 6:
The process repeated and according to its rank position of websites is given in the search result list.
Step 7:
Stop [3]

2.3 Feedback Based Ranking Algorithm:-
Page ranking either focus on link-structure of the Web or similarity of a page with the query to calculate Page Rank values of the pages. Feedback based ranking algorithm takes into consideration the feedback and user interests while calculating the rank.

Ranking algorithm being proposed, combines page similarity, link-structure information with user preference based on Internet domains and implicit user feedback (based on number of clicks and time spent on web page) for page ranking.[4]
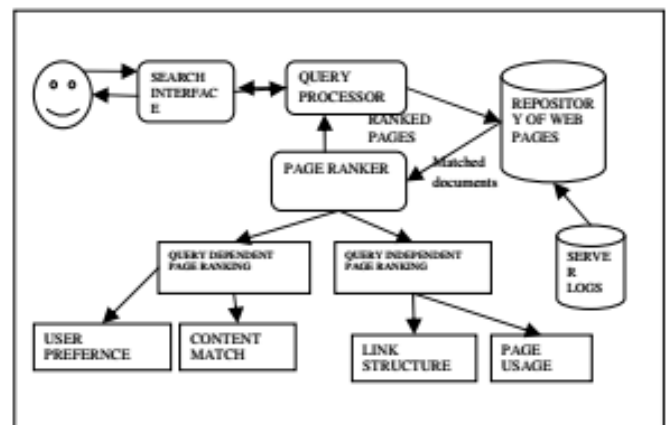
Algorithm
Step 1:
The user provides the search query.
Step 2:
These details are processed by the query processor.
Step 3:
All the web pages that match with the user query are then forwarded from the Repository to the Page ranking module.
Step 4:
It then assigns page rank based on query dependent (content and domain profile) and query independent (number of links and past usage pattern) factors.
Step 5:
The ranked pages are then returned to the user by the query processor. [4]

3. PROPOSED WORK

The Page Ranking algorithm that will be used in our project is based on the basics used in the above mentioned algorithms i.e the user feedback and the time a user spends on a particular web page.

The proposed algorithm is based on keyword search, feedback and the time user spends on a particular web page.

ALGORITHM:

Step 1:
Calculating the time user spends on the web page:-
i] Initially time of all the pages will be set to 0.
ii] After that, whenever a user opens a web page the timer will start.
iii] Timer will be updated by adding the recently fetched timing with the previously saved value of the time.
iv] The average of the time will be taken by dividing the calculated value by 2.

Step 2:
User Feedback:-
i] When a user opens a link and the user is on that page for some predefined period of time, the user needs to rate the document as to how relevant it is to his/her search query.
ii] The rating will be updated by adding the previously stored value of rate + recent rating provided by the user.
iii] This calculated value is then divided by 2 to take the average and this is taken as the final rate for that particular page.

Step 3:
Calculating the Page Rank:-
The page rank will be calculated by adding the time and the rate for a particular web page.
i] Get sum of all the values stored under   column in database. Store it as total_time of every project .

ii] Divide total_time into 5 equal parts.

iii] part1= total_time/5;
    part2 = part1*2;
    part3= part2*3;
    part4= part3*4;
    part5= part4*5;

iv] Now, assign values for each part and calculate the sum of time value and rate.
if(time stored for any project)<=part1
    tr_sum = rating of that paroject +1;

else if(time stored for anyproject)<=part2
    tr_sum = rating of that project +2;
else if(time stored for anyproject)<=part3
    tr_sum = rating of that project +3;
else if(time stored for anyproject)<=part4
    tr_sum = rating of that project +4;
else (time stored for anyproject)<=part5
    tr_sum = rating of that project +5;

v] Calculate average of time and rating and store it in database.

vi] Rank the project according to this average value.

III. System Overview:
The system to be developed will provide a common place to solve all project related issues, by allowing them choose a most efficient topic as they came through various similar project. The system will help students with activities like searching for project related information, finding the supervisor, scheduling meetings as per the timings that suit both supervisors as well as students, scheduling tasks so as to accomplish the short term goals for project.

It is basically a search engine for various project of which the results will be shown by using page ranking algorithm. Along with it, the system will act as a platform with the help of which students and faculty can regulate the project related activities.

It will have 4 users: -

The admin: They will simply maintain website and solve technical issues

The project coordinator: They generate the schedule of projects reviews and due date for project activities and maintaining it.

Guide: They will regulate project activities of their students and help them or can communicate with them

Students:

   Students will be divided into two sections:

   One who wants to upload the project (i.e. they have completed their project and they help others).They can only upload or edit their project details.

 One who wants to search for the projects available, regulate and co-ordinate with their guides.

The different modules available in the application are as Functionalities of various modules are explained below:

Registration::Project coordinator should register themselves to the website by providing proper college credentials (college code, college name, identity, department, etc.). When the registration is successful, project coordinator is given unique id. Project manager would send links to all the students and supervisors thus making his own circle. They should register using that link.

Upload projects::Any user will be able to upload his/her projects over the portal only when it is verified by the authorized member.

Make groups::Every user will be able to create his own circle to have proper communication with his team

Upload documents::Each user will be able to send documents to other registered members of the portal.

Schedule meetings::Users will be able to schedule their meetings as per the convenience of all members of the group

Schedule tasks::Users can schedule their tasks so as to set the deadlines for various activities and get them successfully completed on time

Post queries::Users can post any queries that will be answered by the experts of the domain.

4. CONCLUSION

The final output of this project is a website with multiple users, the website has two sections first is search engine and another is portal. The main objective is to solve the problem of students searching for final year project and a portal to interact. The website will be hosted on World Wide Web, through which user can access the data or projects from anywhere however Internet is a major requirement.

5. ACKNOWLEDGEMENT

We would like to express our gratitude towards our project guide Mrs. Swati Deshpande and Co-guide Mrs. Chintal Maisheri for their kind help and valuable guidance for the selection of this topic and the contents of this report. We also thank them for initial support and counselling for the project work.

6. REFERENCES

[1] IEEE Xplore Digital Library http://ieeexplore.ieee.org/Xplore/home.jsp

[2] Experience of developing a Meta-Semantic Search Engine. IEEE- 2013

[3] Development of a Ranking Algorithm for Search Engine optimization. IJERT -2014.

[4] A Novel User Preference and Feedback based Page Raking Technique. IEEE -2015.

[5] Improving Web Search User Query Relevance using Content Based Page Rank. IEEE-2015

[6] Page Ranking Algorithms used in Web Mining. IEEE-2014.

[7] Programming Collective Intelligence –by Toby Segaran

[8] A Survey on Domain Specific Search Engine - IRD India http://www.irdindia.in/journal_ijacte/pdf/vol2_iss2/5.pdf.