

“Follower”-Tracking Events Of A Person

Aashish Gurmukhani¹, Varsha Keswani², Manisha Parwani³,
Pooja Premchandani⁴, Mrs. Anjali Yeole⁵.

Department of Computer Engineering VESIT, University of Mumbai.

Abstract—Nowadays, To Track, only the events conducted by specific person are crucial issues in today's world of information overflow, with the main goal to provide only relevant information about the events such that users quickly get the correct information when they need it, which indeed takes the help of web crawler. A crawler is an agent that targets a particular topic and visits and gathers only relevant web pages. The web pages are fetched and parsed for retrieving all the hyperlinks by the crawler service, and then the same process is continued recursively using pattern matching algorithm. The main objective of our project "Follower"-Tracking events of a person to follow any popular person if he is conducting a lecture at some location that news (text, pdf, word document ,etc)is immediately visible in that location's website . This same should also be intimated in our website just like the notification.

Keywords—Web Crawling, Searching, Seed, Fetch, Parse

I. Introduction

As the amount of web information grows rapidly an efficient personalization approach that modifies the appearance of a website's content[3] to satisfy a specific user's instructions or preferences is required. It is also essential to keep track of the change of interest of the user from time to time. The search engines[2] currently popular in the market fail to achieve the required balance between personalized search results and user's privacy. Nowadays, To Track, only the events conducted by specific person are crucial issues in today's world of information overflow[4], with the main goal to provide only relevant information about the events such that users quickly get the correct information when they need it. "Follower"-Tracking events of a person plays an important role by providing people with the immediate information about the conducted event's or upcoming event's, when they are really searching for specific person i.e Follower. In case Follower's any event is uploaded in any website, that is to be intimated in our website as notification. This research proposes a use of a program called 'web

crawler' for the retrieval of information from different websites.

II. Literature Survey

The algorithms that are used in the follower tracking project[1]:

1. KMP(Knutt-Morris-Pratt)

In Knutt-Morris-Pratt algorithm Pattern and text are compared in a left to right scan. If a match occurs, the algorithm Searches for the largest suffix of the "first start" this is also a prefix of the pattern and thereby determines how far the pattern can be shifted to the right without missing a possible match. The data we need to find the next shifting position is stored in an auxiliary "next" table which is computed in a preprocessing step by comparing the pattern with itself, the entries contain the information which character of the pattern should be compared to the text character in the mismatch position next. This "next" table is superior to the helper function. Here is a short description of the algorithm we use to compute the next table: We slide a template of the pattern long M and search for the largest prefix of M which is a suffix of M[1...i]. we compute the pattern with itself at each possible shift position, while the characters match, both the pointers are incremented, when a match happens after more then one comparison, we compute next[i] of i-1, if the match occurs already in the first position, next[i] is set to 0, I is incremented, and we check the next shift of the pattern itself.

Input : string m(input word) with m characters and target web file
Output : the number of comparison the algorithm did for matching and the time taken by the algorithm to find the match.

The main algorithm[1] shows nearly the same structure:

```
While(I<n){
  if(pattern.charAt(j)==text.charAt(i)){
    if(j==m-1)
```

```

return I-m+1; //match
j++;
}
else if(j>0){
j=fail[j-1];
} else{
i++;
}
return -1; // no match

}
}

```

III. Proposed Work

Our project proposes the implementation of a smart search engine which generates search results relevant to the follower who will be following any person. The search engine would crawl the web, analyze and rank the pages on the basis of various factors like the frequency and location of keywords within the web page, the duration of existence of the web page, its popularity, the number of out links to the current web page in question.

The most important feature that would set our search engine apart would be the privilege of opting in for an enhanced search experience, influenced by the interests and likings of the follower’s peers on the social media network or any relatives etc based upon their search roles. Search role will be specific based upon the person to whom follower is following. Furthermore, the follower will be notified on and again about various events conducted as well upcoming events. This feature would guarantee a secure and non-intrusive search experience.

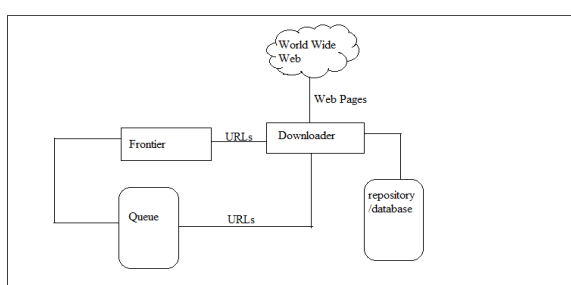


Figure 1: Block Diagram

It Uses Web Crawler to connect to the Website and download data as per the input keyword given by the follower to the search engine, The Software is designed in such a way such that the user can easily interact with the screen through GUI Hence in accordance with the keyword entered by the follower results will be shown using KMP algorithm.

IV. Technologies Used

- 1. Java:** Java is used in a variety of computing platforms from embedded devices and mobile phones on the low end, to enterprise servers and supercomputers on the high end. While less common, Java applets are sometimes used to provide improved and secure functions while browsing the World Wide Web on desktop computers.
- 2. MySQL:** The recorded data is to be stored in a database. This is needed because it is not necessary that there will always be a network available to forward the biomedical data to the server.
- 3. HTML:** HTML describes the structure of a website semantically along with cues for presentation, making it a markup language, rather than a programming language. HTML elements form the building blocks of all websites. HTML allows images and objects to be embedded and can be used to create interactive forms. It provides a means to create structured documents by denoting structural semantics for text such as headings, paragraphs, lists, links, quotes and other items.
- 4. Eclipse** is an integrated development environment (IDE). It contains a base workspace and an extensible plug-in system for customizing the environment. Eclipse is written mostly in Java and its primary use is for developing Java applications, but it may also be used to develop applications in other programming languages through the use of plug-ins.
- 5. Windows Xp:** Microsoft Windows (or simply Windows) is a meta family of graphical operating systems developed, marketed, and sold by Microsoft. It consists of several families of operating systems, each of which cater to a certain sector of the computing industry.

V. Conclusion

The proposed follower tracking events of a person project can connect to the websites ,and download data. The designed system using web crawler is capable of comparing the text found on a link with the input text name of a person, it can be in any pdf or word document or simple text message. The crawler uses pattern recognition and generates the number of times the input text exists in the text found on a link. The information so generated gives an efficient schedule of the events of person using KMP algorithm .

There are two screens User and Admin. Admin can add the links, words and documents .For tracking the person or getting the information, the user has to login through page and then can proceed for the search. search will lbe according to the role viz.

general person search, professors and relatives. then it will be fetched from the admin database. And the results will be given to the user according to the search.

VI. References

[1].Allen Heydon and Mark Najork, “Mercator: A Scalable,Extensible Web Crawler”, Compaq Systems Research Center, 130Lytton Ave, Palo Alto, CA 94301, 2001.

[2]. A Vadivel, Shaila S. G, R. Devi Mahalakshmi and J. Karthika. “Component based Effective Web Crawler and Indexer using Web Services” IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM -2012) March 30, 31, IEEE 2012.

[3]. Weigang Guo, Yong Zhong, Jianqin Xie “A Web Crawler Detection Algorithm Based on Web Page MemberList” 2012 4th International Conference on Intelligent Human Machine Systems and Cybernetics IEEE 2012 URL is irrelevant and place it in Irrelevant links Log.

[4]. StatMarket. Search engine referrals nearly double
<http://websidestory.com/pressroom/pressreleases.html?id=181,gets empty,2003>.

