Recommendation on the Web Search by Using Co-Occurrence

S.Jayabalaji[1], G.Thilagavathy[2], P.Kubendiran[3], V.D.Srihari[4].

UG Scholar, Department of Computer science & Engineering, Sree Shakthi Engineering College, Coimbatore, Tamilnadu, India[1, 2, 3.]

Assistant Professor, Department of Computer science & Engineering, Sree Shakthi Engineering College, Coimbatore, Tamilnadu, India[4]

ABSTRACT: In our day to day, the usage of internet and searching the information should be increases rapidly. Because of this, now a days we have facing the problems like whether the retrieving information would be noise free or not and having many confusions with the usage of keywords to get the exact result. To avoid this problem we are going to propose the concepts called Co-Occurrence and recommendation. These two concepts increases the effectiveness and of the result. By using the recommendation concept we have multiple choices to select the desired thing. The web search increases dramatically [1] user search performance leads to large number of confusions. We examine a general expert search problem: searching experts on the web, where millions of web pages and thousands of names are considered. The two main issues are: Web pages might be of untrustworthy and have more noise; the knowledge evidences spotted in web pages are frequently unclear and ambiguous. The skilled search has been studied in different contexts, e.g., enterprises, academic communities. We propose to influence the huge quantity of co-occurrence information to calculate the significance and status of a person name for a query which is given. So this makes the recommendation system the most important and the trust worthiness of the system will be analyzed in the better way. The personalization will be depended based on the individual user process in the web search mainly worked in E-Commerce application.

Keywords: Co-occurrence, recommendation, diffusion, personalization.

I. INTRODUCTION

Mining the useful data in the web is known as web mining. In web mining, there is 3 partitions they are web content mining, web usage mining, web structure mining, we have chosen web usage mining. It is a data mining technique used to retrieve the web patterns. Recommendation means giving some suggestions about the related item what we are searching and it reduces the accessing time. It is used in the E-Commerce. E-Commerce is an industry where the people can buy and sell the products or services through online. Recommendation supports E-Commerce only in three types. They are Business-Business (B-B), Commercial- Commercial(C-C), Business-Commercial (B-C). Co-Occurrence is the collection of information from the multiple databases. It retrieves the similar information from the different databases. Personalization means the uniqueness finding the uniqueness of the product or item.

The recommendation process is used to reduce the accessing time. When the user gives the keyword in search engine, it searches for information in the data warehouse the keywords get compare in the ontology. If there is any relative information in ontology generator the recommendation retrieve the relative information from ontology generator and it will displays in the web page. In the existing system they told about what are the problems we are facing now? The users having enormous number of web pages and plenty of keywords to search but we don't whether the data can be error free or not. The fig1 shows the process of recommendation [2],
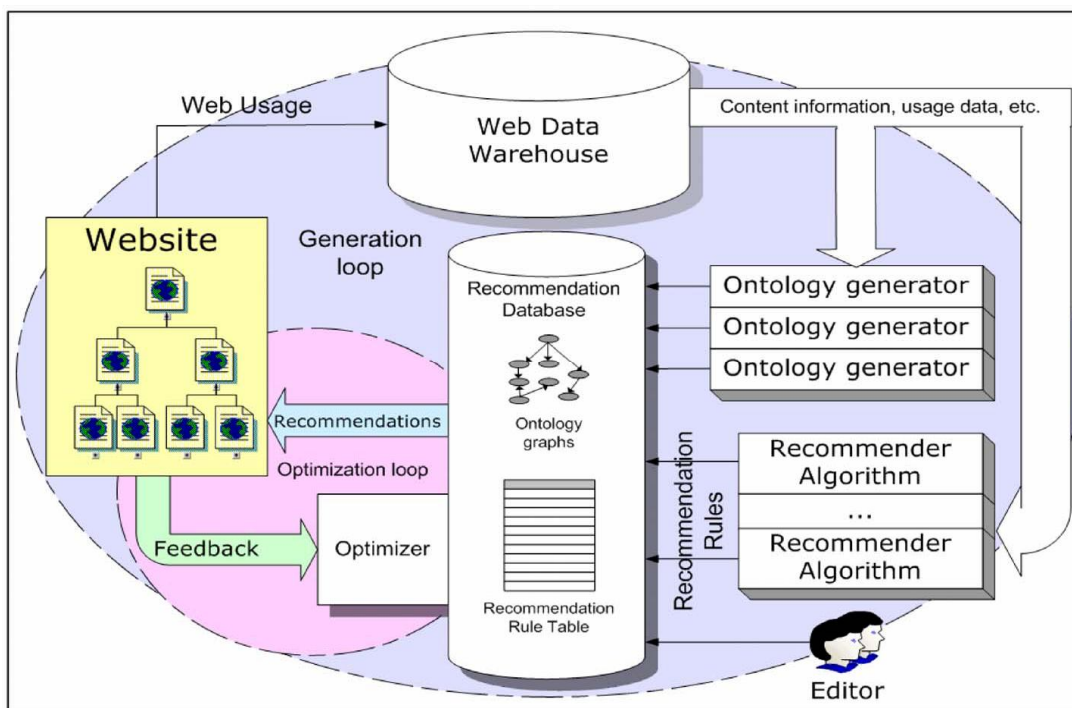
Fig 1 Recommendation Process

If the user want to access a particular information from the web page already there will be an keyword to access the particular page, with the help of the keyword alone we can able to access the particular information but the user gave the own keyword it take lots of time to access the information from the database, when the keyword is get increased the searching time also get increased because the database have huge information.

II. LITERATURE SURVEY

A. *Amazon.com Recommendations Item-to-Item Collaborative filtering*

Amazon is the company who implemented the recommendation concept in the websites. Recommendation algorithms are the best for e-commerce websites [5]. According to the customer interest the algorithm generates the recommended items. In Amazon.com, to personalize the online store we are using a recommendation algorithm for each customer and the storing data changes based on the interest of the customers like displaying the programming titles to a software engineer and baby toys for new mother.

Recommendation algorithms based on e-commerce often operate in a challenging environment. Huge amount of data can be handled by the large retailer, millions of customers and millions of separate catalog items. Customer have limited information because of the few product are purchased by the user. When compared to the new customer old customer have a huge amount of information based on the product purchased. There are three approaches to solve recommendation problems Traditional collaborative filtering, Cluster models, Search-based methods.

- Traditional collaborative filtering – Collecting the similar information from the different database and filtering the useful information according to the user needs.
- Cluster model - The algorithm's goal is to assign the user to the segment containing the   most similar customers. It then uses the purchases and ratings of the customers in the segment to generate recommendations.
- Search-based method – This is the method that treat the recommendations problem as a search for related items.8 Given the user's purchased and rated items, the algorithm constructs a search query to find other popular items by the same author, artist, or director, or with similar keywords or subjects.

In this paper we using the algorithm called Item-item collaborative filtering. The main use of this algorithm is grouping the similar items purchased by the customers. Here we are grouping only the similar items not the similar customers.

B.   *Mining Navigation History for Recommendation*

The history contains a lot of hidden information the history have only a link name in that we cannot see the full details about the information which is accessed by the users. The web page and the user alone know what data is accessed by the users [4]. The information which is hidden is invaluable source it is not necessary to the users but with the help of the history we can know which page is highly accessed by the users from that we can suggest that page. The rating of the pages is given based on the user interests.

There are 2 types of hidden information is discovered from the history.

1. The keyword which is given in the url there must be an some relationship between a set of url, it may similar to each other.

2. If the two users have read similar information we can say that the users have similar interest.

Surflen is an information recommendation system and suggest web pages to the user.

Collaborative filtering is require for the recommendation process. With the help of the recommendation association rule we can find the dependency among the items. The efficiency of the recommendation can be improved by using the recommendation process. A system boot strapping is a common problem in collaborative information recommendation system. If two users have a similar items by these we can say that they have a nearest neighbor.

C.   *A personalized recommender system based on web usage mining and decision tree induction*

Day by day the users searching items and information in the database is getting increasing there may be a chance for the overloaded of information in the database to overcome these overloaded information problems decision tree induction is used [3]. Collaborating filtering is only applicable for the large application like E-commerce it is not applicable for smaller application.

The decision tree induction is like a tree structure. In this the data are classified according to the user's interest and priority of the data give to the web page .the tree contain all the information the data stored in the tree form. The problems identified in this process are

(1) Whether customers for whom we want recommendations (they are called target customers in this paper) are all customers or selective customers.

(2) Whether the objective of recommendation is to predict how much a particular customer will like a particular product (prediction problem), or to identify a list of products that will be of interest to a given customer (top- N recommendation problem).

(3) Whether the recommendation is accomplished at a specific time or persistently.

We suggest three different strategies related with such a choice:

 (1) Recommendation of the most frequently purchased product.

(2) Recommendation of product with the highest click-to-buy rate.

(3) Recommendation of the latest product.

With the help of the click stream we can learn customer preference and product association. To avoid the poor recommendation, the decision tree induction is used.

III. EXISTING SYSTEM

The previous work on expert search is often restricted within specific contexts, e.g., an enterprise quantity, an online forum, or an academic bibliography collection. Recently, the desire to find experts on a variety of daily life topics is increasing. We are observing an increasing search standard that allows users to search for people who can answer their natural language questions. However, this system requires users to register and join a community. In contrast, the web contains a huge quantity of information about people (e.g., personal home pages, blogs, and web news). Therefore, it is possible to build a powerful expert search engine by exploiting the information about people on the web.

IV. PROPOSED SYSTEM

In proposed system we have to implement the concepts like Co-Occurrence, Query keyword, Ranking Algorithm. Co-Occurrence is the process of collecting the scattered information from different database with trustworthiness. Query keyword – with the help of the keyword alone we can able to access the information. If A, B, C are the users searching for particular information in the web page [6]. The user A and B using a different keyword to access the same web page, the user C uses an one of the keywords which is used by A and B it will recommend those pages.

Ranking Algorithm is a process of giving a priority to the web pages in the database. The information which is searched by the user will stored in the database and it will give the priority according to the user interest.

One time ranking algorithm – In the case of one time ranking algorithm, once we fixed the priority for the web page, we can't able to change the priority of the web page.

Iterative re-ranking algorithm – In this algorithm the priority of the web page changes automatically without any manual work. The priority will change according to the user's interest.

Personalization – This process will display the ideal information which is required to the user. It filters the irrelevant information from the database and displays the relevant information to the users.
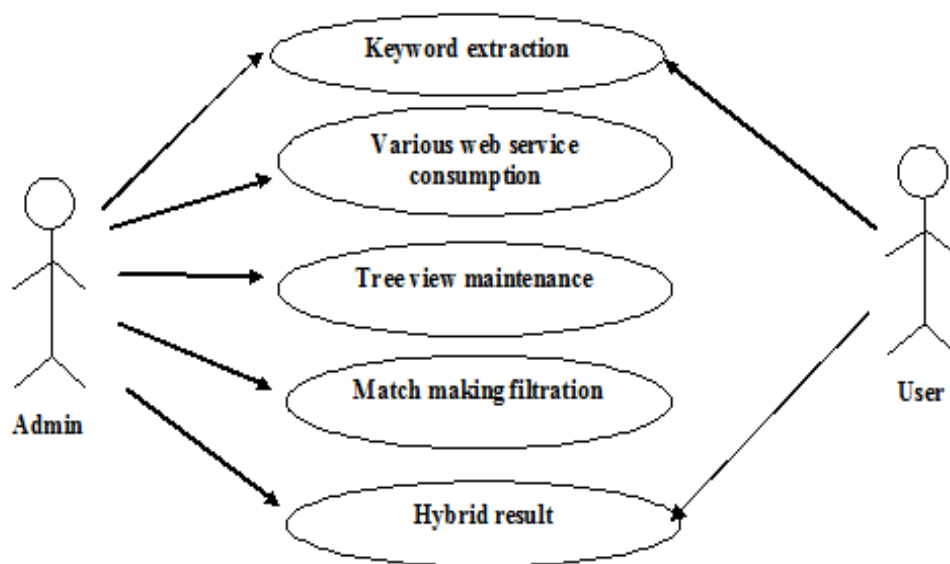
Fig 2 Proposed Method

The Fig 2 represents the overall diagram for the proposed model and the formation of the resulted in the E-commerce application.

V. EXPRIEMENTAL RESULT

The experimental result shows the process in which the web pages for the E-Commerce application is done based upon the ranking concept and the personalization is processed in case of the Match making algorithm this gets processed in many application based upon the retrieval of the information in the recommendation. Initially the process gets started with the user input, then system will analyse the user input based on the match making algorithm. This will be depend upon the web pages which all having the similar result, for example the result for the query can be examine in different websites by means of various search engines. After the process over it will process the better result by means that the system is recommending the user to choose the best result for the given query.

VI. CONCLUSION

Thus the model created for the recommendation process where the user query is established in the various search engine are checked by combining them using algorithms and find out the personalization and trustworthiness of those are also processed. In future, the process going to performed is that, initially need to find the meaning of the particular word by using ontology and then search related to that result. Later, it stores the document which gets download in the local drive. If the search takes more than one word means, it separates the word and makes the searches separately than finally gives the search result. This project acts as a web browser not a website. So in future we are going to develop the project as a website. In generally it is going to access the website less than four. In future we are going to develop the project as to access maximum number of websites.

International Journal of Innovative Research in Computer and Communication Engineering

REFERENCES

1. Ziyu Guan, Gengxin Miao, Russell McLoughlin, Xifeng Yan, Member, IEEE, and Deng Cai, Member," Co-Occurrence-Based Diffusion for Expert Search on the Web" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 5, YEAR 2013.
2. Nick Golovin and Erhard Rahm," Automatic Optimization of Web Recommendations Using Feedback and Ontology Graphs", University of Leipzig, Augustusplatz 10-11, 04109, Leipzig, Germany.
3. Yoon Ho Choa,, Jae Kyeong Kimb, Soung Hie Kima, "A personalized recommender system based on web usage mining and decision tree induction" a Graduate School of Management, Korea Advanced Institute of Science and Technology, 207-43 Cheongryangri-Dong, Dongdaemun,Seoul 130-012, South Korea.
4. Xiaobin Fu, Jay Budzik, Kristian J. Hammond ," Mining Navigation History for Recommendation" Infolab, Northwestern University,1890 Maple Avenue, Evanston, IL 602011(847) 467-1265 fu.
5. "Amazon.com Recommendations Item-to-Item Collaborative Filtering", Greg Linden, Brent Smith, and Jeremy York • Amazon.com Published by the IEEE Computer Society 1089-7801/03/$17.00©2003 IEEE IEEE INTERNET COMPUTING.
6. "A Literature Survey on Semantic Annotation of Ubiquitous Learning Environment", R.Shanmugapriya, Vol. 1, Issue 9, November 2013. ISSN(Online): 2320-9801,ISSN (Print): 2320-9798. Page No :2006-2009.
7. R. Bekkerman and A. McCallum, "Disambiguating web Appearances of People in a Social Network," Proc. Int'l Conf. World Wide Web (WWW), pp. 463-470, 2005.
8. M. Belkin and P. Niyogi, "Laplacian Eigenmaps for Dimensionality Reduction and Data Representation," Neural Computation, vol. 15, no. 6, pp. 1373-1396, 2003.
9. H. Deng, I. King, and M.R. Lyu, "Enhancing Expertise Retrieval using Community-Aware Strategies," Proc. ACM Conf. Information and Knowledge Management (CIKM), pp. 1733-1736, 2009.
10. H. Deng, I. King, and M.R. Lyu, "Formal Models for Expert Finding on DBLP Bibliography Data," Proc. IEEE Int'l Conf. Data Mining (ICDM), pp. 163-172, 2009.
11.

BIOGRAPHY

S.Jayabalaji is an Under Graduate Student in the Dept of Computer Science and Engineering, in Sree Sakthi Engineering College, Coimbatore, under Anna University. Area of Interest is Hardware, Open Source and Web Mining.

G.Thilagavathy is an Under Graduate Student in the Dept of Computer Science and Engineering, in Sree Sakthi Engineering College, Coimbatore, under Anna University. Area of Interest is Web mining, Scripting Languages.

P.Kubendiran is an Under Graduate Student in the Dept of Computer Science and Engineering, in Sree Sakthi Engineering College of, Coimbatore, under Anna University. Area of Interest is Web Mining.

Mr.V.D.Srihari completed his PG in the Dept of Computer Science and Engineering, and currently doing his research in the area of web mining, he is also working as the assistant professor and head of the department in Sree Sakthi Engineering College, Coimbatore, under Anna University. He has guided UG and PG student in his experience. His Area of Interest is Web Mining, Operating System, Open Source Tools, Data Mining, Hardware and data structures.