# Online content optimization with effective event monitoring approach

**Sajithra.N[1], Priyadarshini.D[2]**

M.Phil. Scholar, Department of Computer Science, Sree Narayana Guru College, Coimbatore, Tamil Nadu[1]

Assistant Professor, Department of Computer Science, Sree Narayana Guru College, Coimbatore, Tamil Nadu[2]

**Abstract:** Content optimization has widely used in personalized search engines for better personalized results. It can be generated by the user actions and events on the web search engine. User interaction on the page plays a vital role in recommender systems. Previous studies on recommender systems mainly focused on modeling techniques and feature development, this content optimization is based on general behavior analysis algorithm. It provides user action analysis is critical for a recommender system. The system proposes a novel implicit user feedback and event monitoring schemes for efficient content optimization. For this our system proposes PCO approach. But user interactions in real-world Web applications are unlikely to be as ideal as those assumed by previously proposed models. Our proposed system builds an online dynamic learning framework for personalized recommendation. The main contribution in this paper is an approach of personalizing users' searches to achieve better search result which is based on event monitoring and personalized content optimized search.

**Keywords:** Content optimization, web search engine, Event monitoring, PCO and DKMC.

## 1. INTRODUCTION

In real world all the internet users fully depending on the search engine, search engine is not a small, search engines having the lot of information's, all data's stored with semantic model. Data search and information gathering from the search engine through the internet has placed high demand among the Web Surfer (user). For this needs many popular web search engines like Google provides saving option of individual histories and provide a relevant search to the user. While searching in the web, the result has been generated using the frequently visited sites, i.e. based on the popularity, the search engine providing the result [1]. Many previous approach search or Search intelligence techniques are based on the concept proposed with frequent clicks of the web pages with effective algorithm and technique. In order to improve the frequency based search with additional parameter is the research work, which was failed in the existing. With the explosive growth of information sources available on the World Wide Web and the rapidly increasing speed of adoption to Internet commerce, then it has grow up into a gold mine that encloses or dynamically generates information that is beneficial to E-Searches. Web sites are the most direct link to the user to know the required information from their place [2]. These web sites are offered by the many companies like Google, Yahoo, Wikipedia; etc… these companies has to its current and potential users. The companies can study visitor's activities from the web analysis, and discover the models in the visitor's behavior. These models can be deferred by the web analysis and the result where from the grouping of the data warehouse where offered by the companies [3, 4]. In many popular web search engines like Google provides a saving option of individual histories in order to recover and retrieve frequently viewed pages. But the limitations and convenient based issues over the search engines are the main motivation of the proposed work. Ineffective search history analysis and parameters which determines the interest is so old methods. The goal of this paper is to find users access models robotically and rapidly from the enormous Web log data. Such as frequent access links, regular access page groups and user clustering [5]. In particular, this shows that having a web log analyzer alone is not sufficient to provide a complete picture of Search intelligence. However, web-log analysis provides very good detailed reports on web site traffic statistics. There is a need to mine other information such as competitors, market positions, and most significantly the users and their behaviors on the product and services. For this we aim to provide a Content optimization of personalized search engines for better personalized results. For that user actions and events on the web search engine are monitored. The process of making the personalized webpage based on the user interest and popular data is called *content optimization*. User interaction on the page plays a vital role in recommender systems [6]. Preceding studies on this recommender system have mainly focused on modeling techniques and feature development, this content optimization is based on general behavior analysis algorithm. This argues that suitable user action analysis is critical for a recommender system [7]. Our system proposes a novel user feedback and event monitoring schemes for effective content optimization technique. And proposes a new method named as PCO (Personalized Content Optimization). The main contribution in this paper is an approach of personalizing users' searches to achieve better search. To achieve better item relevance assessment, our proposed system exploits the following parameters. (1). Event monitoring and click behaviors from Web search, (2). Personalized content optimized search.

## 2. LITERATURE REVIEW

Application of data mining techniques to the World Wide Web, referred to as Web mining [8, 9], it has been the concentrate on the several recent research projects and papers to provide a better result. Nevertheless, there is no recognized terminology, leading to puzzlement when evaluate research efforts. This Web mining has been used in two different ways. (i) Web content mining, it is the process of information detection from sources from the World Wide Web. (ii) Web usage mining, this defines the process of mining for browsing of user and patterns which were access. This shows the web mining and provides an overview of the various research issues, techniques, and development efforts which were done. It describes WEBMINER, system for Web usage mining, and comprises the paper by categorizing various research issues.

Mining typical user profiles [10] and URL associations from the huge amount of access logs is an important component of Web personalization. In this paper this defines the notion of a user session as being a temporally compact sequence of Web accesses by a user. This also defines dissimilarity between two Web sessions that organization of a Web site. With this to group the user sessions based on the pair wise differences, this commences the relational fuzzy c-maximal density estimator (RFC-MDE) algorithm. This algorithm is robust and can pact with outliers that are typical in this application. This show real example of the use of RFC-MDE for extraction of user profiles from log data, and compare its performance to the standard non-Euclidean fuzzy c-means.

Large volumes of data are grouped automatically by Web servers and collected in access log files [11, 12]. With the analysis of server utilizing data can present significant and helpful information. The concept of Web Usage Mining is mainly used for pertain the DM techniques to the verdict of usage patterns from Web data and is embattled towards applications. It extracts the secondary data resulting from the relations of the users during definite period of Web sessions from the access [13]. In web usage mining it contains three segments; they are preprocessing, pattern finding, and pattern analysis. In this paper, they applied Kohonen's SOM (Self Organizing Map) to pre-processed Web logs of our university Web server logs (http://www.um.ac.ir/) and extract the regular patterns.

In paper [14] shows an approach for automatically classifying visitors of a web site based on their user access patterns like click and un-click events. User access patterns are observed by discovering clusters of users that display similar information requirements; e.g., access of user's similar pages. It brings the result in an improved understanding of how users spend the site. Hence this mostly depends on what category which an individual user falls into that site; this feature is done by dynamically recommending links for him to navigate.

Designing a web site is a complex problem [15, 16]. Logs of user accesses to a site provide an opportunity to observe users interacting with that site and make improvements to the site's structure and presentation. This proposes adaptive sites which is web sites that learning themselves from user access patterns or frequent visited pages. This adaptive web site can create popular pages more reachable or accessible to the user, and it emphasizes attractive links to the user, and also helps to connect associated pages and grouped similar documents mutually.

In [17, 18] of web users, known as "surfing", they follow links to obtain the knowledge or gather information. Regular navigation often directs to disorientation when users miss the track of the content in which they are navigate and are uncertain how to continue in from their original query, they can drop the point in that situation because of providing irrelevant data by the web sites provider. Analyzing the navigation patterns of web users is most significant to the web site provider. For this they derive Zipf's rank frequency law (i.e. an inverse power law) from an engrossing Markov sequence model of surfers' performance pretentious that less likely navigation track are, on standard, longer than more possible ones. In this model the likelihood of a track is interpreted as the relevance (or "value") of the trail.

In paper [19] analysis of contextual information in search engine query logs enhances the understanding of b users' search patterns. Obtaining contextual information on b search engine logs is a difficult task, since users submit few numbers of queries, and search multiple topics. Identification of topic changes within a search session is a significant branch of search engine and the user performance analysis. The topic identification algorithm's performance becomes uncertain in different cases. By this cases are discover and the reasons underlying the inconsistent performance of automatic topic identification are investigated with statistical analysis and experimental design techniques.

Query logs record the queries and the events of the users by their clicks from the visited sites [[20, 21], which enclose precious information from the interests, the favorite, and the performance of the users, as their implicit feedback to search engine results. Here they use *query-flow graph*, which is for a graph demonstration of the interesting information about latent querying behavior. The query-flow graph is a result of query-log mining and, same time, a valuable tool for it. This Method builds the graph by extracting time textual information and aggregating queries from the different users. Using this approach build a real-world query-flow graph from a large-scale query log and demonstrate its utility in concrete applications like *finding the rational sessions* and *query recommendation [22]*. And has a drawback of guiding query recommendation.

## 3. PROBLEM DEFINITION

The system deals with the problem of how to appropriately define user segments (i.e., divide users into different groups according to their interests) to achieve personalization becomes crucial for effective content optimization. This also deals the query clustering, user interest clustering organizing, result re arranging and page ranking oriented problems. The system has concentrated on the problem of web content extraction according to the

query and desired results of the user [23, 24]. The system should automatically maintain and organizes the user click through logs and clicked citations together for creating event collections. The followings are the main problem deals with the existing system.

- Dynamic and automatic Grouping of user history.
- Automatic identification of query clusters.
- Less performance of extraction results.
- Each web page has its own template and may different in the outline.

Session identification, query clustering and reconstruction of clusters are very tedious. If the link is a non uniform format the existing system suffered to store the histories effectively [25].

## 4. PROPOSED WORK

To the improved performance to the web content optimization we proposed an innovative approach that infers user interest and content optimization for a query by clustering its feedback sessions represented by similar clusters and implicit and explicit user feedbacks. Initially the system introduces feedback sessions which are categorized into two categories one is implicit and explicit feedbacks to be analyzed to infer user search goals rather than search consequences i.e. clicked URLs. With the use of both clicked and unclicked, ones before the last click are considered along with the time spent by the user in every click as user implicit feedbacks and taken into account to construct feedback sessions. Consequently the feedback sessions can suggest user information desires more efficiently. The subsequently method of the suggestion is mapping the feedback sessions to event collections to estimate the goal of user.
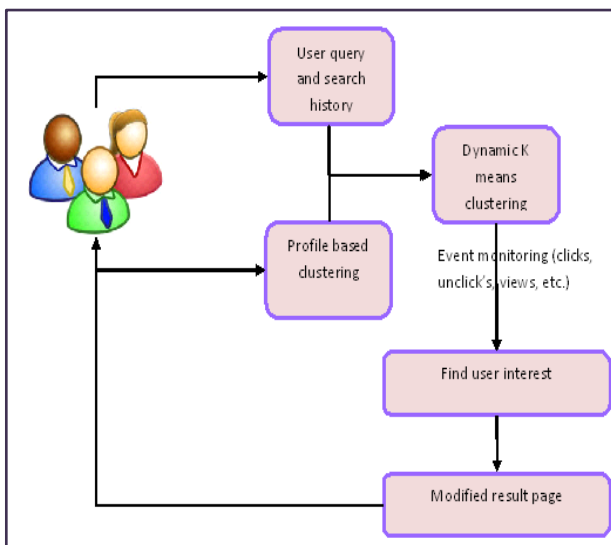
### PROPOSED ARCHITECTURE



Fig 1: Our Proposed Architecture

The above diagrams describe how to use the clicked URLs by the current user together with the given query q to better capture their search intent so the system identifies the set of URLs and clicks, that are clicked by the user after issuing q. After that, it utilize click and the self categorizing cluster to expand the space of queries

considered when it computes the content extraction and relevance ranking R.

The proposed system provides an effective way to store, organize and manipulate the user search histories with effective structure mining techniques. The semantic based search engine produces a query matched results than the priority or important content based. Now a day's data storage and references are very huge in size. Providing filtered data is more important. The proposed system provides an effective summarization and organization of user histories, which is implemented in the web search. Our proposed web search engine presents the main semantic relativity between the given words and terms, and it will generate the semantic measures automatically and it also performs the user history organizations effectively. This kind of summarization, organization and extraction techniques improves the efficiency of the user search over the internet.

This proposed method is an automatic technique to evaluate, and it is estimating the semantic likeness among the words or entities using web search engines, text scraps and a lexical pattern mining schemes. That considers word subsequences in text snippets. The system proposed a new algorithm which is named as "dynamic K means clusters" (DKMC) algorithm which combines more features of web mining concepts.

Our proposed system achieves the following aspects:

- Personal Content Optimization and semantic cluster helps to reduce the risk in user search interest detection
- Fast and improves the accuracy
- Reduces the clustering difficulty.

**Proposed Algorithm:**

**Algorithm:** PCO
**Input:** the user profile and history
**Output:** reconstructed web page
**Steps:**
1. User profile collection.
2. Perform initial clustering process
3. Identify the user implicit feedback and history
4. Get the recommendation score by performing score calculation model
a. Update the cluster Rc by applying dynamic k means clustering
5. Apply data re construction and personalization process
6. Update the page
7. Reconstructed web page

The PCO algorithms used an online learning and content optimization systems, which is based on the sequence mechanism. Personal Content Optimization scheme eliminates a number of the most common web link extraction which is typically found in web results. It places tags in classes depending on their semantic annotations.
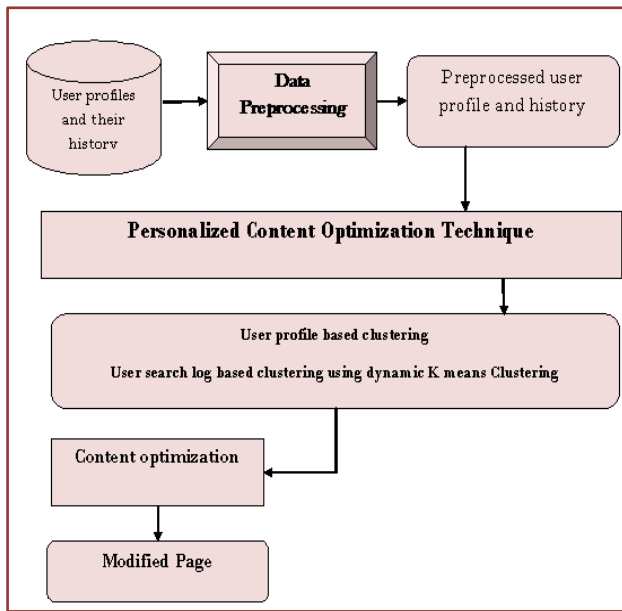
Fig 2: PCO architecture

As a result, this attempts to design a recommender system that achieves content recommendation by automatically estimating user's interest and their needs. The PCO has the following features.

**1. Online dynamic learning Framework.** To attract more users to browse and click content items displayed on the content modules on portal websites, an online learning methodology is necessary because it enable us to model users' behaviors (i.e., clicks, time, downloads and views) on the portal websites as implicit feedbacks and update the recommendation model accordingly in real time (or almost real time), so as to serve more attractive and relevant content to users.

**2. Score Calculation model.** To build efficient online recommendation model, the simple but reliable method is to apply a devoted model for each candidate content item to estimate its attractiveness/ relevance score. Using these dedicated score calculation models; this can rank all items by their respective recommendation scores in the descending order and present the top graded ones to users. From this state of online erudition where real-time user feedbacks are obtainable, the ranking score of an item can be estimated by its click count, which represents a strong signal of attractiveness of this item to users.

**3. Personalization:** Personalization has become very important for content optimization as it provides users with a customized experience of highly attractive and relevant content, so as to improve user commitment, adaptation, and long-term reliability. To begin personalization for content optimization, our learning framework utilizes a dynamic k means clustering strategy. In particular, this group's user into a few unusual groups based on user profiles; for every group of users, the suggestion system provides them with the models which are updated using user actions only by those belonging to the same group.

## 5. EXPERIMENTAL RESULT

The experimental result shows the improvement of our proposed architecture when compare to the existing one. Our system used a dynamic dataset, which can be any number of user weblog created by website

To validate the proposed approaches, this conducts experiments on the data from a real-world content recommendation module. This dynamically collects the events in terms of views and clicks from a random learning bucket of the website.

**Dataset:**

The proposed systems user click through log format



| URL | User ID |
|---|---|
| http://localhost:3366/Home.aspx | Anu |
| http://localhost:3366//sun.aspx | Anu |
| ttp://localhost:3366//moon.aspx | Anu |

Fig 3: USER CLICK THROUGH LOG

**Clicked URLs by User:**

```
www.thesun.co.uk/
www.nineplanets.org/sol.html
www.solarviews.com/eng/sun.htm
en.wikipedia.org/wiki/Sun
www.thesunmagazine.org/
www.space.com/sun/
en.wikipedia.org/wiki/The_Sun_(newspaper)
```

**Un-clicked URL:**

```
imagine.gsfc.nasa.gov/docs/science/know_l1/sun.html
www.nasa.gov/worldbook/sun_worldbook.html
www.enchantedlearning.com/subjects/astronomy/sun/
```

**Result Evaluation**

To evaluate the performance of the proposed schemes, execution time and storage are the main measurement of performance evaluation. Without loss of generality, this defines processing delay and clustering delay for deployed clustering. Processing delay indicates the execution time for clustering to produce frequent items and corresponding interest before page load. Goal detection delay is also evaluated by measuring time spent on processing time on clustering frequent items and interest in the proposed schemes.

| Matrix | Method 1 | Method 2 | PCO |
|---|---|---|---|
| clustering Delay | 3702.09 | 2984.06 | 2108.08 |
| Result Accuracy | 89 | 90 | 94 |
| time | 73.71 | 70.68 | 49.69 |
| Number of iterations | 64.52 | 57.81 | 48.21 |

Table 1: comparison of the proposed system

The above table shows the performance comparison of the proposed method with other existing approaches based on the six different metrics clustering delay, time, processing delay, number of iterations.

Performance comparison of proposed PCO using event collection with existing approaches based on Result accuracy
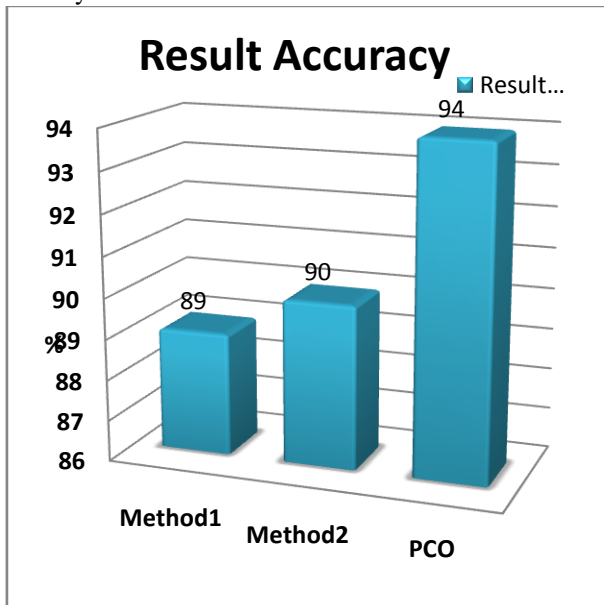


Char 2: Result Accuracy

From the chart it shows the performance measure based on the accuracy of detected cluster and the proposed approach PCO took less time while comparing the other methods and the worst based on the accuracy is method1.
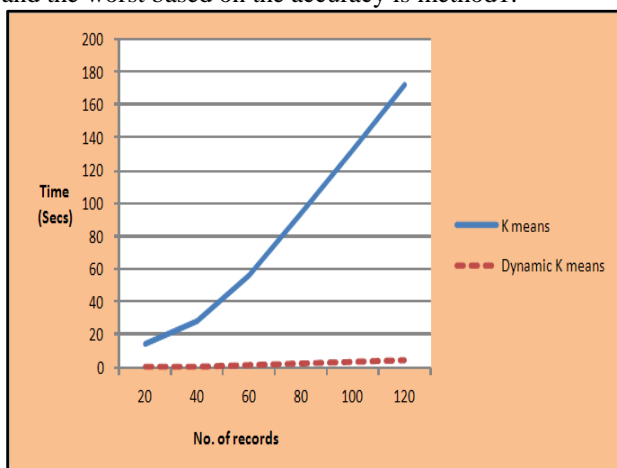


Chart 3:  Performance of Dynamic k means Algorithm in profile Clustering Algorithm with respect to Time Analysis

Experiments were carried out to compare the performances of K-means algorithm and Proposed Dynamic K means Clustering Algorithm by varying the number of the user profiles. The variation of clustering speed with the change in number of records is studied for these algorithms. By comparing the clustering speed for various numbers of transactions, it is observed that the Dynamic algorithm has relatively good performance compared to K-means from the chart 3.

# 6. CONCLUSION & FUTURE ENHANCEMENT

Our proposed PCO scheme has been used to optimize the web content according to the user interest. The system performs the data clustering using dynamic K means clustering algorithm. The system identifies user interest by clustering its implicit feedback sessions represented by event-monitoring with the help of Clicked URLs. The PCO can enrich the URLs with additional textual contents including the titles, snippets and tags and labels. Based on these dynamic cluster information user search interest can then be discovered and represented with some keywords. The main advantage of the proposed method is the implementation of link stay time of user in every link. This can exactly reflect the user interest, which more helps to optimize the web page. In future work, the user search techniques will be extended with some other semantic information.

## REFERNECES

[1]   Web mining: information and pattern discovery on the World Wide Web R. Cooley, B. Mobasher, and J. Srivastava, 8 Nov 1997
[2]   Nasraoui, O. Krishnapuram, R. Joshi, A. Missouri Univ., Columbia, MO 06 August 2002
[3]   Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data J. Srivastava, R. Cooley, M. Deshpande, and P.-N. Tan, 29 September 2009
[4]   From user Access Patterns to Dynamic Hypertext Linking, Tak Woon Yan, Matthew Jacobsen , Feb 1996.
[5]   Adaptive Web Sites: Automatically Learning for User Access Pattern Mike Perkowitz and Oren Etzioni, Seb,1997
[6]   Data Mining of User Navigation Patterns, Mark Levene and Jos´e Borges, August 29, 2000
[7]   Ozmutlu, H. Cenk, and Fatih Çavdur. "Application of automatic topic identification on excite web search engine data logs." *Information Processing & Management* 41.5 (2005): 1243-1262.
[8]   Boldi, Paolo, et al. "Query suggestions using query-flow graphs." *Proceedings of the 2009 workshop on Web Search Click Data*. ACM, 2009.
[9]   Feng, Juan, Hemant K. Bhargava, and David M. Pennock. "Implementing sponsored search in web search engines: Computational evaluation of alternative mechanisms." *INFORMS Journal on Computing* 19.1 (2007): 137-148.
[10]  Ozmutlu, H. Cenk, and Fatih Çavdur. "Application of automatic topic identification on excite web search engine data logs." *Information Processing & Management* 41.5 (2005): 1243-1262.
[11]  W. Chu, S.T. Park, T. Beaupre, N. Motgi, and A. Phadke, "A Case Study of Behavior-Driven Conjoint Analysis on Yahoo! Front Page Today Module," Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD), 2009.
[12]  A. Das, M. Datar, A. Garg, and S. Rajaram, "Google News Personalization: Scalable Online Collaborative Filtering," Proc. 16th Int'l World Wide Web Conf. (WWW), 2007.
[13]  L.M. de Campos, J.M. Ferndez-Luna, J.F. Huete, and M.A. Rueda-Morales, "Combining Content-Based and Collaborative Recommendations: A Hybrid Approach Based on Bayesian Networks," Int'l J. Approximate Reasoning, vol. 51, no. 7, pp. 785-799, 2010.
[14]  D. Downey, S. Dumais, D. Liebling, and E. Horvitz, "Understanding the Relationship between Searchers' Queries and Information Goals," Proc. 17th ACM Conf. Information and Knowledge Management (CIKM), 2008.
[15]  E. Gabrilovich, S. Dumais, and E. Horvitz, "Newsjunkie: Providing Personalized Newsfeeds via Analysis of Information Novelty," Proc. 13th Int'l World Wide Web Conf. (WWW), 2004.
[16]  L.A. Granka, T. Joachims, and G. Gay, "Eye-Tracking Analysis of User Behavior in WWW Search," Proc. 29th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), 2004.
[17]  Z. Guan and E. Cutrell, "An Eye Tracking Study of the Effect of Target Rank on Web Search," Proc. SIGCHI Conf. Human Factors in Computing Systems (CHI), 2007.
[18]  J.L. Herlocker, J.A. Konstan, A. Borchers, and J. Riedl, "An Algorithmic Framework for Performing Collaborative Filtering,"

Proc. 29th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), 1999.

[19] T. Hofmann and J. Puzicha, "Latent Class Models for Collaborative Filtering," Proc. 16th Int'l Joint Conf. Artificial Intelligence (IJCAI), 1999.

[20] R. Jin, J.Y. Chai, and L. Si, "An Automatic Weighting Scheme for Collaborative Filtering," Proc. 29th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), 2004.

[21] T. Joachims, L. Granka, B. Pan, H. Hembrooke, and G. Gay, "Accurately Interpreting Clickthrough Data as Implicit Feedback," Proc. 29th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), 2005.

[22] D. Kelly and N.J. Belkin, "Reading time Scrolling and Interaction: Exploring Implicit Sources of User Preferences for Relevance Feedback," Proc. 29th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), 2001.

[23] Bian, Jiang, et al. "User Action Interpretation for Online Content Optimization."*Knowledge and Data Engineering, IEEE Transactions on* 25.9 (2013): 2161-2174.

[24] D. Kelly and N.J. Belkin, "Display Time as Implicit Feedback: Understanding Task Effects," Proc. 29th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR), 2004.

[25] H.-N. Kim, I. Ha, K.-S. Lee, G.-S. Jo, and A. El-Saddik, "Collaborative User Modeling for Enhanced Content Filtering in Recommender Systems," Decision Support Systems, vol. 51, no. 4, pp. 772-781, 2011.