

Web Search Personalization Based on Browsing History by Artificial Immune System

Hamid Rastegari and Siti Mariyam Shamsuddin

*Soft Computing Research Group,
Universiti Teknologi Malaysia
E-mail: rhamid2@live.utm.my, mariyam@utm.my*

Abstract

Different users have different needs when they submit a query to a web search engine. Common search engines investigate the World Wide Web (WWW) and find many pages according to user query regardless the query submission. The aim of web search personalization is to tailor search results for that particular user. One issue in this field is to gather user information without user intervention. Hence, this paper presents the insight of Artificial Immune Systems (AIS) for user profile construction and probing highly relevant pages in search results. From our investigations, we found conclude that AIS is a suitable personalization tool to knowledge discovery due to its adaptability and learning capability that are required to solve web search personalization.

Keywords: *Personalization of Web Search, Analyzing User Behavior, Artificial Immune System, Construction of User Profile, Browsing History.*

1 Introduction

World Wide Web (WWW) is the largest and most accessible source of information. Usually, web structures are large and sophisticated and users often miss the goal of their inquiry, or receive ambiguous results when they try to navigate through them. Users seek a subject that they need information accordingly. Commonly, the search engines investigate the relevant web and pages according to the user query. Quite often the users find a lot of information for each subject through the web. However, one of the issues is to find the useful

information from search result. Therefore, there are many studies on investigating the importance of personalization in web search engine [1-4].

Personalization of search results is defined as any action to finding more relevant pages in search results list for particular user or a set of users. The objective of a personalization system is to “provide information that users want or need exactly, without expecting from them to ask for it explicitly” [5] .

Web personalization is one of the fastest-growing segments of the Internet economy. Because it can help in reducing information overload and give users a more customized experience of a web site, search personalization can reduce waste time to find requested information on the web. Recently, researchers have used Artificial Immune System (AIS) as a web personalization technique to optimize many problems [6],[7],[8].

AIS is defined by de Castro and Timmis as “adaptive systems, inspired by theoretical immunology and observed immune functions, principles and models, which are applied to problem solving” [9]. The immune system is a vast, complex, interconnected network of agents and processes. While the innate immune system is of great importance to our wellbeing, it is the adaptive immune system that most AIS algorithms take inspiration from. As its name would suggest, the adaptive immune system may change and adapt over time to provide protection against previously unseen dangers. It is this learning and adaptability that AIS algorithms seek to exploit.

This paper provides an overview of the topic of analyzing user behavior in web search personalization. In Section 2, we describe the process of personalization of web search results and the current challenges. Section 3 describes the various techniques used in generating a personalized web search, while in Section 4; the Artificial Immune Systems (AIS) is given. Section 5 provides an association of AIS in the web search personalization, and finally the paper is concluded in Section 6 with a discussion on the current state and future direction.

2 Challenges in Personalization of Search Results

Web search engines help users to discover useful information on the web according to user query. When the same query is submitted by different users, most search engines return the same results regardless the query submission. In general, each user has different requests for his/her query. Gauch [10] mentioned that more than half of the documents returned by search engines are irrelevant information. There are several aspects to the problem [3]. First reason is the problem in synonyms and homonyms terms. Synonyms are words with different spelt and same meaning. Homonyms are words that are spelt the same but have different meanings. Without prior knowledge, there is no way for the search engine to predict user interest from simple text based queries. Secondly, search engines should be deterministic; it should return the same set of documents to all

users with the same query at a certain time. Therefore it is natural that search engines are not designed to adapt to personal preferences.

Personalized system helps every user to overcome the mentioned problems. First of all, the system has to extract interesting keywords for each user. There are two ways to finding interest keywords: explicitly and implicitly. In the explicit approach, user should fill up the registration forms or rate on the visited pages, while implicitly approach finds interesting keywords by examining the historical search and analyzing the user behavior in the web browsing. Second process for personalization system is to apply these keywords for exploring more relevant pages in search results list by filtering and re-ranking techniques.

Web search personalization is still in its infancy [11]. Real-world systems that claim to be doing personalization are often actually offering what we would call customization the ability for users to build profiles of preferences for the content they want to see and the layout it should be displayed in, with users typically choosing from a set number of possibilities. For example, My Yahoo! allows users to specify which news, stock prices, weather, and sports scores they want to be displayed on their My Yahoo! web pages. These preferences are stored in a profile that is used to create the pages each time the user visits [12]. This is what we call customization rather than personalization.

Recently, Google search engine develops a personalization and customization techniques for its user. In fact, Google offered personalized search for only signed-in users for those who have web search history enabled on their Google accounts. This addition in Google enables customization search results for signed-out user based upon 180 days of search activity linked to an anonymous cookie in their browser. It is completely separated from the Google account and web history which are only available to signed-in users.

In the Google personalization system¹, Google makes a user account to keep the search history of the user. It also save the link of pages in search results that user have clicked and viewed. In the next search, the viewed pages in search results are placed on top of page rank. In most situations, this technique benefits the user, but some researchers have mentioned that the best parameters to find the user interested keywords are by calculating the elapsed time and number of clicks on the web pages [13], [14], [15].

In this article, we present alternative approach for personalization of search results by analyzing the number of clicks and elapsed time in browsing history and user past searching. We consider time elapsed to find interesting documents for particular user. In our proposed method, AIS is applied as a model to calculate affinity function between the user interested keywords and the search results.

¹ <https://www.google.com/psearch>

3 Current Approaches on the Search Personalization

Current information retrieval and data mining research have explored the enhancement of user's web experience from several directions. One direction is to create a better structural model of the web, such that it can interface more efficiently with search engines [16]. Another approach is to model user behavior as to predict user's interests better [3]. In this paper, we review relevant studies with the most significant research and focuses on analyzing the user's behavior to construct the user profile for personalization search results. This is presented in the next sub-sections.

3.1 Personalized Browser

Jung [13] developed a customized browser that called Kixbrowser to save users' explicit rating for visited web pages and user activities like: mouse clicks, highlight, key input, size, copy, rollover, mouse movement, add to bookmark, select all, page source, print, forward, stop, duration, the number of visits (frequency), and recently users' browsing time. The author has used linear and nonlinear regression models to predict the explicit rating. From this study, it was proven that the number of mouse clicks is the most accurate indicator for predicting a user's interest level.

Curious Browser was another web browser that recorded the user activities and explicit ratings of users. This browser was used to record mouse clicks, mouse movement, scrolling and elapsed time [14]. The results indicate that the time spent on a page, the amount of scrolling on a page, and the combination of time and scrolling has a strong correlation with explicit interest. In these researches, Jung [13] mentioned that the most accurate indicator is mouse click, but Claypool [14] found that duration and scrollbar movement are very predictive of a user's interest.

Powerize is a system for information retrieval and filtering based on contents and used an explicit user interest model [16]. The authors also reported a way to develop the implicit feedback technique of user modeling for the proposed system. While, Goecks [17] suggested a different method to develop an intelligent web browser that found user's interest without the need for explicitly rating pages. They calculated mouse movement and scrolling activity in addition to user browsing activity. On the other hand, Pan [18] proposed a new method for mining user interest pages. They measured eye-tracking to determine how the displayed web pages are actually viewed. Their experimental environment was restricted to a search results.

3.2 Collaborative Filtering

Collaborative filtering is one of the successful techniques in recommendation systems. The term collaborative filtering is proposed by Goldberg [19]. In this technique, some people collaborates another person by recording the read web

pages to filter and to recommend web documents for the new user. Initially, Goldberg [19] implemented Tapestry system based on collaborative filtering for recommendation system. Nevertheless, recommender systems for large communities generally cannot depend on known individual. Hence, the framework in Tapestry is not appropriate for large communities.

The GroupLens [20], was an automated system for collaborative filtering which filtered Usenet news using the k-nearest neighbor-based algorithm. In this system, a subset of appropriate k users is chosen based on their similarity to the active user, and a weighted aggregate of their rating is used to generate predictions for the active user.

While the two mentioned systems as above rely on explicit ratings, some other systems rely on implicit ratings. For example, Morita and Shinoda [15] exploited “time-spent-reading” as a main parameter in implicit ratings. PHOAKS (People Helping One Another Know Stuff) also uses implicit ratings to construct a recommender system by examining Usenet news postings to find “endorsements” of Web sites [21]. It creates and endorsed a listing of the top web sites in each newsgroup. Some recommender systems also explore user preferences transparently without any extra effort from the users like the recommender systems relying on implicit ratings described above.

In addition, at the E-commerce sites like Amazon.com, CDnow.com and MovieFinder.com, automated collaborative filtering systems have been used with considerable success.

Collaborative filtering can be represented as the problem of predicting missing values in a user-item ratings matrix. Table 1 shows a simplified example of a user-item ratings matrix. In this case “Item i” is predicated to the “user a”

Table 1: User-item ratings matrix for collaborative filtering

	Item 1	Item 2	Item 3	...	Item i	...	Item n
User 1	2	3	-		5		1
User 2	-	3	1		1		2
...							
User a	2	1					
...							
User k	5	1	-		1		-

In the neighborhood-based algorithm, a subset of users is first chosen based on their similarity to the active user, and a weighted combination of their rating is then used to produce predictions for the active user [22]. The algorithm can be summarized in the following steps:

- Weight all users with respect to similarity to the active user. This similarity between users is measured as the Pearson correlation coefficient between their rating vectors.
- Select n users that have the highest similarity with the active user. These users form the neighborhood.
- Compute a prediction from a weighted combination of the neighbor's ratings.

In the recent works, Gao et al. [23] proposed a recommendation method for personalized service in digital recourse which unified partition-based collaborative filtering and meta-information filtering. In partition-based collaborative filtering the user-item rating matrix can be partitioned into low-dimensional dense matrices using a matrix clustering algorithm. The unified method was applied to a digital resource management system. A related problem is that if a user does not rate anything, he/she will receive no recommendation from their system. Similarly, an item cannot be recommended if no user has rated it. A meta-information filtering method can solve this problem.

Another group of researchers have studied on recommender system benefited from artificial Immune techniques. In this case, Acilar et al. [6] proposed a new collaborative filtering method based on artificial immune network as a solution for sparsity and scalability problems.

3.3 Content-Based Recommendation

A content-based recommendation system compares representations of contents and provides recommendations included user interesting contents. This approach use probabilities and envision of the collaborative filtering to find the contents that user is interested in. This model performs three machine learning algorithms as follow:

- Bayesian network
- clustering
- rule-based models

In addition, some authors combine collaborative filtering with content information to provide better recommendation system. Fab [24] constructed a personal filter along with a communal "topic" filter by using relevance feedback to. Web pages are initially ranked by the topic filter and then sent to user's personal filters. The user then provides relevance feedback for that web page, and this feedback is used

to modify both the personal filter and the originating topic filter. Base [25] integrated content and collaboration in a framework where they treat recommendation as a classification task. Melville [26] addressed drawbacks of collaborative filtering systems in their recommender system by exploiting content information of items already rated.

3.4 Reranking search results

Page [27] proposed the first personalized web search by modifying the global PageRank algorithm with the input of bookmarks or homepages of a user. Their work mainly focuses on global “importance” by taking advantage of the link structure of the web. Brin et al. [28] suggested the idea of biasing the PageRank computation for the purpose of personalization, but it was never fully explored. Bharat and Mihaila [29] suggested an approach called Hilltop, which generates a query-specific authority score by detecting and indexing pages that appear to be good experts for certain keywords, based on their links. Hilltop is designed to improve results for popular queries; however, query terms for which experts were not found will not be handled by the Hilltop algorithm. Haveliwala [30] used personalized PageRank scores to enable “topic sensitive” web search. They concluded that the use of personalized PageRank scores can improve web search, but the number of hub vectors (e.g., number of interesting web pages used in a bookmark) used was limited to 16 due to the computational requirements. Kamvar [31] determined that PageRank could be computed for very large subgraphs of the web on machines with limited main memory. Jeh and Widom [2] scaled the number of hub pages beyond 16 for finer-grained personalization.

3.5 Personalization based on browsing History

Another approach for web personalization is to recognize user preference by review of user browsing history and use these preferences to find relevant pages in the search results.

UCAIR [32] is a personalized search agent as a web browser plug-in that designed for web search engine. This system captures user information by saving search history and builds user model to reranking search results. In the other research, Sieg [33] presented an approach to construct ontological user profiles by assigning interest scores to existing concepts in domain ontology. They explained that re-ranking the search results based on the interest scores and the semantic evidence in an ontological user profile are effective in presenting the most relevant search result to the particular user.

In the other way to personalize web search based browsing history, Dou [34] proposed a framework that enables large-scale evaluation of personalized search. In their framework, user clicks on the web pages is used to recording in the search engine logs as a simulate user experiences. In fact, after issue a query to search engine, the user usually checks documents in a result list from top to down. The

user clicks one or more documents that look relevant and skips those documents that the user is not interested in. If a specific personalization method can rerank relevant documents for a user higher in results list, the user would be more satisfied. Hence, it can be a judgment to evaluate search accuracy. Since click-through data can be done at low cost, it is possible to do large-scale evaluation.

4 Artificial Immune System

Immunology can be defined as the study of the defense mechanism that confers resistance against diseases. Immune system is defined a natural organization that protect the human's bodies against the constant attack of external micro organisms. The immune system consists of a complex set of cells and molecules and it is a natural, rapid, and effective defense mechanism for a given host against infections. Natural immune system has the cells which are capable of pattern recognition, diversity, autonomy, noise tolerance, self-organization, learning, gaining memory, fault detection, optimization etc [9]. For benefiting these characteristics of immune system, a new research field is emerged called artificial immune system.

The term artificial immune system refers to a group of computational intelligence techniques that are inspired by and attempt to emulate the information processing capabilities of the biological immune system. Dasgupta provides one of the earliest definitions for the term: "An Artificial Immune System is an intelligent methodology, inspired by the natural immune system, for real-world problem solving" [35]. DeCastro and Timmis provide more detail: "Artificial immune systems can be defined as abstract or metaphorical computational systems developed using ideas, theories, and components, extracted from the immune system" [9]. AIS aim to solving complex computational or engineering problems, such as pattern recognition, elimination, and optimization. This distinguishes AIS from computational models used in biology to simulate and better understand the natural immune system itself (for more information see [9]).

This section, briefly mentioned some ability of AIS corresponding with web search personalization process.

4.1 Pattern Recognition

The natural immune system displays sophisticated pattern recognition capabilities. Self and non-self bias and the other means the immune system uses to identify and respond to threats are exercises in pattern recognition. The immune receptors can identify complex molecular patterns, with the affinity giving a measure of the exactness of the match. The ability to recognize patterns of data without training examples is an important property that found in artificial immune system [7]. It can be use to classification of search results.

4.2 Learning Capability

If we consider a learning system to be one that can perform a task without having to be explicitly programmed for the specifics of that task, then the natural immune system qualifies as a learning system. The innate immune system appears to be largely ‘preprogrammed’ by the organism’s genetic code. However, the adaptive immune system is not explicitly ‘programmed’ for its task. The huge number of possible pathogens, combined with the rapidity with which microorganisms and viruses change, make an enumeration in the genetic code of all possible threats infeasible. Any enumeration would inevitably have ‘blind spots’ consisting of unrecognizable threats. An immune system that could not learn and adapt to new threats would be defeated by some pathogen that found and exploited its blind spots.

4.3 Diversity

Diversity is a feature of the immune system, and a key feature in the ability to recognize and response to a continuously changing environment. It means that the immune system cells can adapt to different components and structure [36]. Like the immune system, the web pages also are diverse. It carries many different information formats and structures, from simple text to PDF formats. The immune system includes a set of various cells with own specialized function and is able to recognize a huge number of different types of antigen. These metaphors could be extracted to produce a system in which different types of cell support different types of data and therefore the ability to identify information contained in these diverse media is a great advantage. Finding relevance information in the web pages with different structure needs a powerful tool to overcome this problem. We have found the artificial immune system as an appropriate and adaptable model to personalized search results in this paper.

5 Artificial Immune System in Personalized Search

The personalization process is traditionally done in several steps with only few variations. The steps can be summarized below:

- Retrieve the users activities represented as log files stored on web servers.
- Preprocess the log files to remove any irrelevant data.
- Discover the usage patterns using a web usage mining algorithm.

The traditional main steps above have been used to discover usage patterns within one specific period of time, but they can be reapplied periodically on the web data to try to capture the changes in navigation patterns. Reapplying the steps periodically can either be performed on the whole data including the newly coming logs, or only on the new log files. However, there are some concerns using this approach [37].

When personalization system uses web log file for pattern discovery, the main issue for using server side data is reliability [37]. Client side data are collected from the host that is accessing the web site. This data are more reliable and accessible. One of the most common techniques for acquiring client side data is to send out a remote agent, implemented in Java or JavaScript [38]. These agents are embedded in web pages, for example as Java applets, and are used to collect information directly from the client, such as the time that the user is spending and leaving the web site, a list of sites visited before and after the current site, and the user's navigation history.

In addition of reliability, Scalability is second issue for server side data. Since a usage data stream can grow to be huge, trying to discover the new behaviors from the accumulated log files each time will require significant computational resources, and could even be impractical or impossible for websites with huge traffic.

After submission a query to a search engine by a user, the search engine returns search results according to the submitted query. In the results list, the user may select number web pages according to request information. Also, the user may access more web pages by following the hyperlinks on the selected web page and continue to browse. In the proposed approach, the system monitors user's browsing history and updates user profile. When the user submits a query in the next time, personalization system re-ranks the search results based keywords on user profile. Figure 1 shoes the proposed framework for personalization of search results by analyzing user's browsing history to construct user profile.

To improve the relevancy of the search result, some researcher use query expansion by modified user query and use synonymous words [39-40]. This method to find appropriate words to include in a query is to compare the previous user queries using semantic similarity measures. If there exist a previous query that is semantically related to the current query, then it can be suggested either to the user or internally used by the search engine to modify the original query [41].

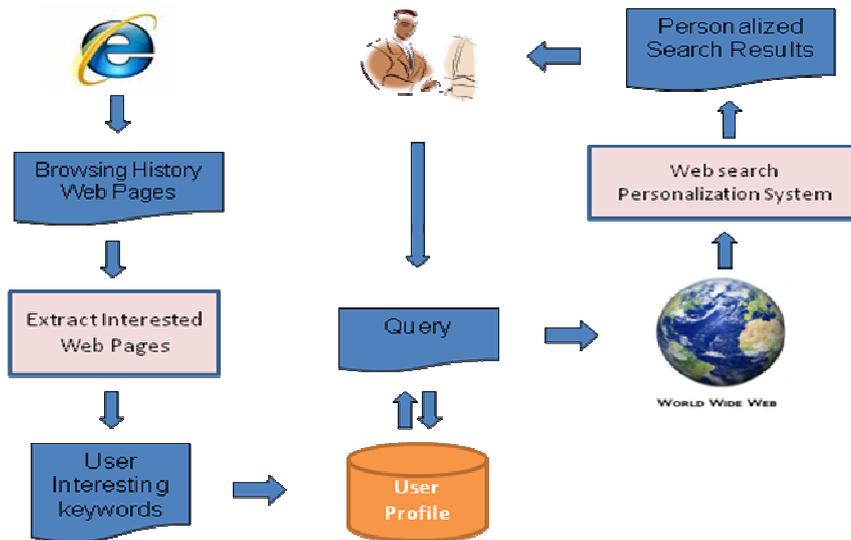


Fig. 1: Web search personalization framework.

5.1 WordNet²

WordNet is a useful tool for archive of words, phrases and relationships between them. WordNet is described as an attempt to map the human understanding of words and its synonyms. WordNet has proven to using in data mining and web mining and it is used to improve the performance of an information retrieval system [42]. In the WWW, different sites for a subject use different terms that they have the same meaning. For instance, a user may looking for a car and submits a query like “buy a car”, while some web sites used this term “purchase a vehicle”. In proposed framework, we use WordNet for creation a synonym word vector to overcome this problem.

When user submits a query, system finds synonyms of the query phrases by using WordNet, and makes a synonym word vector (synset). A search engine like Google receives this vector and search WWW. Search engine searches web for all words or phrases in synonym vector, so gathers a complete set of web pages and delivers to web search personalization system. Web search personalization system selects relevant pages based on user interesting words in user profile.

5.2 User Profile Construction

The first step in the web personalization process is gathering of the relevant data through the browsed web pages, which will be analyzed to provide useful information about the users’ behavior. There are two main sources of data for web

² <http://wordnet.princeton.edu/>

usage mining, corresponding to the two software systems interacting during a web session: data on the web server side and data on the client side.

In the web server side, data are collected and stored in web log files. They consist primarily of various types of logs generated by the web server. These logs record the web pages accessed by the visitors of the site. Web mining tools use web server log files as the main data source for discovering usage patterns. However, log files cannot always be considered a reliable source of information about the usage of a site. The problem of data reliability becomes particularly serious for web personalization, where it is important to identify individual users, in order to discover their interests.

In the client side, gathering data about the user for creating his profile can be done both implicitly and explicitly. The simplest method for collection of data is to collect the user's preferences explicitly through forms, questioners, rating of search results, value elicitation and preference feedback [43].

Although explicitly entered profile information is potentially of "high quality" and provides reliable information about the user, but studies have shown that reluctance and lack of motivation on the user's part to provide information makes the explicit collection of sufficient data for the profile difficult [44].

Therefore, it is required that the user data be collected implicitly by inferring preferences from the user's activities in web browsing. This can be implemented by using the following techniques:

- Monitoring the user's past search queries.
- The search results actually clicked on by the user, clicking and spending time on a link confirms its relevance to the user.
- Monitoring the user's browsing patterns.
- Background information about the user, the IP address of the user gives an idea of his geographical location.

Implicit and explicit data collection methods can be used in conjunction with one another, potentially giving the best of both the worlds [44]. Time spent on visited web pages can be calculated for implicit data collection. Experimental studies in [15] proved that a user usually jumps to another page quickly, if a page is not interesting. However, a quick jump might be caused by the short length of the page; hence the user's interest might be more appropriately approximated by the time spent on a page normalized by the page's length. The proposed approach use term frequency and word density to extract interest keywords in the web pages based on the equation 1.

$$TF_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (1)$$

Where ni,j donates the number of occurrences of the *term* i in the web document dj , and the denominator is the sum of number of occurrences of all terms in the web document dj . In this paper, we used percentage of the terms in the web documents and called word density. Table 2 shows the snapshot of a user profile that the user submitted the “intelligent data mining” query in the Google search engine and clicked on some of links. We analyzed the user behavior in the interest web documents based on the time spent and number of clicks.

Table 2: A snapshot of the user profile

Keyword	Count	Word Density	Weight
data	36	0.078	908.83
mining	23	0.050	751.37
software	9	0.020	63.70
mysql	7	0.015	49.40
application	6	0.013	42.25
source	5	0.011	35.43
workgroup	7	0.014	28.00
analysis	5	0.010	2.50
medicine	5	0.010	2.50
medical	5	0.001	0.03
applications	5	0.025	116.38
intelligent	5	0.025	116.38
results	11	0.015	142.12
techniques	11	0.015	142.12
knowledge	9	0.012	115.94
visualizations	9	0.013	92.30
knowledge	10	0.013	92.30

In this table, “weight” indicates the measure of user interest. We use the user profile to find more relevant web documents in next user query by calculation the affinity measure.

5.3 Affinity function

This section descripts representation of the artificial immune cell job as an agent in the web personalization. The main task of artificial immune cells is to present relevant information in the search result. This information includes a summary of relevant keywords for every web page which is called antigens that will be detected by the cell to determine whether it is relevant to the user’s subject. All the summary of relevant keywords and their hyperlinks for every web page will be saved in a text file for checking. To determine whether the web pages are

relevant, every cell must carry a set of words relevant (user keywords) to the user's subject. If system has detected the current web page is relevant, the cell will remove it from text file and it will be saved into the user interest file.

In the recent work, Secker et al. [8] developed an artificial immune system for interesting information discovery on the web (AISIID). This system applies Artificial Immune System for the collection and ranking of web pages judged to be interesting to the user. AISIID uses a population of immune cells, and again processes inspired by clonal selection, to discover interesting web pages. The user specifies a small collection of web pages that summaries own knowledge on the search subject. Starting on one of the user specified pages; each cell is given a position on the web and is free to move, following hyperlinks that may lead it to other interesting web pages. Each web page it encounters is regarded as an antigen and is therefore available for an affinity evaluation.

The affinity function is a mathematical formulation for measurement of similarity between user interesting keywords in user profile and keywords in search result pages. In fact affinity function is rate of antibody for matching with antigen and removes it. Equation 2 calculates the affinity function for finding the similarity between search results and user profile.

$$\text{affinity} = \frac{(\text{relevance} \cdot \alpha) + (\text{interest} \cdot \beta)}{2(\alpha + \beta)} \quad (2)$$

To calculate the affinity, we need to achieve two value relevance and interest. Relevance is similarity between query word vector that submitted to the search engine and web page keywords in search results. Interest parameter used to find the similarity between user word vector in user profile and web page keywords. Alfa is measurement of web page affinity with new knowledge and Beta is measurement of web page affinity with user knowledge saved in use profile. We mention that if affinity between a query submitted to search engine and user keywords in user profile be zero value, it means this is a new subject for user. In other word:

If Affinity (UWV, QWV) = 0 then Query is a new subject, update user profile

The result is the affinity between the antigen and the immune cell and by definition will return a real number in the range [0, 1]. The relevance of a page is calculated as shown in Equation (3) where the number of words in the QWV (query word vector) that also appear on the Web page are encountered and then normalized by the length of the QWV.

$$\text{relevance} = \frac{\sum_{i=1}^{|QWV|} \delta_i}{|QWV|} \quad \text{where } \delta_i = \begin{cases} 1 & \text{if } QWV_i \in WPK \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

Where QWV_i is the i th component (word) of vector QWV and WPK is the set of keywords in the Web page result. Likewise, the calculation for interestingness is shown in Equation (4), the UWV (user word vector) is compared against the Web page and the count of the number of words present in both the Web page and the UWV is normalized by the length of the UWV .

$$interest = \frac{\sum_{i=1}^{|UWV|} \delta_i}{|UWV|} \quad \text{where } \delta_i = \begin{cases} 1 & \text{if } UWV_i \in WPK \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

Affinity function is used for calculating the similarity between the search results and user profile to finding most relevant web documents. Table 3 depicts reranking the search results from the “rule generator” as a next query in the user profile that mentioned above. In this table, after calculation the affinity function for 50 links in search results, we ordered this list based on value of affinity descendingly and selected the 10 links in top of the ordered list.

Table 3: Personalized results by using the affinity function

Ranking List	URL	Affinity	Google Rank
1	www.prosoxi.gr/.../top-5-online-htaccess-mod-rewrite-rules-generator	0.6410	16
2	markmail.org/message/b557laxqne75qcnw	0.4609	37
3	stackoverflow.com/questions/3694740/rewrite-rule-generator	0.4417	39
4	ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1052348	0.3750	15
5	www.downv.com/Linux-software-download/iptables-rule-generator	0.3043	28
6	www.searchenginegenie.com/mod-rewrite-generator.php - United States	0.2963	7
7	www.ditii.com/2010/03/11/css3-cross-browser-rule-generator/	0.2804	19
8	www.downv.com/Linux-software-download/iptables-rule-generator	0.2667	21
9	vrt-sourcefire.blogspot.com/.../introduction-to-shared-object-rules.html	0.2581	14
10	www.ncbi.nlm.nih.gov/pubmed/21030735	0.1667	9

As a benefit of this work, figure 2 shows the comparison between the proposed approach and common search engines like Google. This figure distinguishes that the personalization of the search results can optimize the search results by analyzing the user behavior and construction the profile for each user.

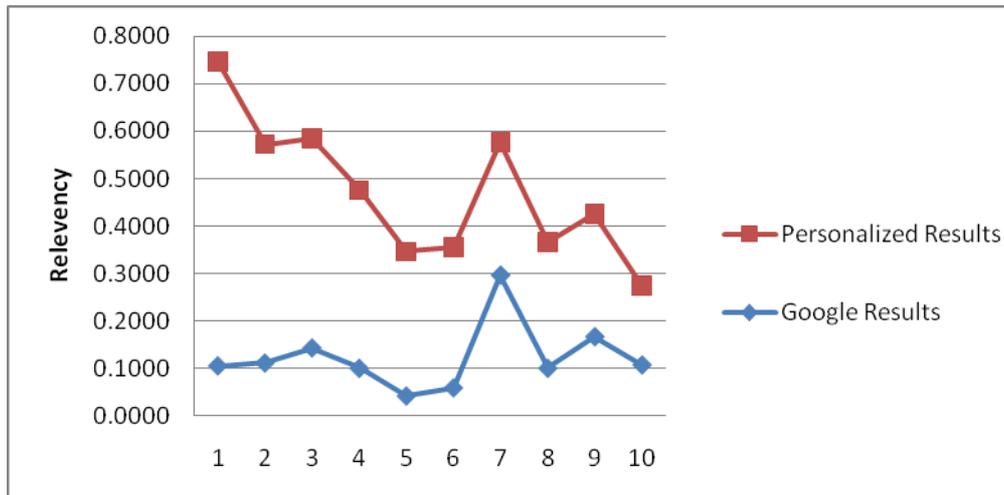


Fig. 2: Comparison the personalized results and Google results for a special query

6 Conclusion

In this paper, we have given the review of web browsing history to extract user's preferences and create a user profile implicitly. The user profile applied to adapt the search results according to the user needs. This approach is quite interesting since it allows each user to perform a fine-grained search by capturing changes in each user's preferences.

This work considers AIS as an alternative model to find more relevant pages in the search results. It is suitable for intelligent search personalization to mine the relevant documents in the search results based on the user preferences. We defined affinity function for detection of user interesting page on web.

Based on the experimental result, web search personalization can improve the search results for particular user in the most of time, but it need a intelligent techniques to update the user profile. For future work, location information and hybridization with collaborative filtering and clustering methods will be considered accordingly.

ACKNOWLEDGMENT

The authors would like to thank *Soft Computing Research Group (SCRG)* for their continuous support in making this research a success.

References

- [1] Sugiyama, K., K. Hatano, and M. Yoshikawa, "Adaptive web search based on user profile constructed without any effort from users", *Proceedings of the 13th international conference on World Wide Web*, ACM: New York, NY, USA, (2004), pp.675-684.
- [2] Jeh, G. and J. Widom, "Scaling personalized web search", *Proceedings of the 12th international conference on World Wide Web*, ACM: Budapest, Hungary, (2003), pp.271-279.
- [3] Tanudjaja, F. and L. Mui, "Persona: A Contextualized and Personalized Web Search", *IEEE Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS'02)*, (2002), pp. 1232-1240.
- [4] Teevan, J., S.T. Dumais, and E. Horvitz, "Personalizing search via automated analysis of interests and activities", *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, ACM: Salvador, Brazil, (2005), pp.449-456.
- [5] Maurice, D.M., S.A. Sarabjot, and G.B. Alex, "Personalization on the Net using Web mining: introduction". *Commun. ACM*, **43**(8), (2000),pp. 122-125.
- [6] Merve Acilar, A. and A. Arslan, "A collaborative filtering method based on artificial immune network". *Expert Systems With Applications*, **36**(4), (2009), pp. 8324-8332.
- [7] Nazri, M., et al., "A Hybrid Approach for Learning Concept Hierarchy from Malay Text Using GAHC and Immune Network", in *Artificial Immune Systems*. (2009). pp. 315-328.
- [8] Secker, A., A.A. Freitas, and J. Timmis, *AISIID*: "An artificial immune system for interesting information discovery on the web". *Applied Soft Computing*, **8**(2), (2008), pp. 885-905.
- [9] Castro, L.R.d. and J. Timmis, *Artificial Immune Systems: A New Computational Intelligence Paradigm*, Springer-Verlag New York, Inc. 368, (2002).
- [10] Gauch, S., J. Chaffee, and A. Pretschner, "Ontology-based personalized search and browsing", *Web Intelligence and Agent Systems*, **1**(3), (2003), pp. 219-234.
- [11] Jeevan, V.K.J. and P. Padhi, *A selective review of research in content personalization*. *Library Review*, (2006). **55**(9): pp. 556-586.
- [12] Manber, U., A. Patel, and J. Robison, "Experience with personalization of Yahoo! Commun". *ACM*, **43**(8), (2000), pp. 35-39.
- [13] Jung, K., *Modeling web user interest with implicit indicators*, MSc Thesis, Florida Institute of Technology, (2001).
- [14] Claypool, M., et al., "Implicit interest indicators", *Proceedings of the 6th international conference on Intelligent user interfaces*. ACM: Santa Fe, New Mexico, United States, (2001), pp.33-40.

- [15] Morita, M. and Y. Shinoda, "Information filtering based on user behavior analysis and best match text retrieval", *Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval*, Springer-Verlag, (1994), pp.272-281.
- [16] Kim, H.-r. and P. Chan, "Personalized Search Results with User Interest Hierarchies Learnt from Bookmarks", *Advances in Web Mining and Web Usage Analysis*, (2006), pp. 158-176.
- [17] Goecks, J. and J. Shavlik, "Learning users' interests by unobtrusively observing their normal behavior", *Proceedings of the 5th international conference on Intelligent user interfaces*, ACM: New Orleans, Louisiana, United States, (2000), pp.129-132.
- [18] Pan, B., et al., "The determinants of web page viewing behavior: an eye-tracking study", in *Proceedings of the 2004 symposium on Eye tracking research & applications*, ACM: San Antonio, Texas, (2004),pp.147-154.
- [19] Goldberg, D., et al., "Using collaborative filtering to weave an information tapestry". *Commun. ACM*, **35**(12), (1992), pp. 61-70.
- [20] Konstan, J.A., et al., "GroupLens: applying collaborative filtering to Usenet news". *Commun. ACM*, **40**(3), (1997), pp. 77-87.
- [21] Terveen, L., et al., "PHOAKS: a system for sharing recommendations. *Commun. ACM*", **40**(3), (1997), pp. 59-62.
- [22] Kazunari, S., H. Kenji, and Y. Masatoshi, "Adaptive web search based on user profile constructed without any effort from users", *Proceedings of the 13th international conference on World Wide Web*, ACM: New York, NY, USA, (2004), pp. 675 - 684.
- [23] Gao, F., et al., "Personalized Service System Based on Hybrid Filtering for Digital Library". *Tsinghua Science & Technology*, **12**(1), (2007), pp. 1-8.
- [24] Balabanovi, M. and Y. Shoham, "Fab: content-based, collaborative recommendation". *Commun. ACM*, **40**(3), (1997),pp. 66-72.
- [25] Basu, C., H. Hirsh, and W. Cohen. "Recommendation as classification: Using social and content-based information in recommendation", *Fifteenth National Conference on Artificial Intelligence*, (1998), pp.714-720.
- [26] Melville, P., R.J. Mooney, and R. Nagarajan, "Content-boosted collaborative filtering for improved recommendations", *Eighteenth national conference on Artificial intelligence*, American Association for Artificial Intelligence: Edmonton, Alberta, Canada, (2002), pp.187-192.
- [27] Page, L., et al., *The PageRank Citation Ranking: Bringing Order to the Web*, Technical Report, (1999).
- [28] Brin, S., R. Motwani, and T. Winograd, What can you do with a web in your pocket, *Data Engineering Buletin*, **21**(2), (1998), pp.37-47.
- [29] Bharat, K. and G.A. Mihaila, When experts agree: using non-affiliated experts to rank popular topics. *ACM Trans. Inf. Syst.*, **20**(1), (2002), pp. 47-58.

- [30] Haveliwala, T.H., et al., *Evaluating strategies for similarity search on the web*, in *Proceedings of the 11th international conference on World Wide Web*, ACM: Honolulu, Hawaii, USA, (2002), pp. 432 - 442.
- [31] Kamvar, S., T. Haveliwala, and G. Golub, "Adaptive methods for the computation of PageRank". *Linear Algebra and its Applications*, 386: pp. 51-65, (2004).
- [32] Shen, Y., L. Xing, and Y. Peng. "Study and application of Web-based data mining in e-business". *Proceedings - SNPD 2007: Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing*, Qingdao, (2007), pp.812-816.
- [33] Sieg, A., B. Mobasher, and R. Burke. "Web search personalization with ontological user profiles", *International Conference on Information and Knowledge Management, Proceedings*, Lisboa, (2007), pp.525-534.
- [34] Dou, Z., et al., "Evaluating the Effectiveness of Personalized Web Search". *IEEE Transactions on Knowledge and Data Engineering*, **21**(8), (2009), pp: 1178-1190.
- [35] Dasgupta, D. *Artificial neural network and artificial immune systems: similarities and differences*, *IEEE Proceeding on Systems, Man, and Cybernetics*, (1997), pp.873-878.
- [36] Nurulhuda Firdaus Mohd, A. "Profile Adaptation in Adaptive Information Filtering: An Immune Inspired Approach", *Proceeding of IEEE conference on Soft Computing and Pattern Recognition*, (2009), pp.414-419.
- [37] Pierrakos, D., et al., "Web usage mining as a tool for personalization: A survey", *User Modelling and User-Adapted Interaction*, **13**(4),(2003),pp. 311-372.
- [38] Shahabi, C., et al., *Yoda: An Accurate and Scalable Web-Based Recommendation System*, in *Cooperative Information Systems*. (2001), pp. 418-432.
- [39] Mitra, M., A. Singhal, and C. Buckley, "Improving automatic query expansion", *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, ACM: Melbourne, Australia, (1998), pp.206-214.
- [40] Biancalana, C., A. Lapolla, and A. Micarelli, "Personalized Web Search Using Correlation Matrix for Query Expansion", *Web Information Systems and Technologies*, (2009), pp.186-198.
- [41] Bollegala, D., Y. Matsuo, and M. Ishizuka, "Measuring semantic similarity between words using web search engines", *Proceedings of the 16th international conference on World Wide Web*, ACM: Banff, Alberta, Canada, (2007), pp.757-766.
- [42] Fellbaum, C., *WordNet: an electronic lexical database*. 1998, MIT Press.

- [43] Anand, S.S. and B. Mobasher, "Introduction to intelligent techniques for Web personalization". *ACM Transactions on Internet Technology*, **7**(4), (2007).
- [44] Keenoy, K. and M. Levene, "Web Search Personalization, in Intelligent Techniques for Web Personalization". *Lecture Notes in Computer Science*, Springer-Verlag, Vol. 3169, (2005), pp:201-208.