

# Web Page Evaluation based on Implicit User Reactions and Neural Networks

Donghoon Lee<sup>1</sup>, Jaekwang Kim<sup>1</sup> and Jee-Hyong Lee<sup>1</sup>

<sup>1</sup> Department of Electrical and Computer Engineering, Sungkyunkwan University  
2066 Seobu-ro, Jangan-gu, Suwon-si, Gyeonggi-do 440-746, Republic of Korea

## Abstract

This paper proposes a method for evaluating web pages by considering implicit user reaction on web pages. Usually users spend more time and make more reactions, such as clicking, dragging and scrolling, while reading interesting pages. Based on this observation, a web page evaluation method by observing implicit user reaction is proposed. The system is designed with Ajax for observing user reactions, and neural networks for learning correlation between user reactions and usefulness of pages. The amounts of each type of user reactions are inputted to neural networks. Also the numbers of characters and images of pages are used as inputs because the amount of users' behaviors has a tendency to increase as the length of pages increase. The experiment is conducted with 113 people and 74 pages. Each page is ranked by users with a questionnaire. The proposed method shows more close ranking results to the user ranks than Google. That is, our system evaluates web pages more closely to users' viewpoint than Google. Although our experiment is limited, our result shows powerful potential of new element for web page evaluation. Some approaches evaluate web pages with their contents and some evaluate web pages with structural attributes, particularly links, of pages. Web page evaluation is for users, so the best evaluation can be done by users themselves. So, user feedback is one of the most important factors for web page evaluation. This paper proposes a new method which reflects user feedbacks on web pages.

**Keywords:** Web page evaluation, User reaction, Neural networks, Ajax.

## 1. Introduction

With the emergence of the Internet, people have been able to access and create the data on the Internet using their personal computer in their home or offices. They can even publish their thoughts and ideas in the Internet freely. This great change has generated a tough problem, searching the right information in the flood of data which exist in the World Wide Web that a system of interlinked hypertext documents accessed via the Internet. In addition, the amount of information in the Web is growing extremely fast, and it is not easy to find right information without help of search engines in the Web.

Usually, search engines have their own web page rank algorithms to recommend users web pages close to query keywords. Web page rank algorithms can be classified into two categories. One category includes the approaches that evaluate web pages with their contents. Those evaluate web pages by checking the relevance of web page content to query keywords [1][2]. The other includes the approaches that evaluate web pages with structural attributes, particularly links, of pages [3][4][5]. Those algorithms assume that pages with more in-links are more useful. One of demerits of those approaches is

mainly depending on web pages. Web page evaluation is for users, so the best evaluation can be done by users themselves. However those approaches hardly reflect users' evaluation or feedback on web pages.

One of frequently used approaches to obtain users' feedback is page view counts. Page view counts are based on the counting how many times users take a look at each page. The reason of many web sites widely use page view counts, is originated from its simplicity and easiness. However, page view counts have several weak points. One of them is that this method cannot reflect user's satisfaction properly. When users search information in the Web, they generally choose pages after reading only the titles or the small parts of contents. Those pre-furnished parts of pages may not enough for users to determine whether the pages are useful for them, so that users often make wrong choices: unrelated pages, changed pages and phishing sites. For this reason, page view counts cannot properly reflect the value of web page contents or users' satisfaction with pages [6]. Another weak point is an abusing problem. People can easily raise page view counts of specific elements by using various methods such as proxy, multi account, many hacked computers, and even other people. This problem causes the reduction of system confidence.

In this paper, we propose a new web page evaluation method by considering users' feedbacks to web pages. Our idea is started from one observation that interesting and useful web pages more stimulate users' reaction [7][8][9][10]. If a page has useful and interesting content, users not only visit the page much, but also spend much time or take various reaction, such as clicking, dragging and scrolling. This means that the usefulness of a page can be evaluated by observing users'

---

Manuscript received Mar. 13, 2012; revised Jun. 7, 2012; accepted Jun. 18, 2012

\*Corresponding Author: Jee-Hyong Lee (jhlee@ece.skku.ac.kr)  
This work was supported by the IT R&D program of MKE/KEIT. [KI001810041244, SmartTV 2.0 Software Platform].

© The Korean Institute of Intelligent Systems. All rights reserved.

behaviors while users are reading pages. To implement this idea, we first collect the data of users' behaviors while reading pages. To collect the data of users' behaviors, such as mouse clicks, mouse point movements and visiting time, we use Ajax which is one of Web 2.0 technology. Ajax can send messages to web servers on the background of web browsers without interrupting users. Since Ajax is operating within web browsers based on the standard web development techniques, users need not to install any agents or programs in their PC.

Next, in order to discover the correlation between the usefulness of web pages and users' reaction to the pages, we use neural networks. Neural networks are used to evaluate web pages with collected data in our system. In order to collect data to train neural networks, we conduct an experiment with volunteers. Volunteers read prepared web pages and answered with the degree of satisfaction to the pages. While they read pages, we observed their implicit reactions. For the evaluation, we compare our result with Google.

Section 2 shows related work and Section 3 describes the proposed web page evaluation system. The experiment and the analysis are presented in Section 4 and 5, respectively. Section 6 concludes this paper.

## 2. Related Works

There are many web page rank algorithms. Those can be classified into two categories. The first includes approaches that evaluate web pages based on the contents. Most of those approaches ranked documents according to their relevance to a given search query. One of them is BM25 which ranks a set of documents based on the frequency of the query keywords in each document [11]. There are many versions with slightly different components and parameters such as adopting some weight mechanisms which give more weights on keywords surrounded by head tags in HTML documents [1][2].

The second includes approaches that evaluate web pages based on structural attributes, particularly links [3][4][5]. Those algorithms assume that pages with more in-links as more useful, and out-links by more useful pages as more important than out-links by less useful pages. Some approaches use not only HTML structures but also the visual layout of pages. Those algorithms assume that blocks in visual layout of a page are not equivalent. For example, the headline in a news web site is much more attractive to users, while users hardly pay attention to the advertisement or the copyright [12].

Most approaches for web page ranking are heavily dependent on web pages. The most proper evaluation of web pages can be done by users, but those approaches hardly have a mechanism for reflecting user's evaluation or feedbacks on web pages. Among many types of feedback from users, implicit reactions to web pages are known to have a relation to usefulness of web pages [7][9][10][13]. Users have a tendency to spend more time or make more reactions to the page such as mouse clicks, page scrolling, etc., if they think a web page is

useful for them. Thus users' implicit reactions can be useful information for web page evaluation. However, studies related to implicit user reactions focused on only the relation between user reaction and usefulness of web pages but not on the web page evaluation based on user reactions. Studies on evaluating web pages based on user reactions are hardly found in literature.

In this paper, we introduce a new web pages evaluation algorithm based on user reaction. We approach with Ajax technique which is known as one of famous Web 2.0 techniques. Ajax is an abbreviation Asynchronous JavaScript and XML [13].

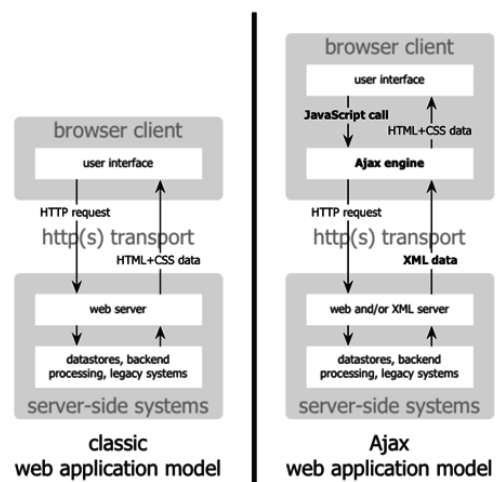


Fig. 1. The traditional model for web applications (left) compared to the Ajax model (right) [13]

Fig. 1 shows the differences between the classic web application model and the Ajax model. 'Ajax engine' is the core module of Ajax. It can be executed without awareness of users, unlike usual HTTP request execution. This merit gives us a chance to collect users' reaction to web pages, such as clicks and drags, without intervention to users [13][14]. Ajax is running on web browsers, so users do not have to install any software. Recently, users do not prefer any installation of programs in their computers because of the security problems. Ajax module is coded in JavaScript and included in web pages. So it runs on web browsers without installation, and thus it is also fast and lightweight. Ajax is using XMLHttpRequest object supported by web browsers. XMLHttpRequest performs the same HTTP requests as a link is clicked. But Ajax can send HTTP requests much faster because it needs fewer amounts of header information and network resources.

## 3. Design of Evaluation System

Our system consists of three main modules: Ajax Log Module, JavaScript Filter and Log Server. JavaScript filter is embedded in the web server, which attaches the Ajax Log Module to the header of requested web pages as shown in Fig. 2. When the

web page to which the Ajax Log Module is attached is loaded on the user's web browser, the Ajax Log Module runs on the background of web browser. The running Ajax Log Module collects user behavior data and automatically sends to the Log Server as shown in Fig. 3.

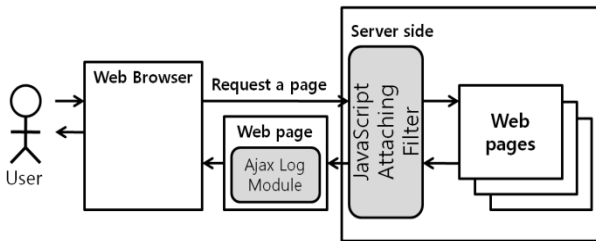


Fig. 2. Process of JavaScript Filter

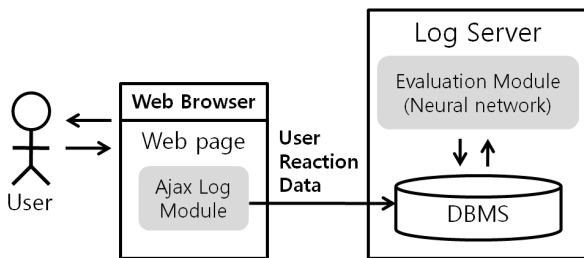


Fig. 3. Ajax Log Module sends user action data to log server

The Ajax Log Module is the collector module which collects users' reaction to web pages using JavaScript while users read pages. Then the collected data is sent to the log server periodically by the Ajax Log Module. JavaScript can catch event messages, accumulate counters, and send messages using Ajax technique. JavaScript can catch various events, but we use only some events that are related to users' behaviors that most people generate while reading such as clicks, scrolling, etc.

Our Ajax Log Engine collects the following events of web browsers:

- **Visiting time:** The visiting time that users spend for reading web pages is relatively one of very important indicators which reflect user's interest in pages. It usually increases if the contents of pages are close to users' interest.
- **Mouse movements:** It is the most behavior of users because users always move their mice while using computers. So we can know, with this event, whether users are using computers or not.
- **Mouse button clicks:** Most of recent web pages have interactive objects: links, movies, buttons, etc. So, users will generate many clicks if they are carefully reading.
- **Mouse wheel movements:** This reaction is directly related to page scrolls, if users are reading web pages, they slowly scroll up pages. If they want to take a look at pages, they may scroll up fast.

- **Key typing except page up/down keys:** This is one of most important behaviors with computers.
- **Page up/down key pressing, i.e. page scrolls:** This reaction is also related to page scrolls.
- **Object selecting:** We select this reaction with similar reason to mouse button clicks.
- **Object dragging:** It is also one of important behaviors following object selections.

The JavaScript Filter is the module that attaches the Ajax Log Module to the HTTP responses. This module is for easy installation of the Ajax Log Module. In our system, each page should embed the Ajax Log Module to collect users' behaviors. We need to devise how to easily modify web pages so that pages embed the log module. Modifying every web page is not easy in a big system. To solve this problem with less effort, we design this module, JavaScript Filter. Using this filter, the Ajax Log Module was automatically injected to the header part of requested web pages. After the filter adds the Ajax Log Module to the requested page, the modified page is transferred and loaded on the web browser. By using this module, we do not have to manually modify web pages.

The JavaScript filter attaches the following code to every page:

```
<script type='text/javascript'>
    var pageSig = '34234234'; // page ID
    var thisUrl = 'http://...'; // URL of this page
</script>
<script src='/js/ahCollector.js'></script>
```

In the code, 'ahCollector.js' is the file of the Ajax Log Module which contains JavaScript code. The variable pageSig is the page ID which is randomly created on every new request. The variable thisUrl is the URL of the page. The values of these two variables are also sent to the log server when the Ajax Log Module works. These variables are recorded with user reaction data and used for the web page evaluation.

The main goal of the Log Server is the web page evaluation. We locate the evaluation module and DBMS on the Log Server. The evaluation module is a standalone batch program which runs once a day and evaluates web pages based on the collected users' reaction data. And DBMS is a storage which has every data used in our system: collected user reaction data, information of web page, and volunteers' identification data.

In order to evaluate the usefulness of web pages based on the collected users' reaction, we build a neural network as a model for the correlation between the users' reaction and the usefulness of web pages. Neural networks are one of machine learning methodologies, and generally used to model patterns hidden in the complicated data. Neural networks involve a

network of simple processing elements, which can exhibit complex global behavior determined by the connections between the processing elements and the element parameters. Through the weight adjustment of connections, neural networks can learn the past data, or the patterns in the data[15][16]. The details of the neural network which is used in our experiment are described in Section 4.

The Log Server based on the DBMS is the main storage of the collected data sent by the Ajax log module. It stores the pre-retrieved content information such as the numbers of characters and images in the web pages. The pre-retrieved information of the web pages is also used as the input features of the neural network. The server also performs bookkeeping for the evaluation of our method by survey, such as staring, stores the some identification of volunteers, the results of questionnaires, etc.

#### 4. Experiment

For the experiment, seventy four web pages are used and 115 people participated. We collected web pages from Google. The keywords used in search are “recommend Korean domestic sight” and “recommend movie.” The number of pages used for experiment is 37 for each keyword. The pages are collected from the top of the search result of Google. We divide the pages and the users into two groups for training the neural network as follows:

**Group A:** 59 people, who will read 24 pages from ‘recommend movie’ keyword and 13 pages from ‘recommend Korean domestic sight’ keyword.

**Group B:** 56 people, who will read 13 pages from ‘recommend movie’ keyword and 24 pages from ‘recommend Korean domestic sight’ keyword (People and pages are exclusive to group A).

The volunteers in each group log in the experiment system using two different URLs. In the login page, the volunteers must answer small questionnaires to check their identity. Then the system shows the list of the web pages, the order of the pages in the list is the same as the result list of Google. The list consists of the titles of web pages and the links to the web pages. Before the experiment, we notice to the volunteers that there are the two lists which are the search result using each keyword and ask to read whatever pages they think useful. After reading pages, they are requested to answer one question: “Is this page useful to you?”. Each person should answer to the question for evaluating the web page they read. The grade of answer is 1 to 5. As the grade is higher, a page is more useful to users.

The neural network in our system has three layers: one input layer, one hidden layer, and one output layer. The number of nodes of the input layer is 7. The nodes of input layer are 5 user reactions: the amounts of visiting time, mouse movements,

mouse button clicks, mouse wheel movements and object selecting. Among the observed 7 behaviors, key typing and page up/down are excluded because users generated little these behaviors while reading. Users use mainly mice and little keyboard for web navigation. The other inputs are the information on each page: the number of characters and the number of images in each web page. If a page is short, the amount of user behaviors in the page may relatively small even though it is close to users’ interest. So, we need to add the factors which can reflect the length of pages. Since most web pages are composed of text and images, we choose the numbers of characters and images as the inputs. The number of nodes of the hidden layer is 8 and the number of nodes of the output layer is 1. The result of the output node is the evaluation value of our system.

The users’ reaction data of 24 pages from ‘recommend movie’ keyword of group A and 24 pages from ‘recommend Korean domestic sight’ keyword of group B are used for the training of the neural network. The data of 13 pages from ‘recommend Korean domestic sight’ keyword of group A and 13 pages from ‘recommend movie’ keyword of group B are used for the test of the neural network. The neural network was trained for 30,000 epochs and the learning rate was 0.05.

To remove outliers, we filter out user data which satisfy the following condition:

- All data of people who does not read more than half of given pages are excluded for training the neural network.
- User reaction data of which amount is upper than top 2% or lower than bottom 2% are excluded.
- The data of which the visiting time is less than three seconds are excluded.

#### 5. Result

Table I is the comparison result of the output of the neural network and the users’ evaluation for the ‘recommend movie’ keyword. The first column is for page IDs and the second shows the rank of each page by users. The rank is the order by the users’ evaluation from survey. The third column is the order by the neural network. The last column is the order presented by Google. We regard it as the rank of page evaluations by Google. For example, the page with ID 40 is ranked as the 1<sup>st</sup> by the users, the 1<sup>st</sup> by neural networks, and also 1<sup>st</sup> by Google.

TABLE 1. COMPARISON OF THE SEARCH RESULT WITH “RECOMMEND MOVIE”

Page ID	Rank by Users	Rank by Neural Network	Rank by Google
40	1	1	1
47	2	5	3
63	3	9	8

44	4	6	2
66	5	3	9
69	6	2	10
75	7	11	12
50	8	8	4
79	9	13	13
60	10	7	7
53	11	10	5
72	12	4	11
57	13	12	6

To compare the results of our method and Google, we calculate the distance to the user answer. Eq. (1) is the equation for the distance between the rank by the users and the rank by the neural network, and Eq. (2) is for between the rank by the users and the rank by Google.

$$\sum_{p \in \text{pages}} |r_p^U - r_p^{NN}| = 38 \tag{1}$$

$$\sum_{p \in \text{pages}} |r_p^U - r_p^G| = 46 \tag{2}$$

### 6. Conclusion

In this paper, we have introduced a web page evaluation system which considers users’ reaction on the web. Our newly designed system can collect users’ reaction data using Ajax and evaluate the usefulness of web pages using neural networks. Owing to Ajax, we collected several kinds of user reaction data without user intervention and additional software. The neural network, which we used for web page evaluation with users’ reaction, effectively worked. We also used the amount of contents of the web pages for the input of the neural network, so that our system could also learn the relation among the usefulness of web pages, the amount of user reactions and the amount of content of web pages. Although our experiment is limited to two keywords, small number of web pages and volunteers, our result shows powerful potential of new element for web page evaluation.

### References

[1] S. Robertson, H. Zaragoza, and M. Taylor, “Simple BM25 Extension to Multiple Weighted Fields,” *In the Proceedings of the 13th ACM International Conference on Information and Knowledge Management*, Washington D.C., USA, ACM, New York, pp.42-49, 2004.

[2] M. Cutler, Y. Shih, and W. Meng, “Using the Structure of HTML Documents to Improve Retrieval,” *In the Proceedings of the USENIX Symposium on Internet Technologies and Systems on USENIX Symposium on Internet Ethnologies and Systems*, California, USA, USENIX Association, Berkeley, CA, pp.241-251, 1997.

[3] L. Page, S. Brin, R. Motwani, and T. Winograd, “The PageRank Citation Ranking: Bringing Order to the Web,” *Technical report, Stanford Digital Library Technologies Project*, Stanford University, Palo Alto, 30 Oct., 1998.

[4] J.M. Kleinberg, “Authoritative Sources in a Hyperlinked Environment,” *Journal of the ACM*, vol. 46, issue 5, pp.604-632, 1999.

[5] S. Brin, and L. Page, “The Anatomy of a Large-scale Hypertextual Web Search Engine,” *In the Proceedings of the 7th International Conference on World Wide Web*, Brisbane, Australia, Elsevier Science Publishers Amsterdam, The Netherlands, pp.107-117, 1998.

[6] S. Bhat, M. Bevans, and S. Sengupta, “Measuring Users’ Web Activity to Evaluate and Enhance Advertising Effectiveness,” *Journal of Advertising*, vol. 31, no. 3, pp.97-106, 2002.

[7] M. Claypool, P. Le, M. Wased, and D. Brown, “Implicit Interest Indicators,” *In the Proceedings of the 6th International Conference on Intelligent User Interfaces*, Santa Fe, New Mexico, USA, 2001, ACM, New York, pp.33-40, 2001.

[8] R. Badi, S. Bae, J.M. Moore, K. Meintanis, A. Zacchi, H. Hsieh, F. Shipman, and C.C. Marshall, “Recognizing User Interest and Document Value from Reading and Organizing Activities in Document Triage,” *In the Proceedings of the 11th International Conference on Intelligent user Interfaces*, Sydney, Australia, ACM, New York, pp.218-225, 2006.

[9] R. Atterer, M. Wnuk, and A. Schmidt, “Knowing the User’s Every Move? User Activity Tracking for Website Usability Evaluation and Implicit Interaction,” *In the Proceedings of the International World Wide Web Conference Committee (IW3C2)*, Edinburgh, Scotland, ACM, New York, New York, pp.203-212, 2006.

[10] T. Joachims, L. Granka, B. Pan, H. Hembrooke, F. Radlinski, and G. Gay, “Evaluating the Accuracy of Implicit Feedback from Clicks and Query Reformulations in Web Search,” *ACM Transactions on Information Systems (TOIS)*, vol. 25, no. 2, Article 7, 2007.

[11] S.E. Robertson, and S. Walker, “Some Simple Effective Approximations to the 2-Poisson Model for Probabilistic Weighted Retrieval,” *In the Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Dublin, Ireland, Springer, NY, pp.345-354, 1994.

[12] R. Song, H. Liu, J.-R. Wen, and W.-Y. Ma, “Learning Block Importance Models for Web Pages,” *In the Proceedings of the 13th International Conference on World Wide Web*, NY, ACM, pp.203-211, 2004.

[13] J.J. Garrett, “Ajax: A New Approach to Web

Applications,” available at: <http://www.adaptivepath.com/ideas/essays/archives/000385.php> (accessed 2009.10.06 ), 2005.

- [14] R. Baraglia, and F. Silvestri, “Dynamic Personalization of Web sites without User Intervention,” *Communications of the ACM*, vol. 50, no. 2, pp.63-67, 2007.
  - [15] R. Callan, “Artificial Intelligence,” *Palgrave MACMILLAN*, Basingstoke, Hants, 2003.
  - [16] E. Alpaydin, “Introduction to Machine Learning,” *The MIT Press*, Cambridge, MA 2004.
- 



**Donghoon Lee**

Donghoon Lee received his B.S., M.S. degree from Sungkyunkwan University at Suwon, Korea in 2004, and 2009 respectively. He is currently a researcher in LG Electronics. His research interests include intelligence system, user modeling and web data mining,.



**Jaekwang Kim**

Jaekwang Kim received his B.S., M.S. degree from Sungkyunkwan University at Suwon, Korea in 2004, and 2006 respectively. Jaekwang’s research interests include networks, security, and intelligent system. He is currently a Ph.D. candidate student at the Department of Electrical and Computer

Engineering, Sungkyunkwan University, and has also worked as a researcher of the Institute of Information and Communications Technology, Sungkyunkwan University. Mr. Kim is a member of the Korean Institute Intelligent Systems and a student member of the IEEE. Mr. Kim was a pioneer of the system consultant group at Sungkyunkwan University School of Information and Communication Engineering.



**Jee-Hyong Lee**

Jee-Hyong Lee received his B.S., M.S. and Ph.D. degrees in Computer Science from KAIST (Korea Advanced Institute of Science and Engineering), Taejon, Korea, in 1993, 1995 and 1999 respectively. From 2000 to 2002, he was an International Fellow of SRI International, Menlo Park, CA. He served as the Publicity Chair of FUZZ-IEEE in 2009 and the Treasurer of Ubicomp in 2008. He is currently an Associate Professor in the School of Information and Communication Engineering, Sungkyunkwan University, Suwon, Korea. His areas of research include data mining, user modeling and intelligent information systems.