International Journal of Engineering Inventions
e-ISSN: 2278-7461, p-ISSN: 2319-6491
Volume 5, Issue 11 (Dec 2016) PP: 28-31

A Personalized Trend-based Webpage Rank Schema Based on Two Phase Personalization

Dae Young Choi
Department of MIS, YUHAN University, South Korea

Abstract:-Web personalization is generally achieved by using Web cookies, Web server log, Google's Web history, etc. They largely focused on user's preferences. However, they have some constraints on seizing both collective and private intelligence in big data provided by ubiquitous smart spaces such as SNS, Web, etc. To handle both collective and private intelligence in the process of Webpage Rank, we propose a Personalized Trend-based Webpage Rank (PTWR) schema based on two phase personalization. Proposed two phase personalization includes collective personalization (phase 1) using weighting order of keywords from big data such as SNS, Web, etc., and private personalization (phase 2) applying SNS users' characteristics, preferences, etc. Thus, Web search engine can automatically provide users with personalized trend-based Webpage ranking. It is particularly useful to find current focal issues such as trend, what's new, etc., on the Web and SNS.

A questionnaire survey on national IT investment had circulated to interested parties throughout the research institute, IT industry, government but there was a little response. In addition, the result was sometimes biased views on the government IT investment. So, we consider big data analysis on national IT investment. To do the questionnaire survey on national IT investment, the proposed PTWR schema based on big data is applied. As a result, we present trend priority on IT items for the national IT investment.

Keywords: *Personalized Trend-based Webpage Ranking (PTWR), SNS users' characteristics, Two phase personalization*

I. INTRODUCTION

In recent years, social networking is enormously increasing. The rapid adoption of smartphones and SNS (Social Network Services) is driving up the usage of social networking. 'Big data' refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze [7]. Big data generally includes semi-structured data, unstructured data, etc. It does not conform with the formal structure of data models associated with existing relational databases or other forms of data tables. Big data analytics is the process of examining large amounts of data of a variety of types to uncover hidden patterns, unknown correlations and other useful information. Such information can provide competitive advantages over rival organizations and result in business benefits such as more effective marketing and increased revenue. Big data sources may include credit card usage data, Web server logs, Internet clickstream data, social media activity, mobile phone call detail records, information captured by sensors, photos in SNS, etc. In this paper, however, the domain of big data is restricted to SNS data such as Facebook, Twitter, and search terms (keywords) on the Web. Existing Web personalization is generally achieved by using Web cookies, Web server log, Google's Web history, a community-based personalizing Web search approach, etc [2-5, 9, 11]. They largely focused on user's preferences. However, they have some constraints on seizing both collective and private intelligence in big data provided by ubiquitous smart spaces such as SNS, Web, etc. To handle both collective and private intelligence from big data such as SNS, Web, etc., in the process of Webpage Rank, we propose a Personalized Trend-based Webpage Rank (PTWR) schema based on two phase personalization. It is particularly useful to find current focal issues related to a user's query on the Web. Thus, Web search engine can automatically provide users with personalized trend-based page ranking

II. TWO PHASE WEB PERSONALIZATION

Using Web search results based on existing ranking algorithms, in phase 1 (collective personalization), Webpage rank is computed by applying weighting order of keywords from big data such as SNS, Web, etc. In this case, weighting order of keywords is mainly derived from text analysis on SNS data such as Facebook, Twitter, and search terms (keywords) on the Web. Vector space model based on term frequency is generally used for text analysis in Web search engines [6, 8]. Some text analysis tools for Web and SNS have already developed within its domain. STAT Search Analytics tracks around 79,000 keywords in Google, Yahoo!, and Bing. Web-based tool that allows anyone to monitor Twitter in real-time for mentions of any words or phrases they choose [10]. In phase 2 (private personalization), Webpage rank tuning is obtained by applying

personalized attributes pattern used in similar SNS users such as age, sex, hobby, etc., as shown in Fig. 1. Thus, Web search engine can automatically provide users with personalized trend-based Webpage ranking.

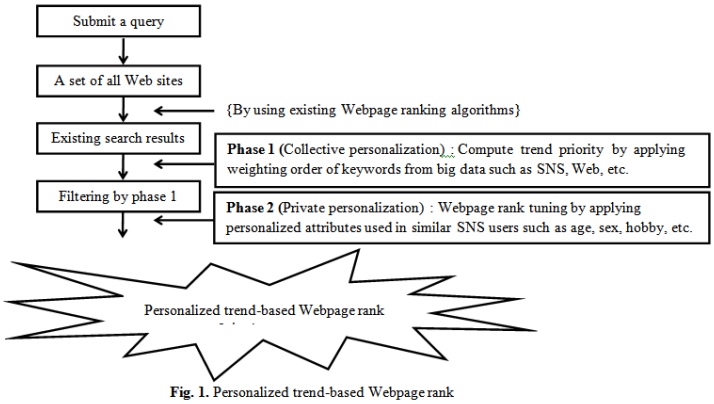

Fig. 1. Personalized trend-based Webpage rank

III. PTWR SCHEMA BASED ON TWO PHASE PERSONALIZATION

Webpage ranking algorithm plays an important role in Web search engines [1, 2]. In existing Web search engines, however, detailed information regarding page ranking algorithms used by major search engines is not publicly available. They often give users many Webpages irrelevant to user's query. They often tend to return unranked random samples in response to user's query [2]. In addition, they do not properly provide both collective and private personalization. To tackle these problems, we propose PTWR schema based on two phase personalization.

PTWR SCHEMA :

PHASE 1 **(**Collective personalization): Based on Web search results using existing ranking algorithms, term weighting [8] can be computed based on the number of occurrences of the term in a text on SNS data such as Facebook, Twitter, and search terms (keywords) on the Web. They become trend priority such as what's new, real-time keywords on the Web, SNS, etc.

PHASE 2 **(**Private personalization): Rank tuning by using personalized attributes pattern used in similar SNS users such as age, sex, nationality, hobby, etc. Thus, personalized trend-based Webpage ranking is obtained (See Fig. 1).

Big data generally includes semi-structured data, unstructured data, etc. To enhance the capability of matching between a user's query and focal keywords on big data in the Phase 1, a fuzzy search [12] based on a fuzzy matching program needs to be considered. Fuzzy search returns a list of results based on likely relevance even though search argument words and spellings may not exactly match. Exact and highly relevant matches appear near the top of the list. The program can also find synonyms and related terms, working like an online thesaurus or encyclopedic cross-reference tool [12]. Applying personalized attributes pattern in the Phase 2, personalized trend-based Webpage ranking can be displayed from the highest rank to the lowest rank. Thus, Web search engine can automatically provide users with personalized trend-based Webpage ranking. It is particularly useful to find current focal issues related to a user's query on the Web and SNS. Together with existing Webpage ranking algorithms, the proposed approach can be used as a complementary and synergistic rather than competitive.

IV. AN EXPERIMENT ON PTWR SCHEMA

To quantify accurately the current state of opinions and the reasons for national IT investment increasing the rate of economic growth, we decided to circulate a questionnaire. A questionnaire survey on national IT investment had circulated to interested parties throughout the research institute, IT industry, government but there was a little response. In addition, the result was sometimes biased views on the government IT investment. So, we consider big data analysis on national IT investment. This experiment was conducted under the following conditions : First, use open API, crawling tool, etc., to collect keyword-related data. Second, boundary of crawling is domestic level. Third, crawling targets for big data are Facebook, Twitter, blog, Internet portal. We try to find IT items' trend priority ranking for national IT investment. After keywords (IT items) selection process, we obtain some keywords for IT items such as 'IoT', 'big data', 'VR', etc. By applying these keywords to big data analysis tool, we obtain raw data in Excel format. After cleansing, transforming and processing the raw data, we obtain the trend priority ranking on IT items for national IT investment as in Table 1.

Table 1. Trend priority ranking on IT items

Priority	Keywords (IT items)	Female (%)	Male (%)
1	IoT	53	47
2	Big Data	36	64
3	Virtual Reality (VR)	61	39
4	Augmented Reality(AR)	48	52
5	Healthcare	43	57
6	Mobile App.	59	41
7	Smart Home	45	55
8	Info. Security	45	55
9	Cloud	47	53
10	Platform	57	43

Now, these keywords for IT items can be implicitly used as a trend search anchor. As a result, Web search engine can automatically provide users with trend-based Webpage ranking as shown in Fig. 1. That is, by applying weighting order (priority) of keywords based on big data, page rank results by using existing Webpage rank algorithms can be adjusted depending on trend priority in big data (See Phase 1(Collective personalization) in Fig.1). By applying trend keywords based on big data to existing SERP(search engine results page), Webpages with the trend keyword have more possibility to be in the upper part of the SERP, compared to the SERP of existing Webpage rank algorithms. Proposed approach uses trend priority to enhance Webpage ranking. Thus, it is particularly useful to find current focal issues such as trend, what's new, etc., on the Web.

As shown in Phase 2 in PTWR schema, if we add personalized attributes pattern used in similar SNS users such as age, sex, culture, locality, nationality, hobby, etc., to the proposed trend-based Webpage ranking, the SERP becomes rank tuning of results returned by existing search engines. The proposed Webpage rank mechanism in Fig. 1 can be explained by SQL-like language : SELECT * FROM {SERP based on existing ranking algorithms} [ORDER BY {trend priority of focal keywords selected from big data} AND {rank tuning by using personalized attributes pattern used in similar SNS users such as age, sex, nationality, hobby, etc.}]. We note that existing Web search engines tend to ignore the importance of [ORDER BY] part. The [ORDER BY] part has 2 phases. In phase 1, Webpage rank adjusted by trend priority of focal keywords selected from big data. In phase 2, rank tuning by using personalized attributes pattern.

We briefly summarize the differences between the proposed Webpage ranking and existing Webpage ranking in Table 2.

Table 2. Comparisons

	Proposed Webpage Ranking	Existing Webpage Ranking
Personalization method	Focal keywords selected from big data + SNS users' characteristics	Web cookies, Web server log, etc.
Webpage rank	Trend-based page rank	Link information, bookmarks, etc.
− For the better personalization, both personalization methods are complementary rather than competitive.		

V. CONCLUSION

We propose PTWR schema based on two phase personalization. Using the proposed schema, Web search engine can automatically provide users with personalized trend-based Webpage ranking. It is particularly useful to find current focal issues related to a user's query on the Web. It is a further step toward personalized trend-based Webpage ranking. We present an experiment on trend-based Webpage rank by using big data. Future works may be extended in several directions based on the current proposal : First, a detail algorithm for rank tuning in Phase 2 could be developed. Second, to enhance Web search results, a local search method [2] on the proposed two phase personalization could be developed. In other words, the proposed approach could be enhanced by applying location-based social networks data. Third, a sophisticate algorithm for integrating trends (keywords) from many big data channels such as Facebook, Twitter, Internet, etc., could be developed.

VI. ACKNOWLEDGEMENTS

The author would like to thank prof. L. A. Zadeh, University of California, Berkeley, for his inspirational address on the perceptual aspects of humans in the BISC seminars and group meetings

REFERENCES

[1] Carriere, J and Kazman, R. WebQuery : searching and visualizing the Web through connectivity, Proceedings of the sixth international conference on the World wide Web, 1997.
[2] Choi, D. Y. Personalized local internet in the location-based mobile Web search, Decision Support Systems, vol. 43, No.1, pp. 31-45, 2007.
[3] Choi, D. Y. A Prototype Model for Handling Fuzzy Query in Voice Search on Smartphones, Advanced Science Letters, Vol. 20, No. 1 (2014) 222-225.
[4] Choi, D.Y. , RA, I.K. Toward a voice interface and personalized local Web search in smartphones, RCIS 2010, 641-645.
[5] Fensel, D and Musen, M. A. The semantic Web : A brain for humankind, IEEE intelligent systems, pp. 24-25, (March/April, 2001).
[6] Kao, B., Lee, J., Ng, C. Y., Cheung, D. Anchor Point Indexing in Web Document Retrieval, IEEE trans. on SMC (part C) 30(3) (2000) 364-373.
[7] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A. H. Big data: The next frontier for innovation, competition, and productivity, McKinsey Global Institute, 2011
[8] Salton, G., Wong, A., Yang, C. S. A vector space model for automatic indexing, Communications of the ACM Volume 18 Issue 11, Nov. 1975, 613-620
[9] Smyth, B. A community-based approach to personalizing Web search, IEEE Computer (August, 2007), pp. 42-50, 2007.
[10] http://monitter.com/
[11] http://searchcrm.techtarget.com/definition/personalization
[12] http://whatis.techtarget.com/definition/fuzzy-search

Dae Young Choi received the B.S., M.S., and Ph.D. degrees in computer sciences from Sogang University, Seoul, South Korea, in 1985, 1992, and 1996, respectively. He was a Research Fellow at the Korea Institute for Defense Analyses (KIDA), Seoul, from 1985 to 1990. He received a postdoctoral fellowship from the Korea Science and Engineering Foundation (KOSEF) in 2000. He was with the Berkeley Initiative in Soft Computing (BISC) Group, Dept. of EECS, UC, Berkeley, as a Visiting Scholar, in 2001. He was also selected as a scholar to be sent abroad under the Professor Dispatching Scheme by the Korea Research Foundation (KRF) in 2004. He was with the Dept. of CSE, Univ. of Colorado at Denver as a Visiting Professor in 2004–2006. Dr. Choi was listed in Marquis Who's Who in Science and Engineering in 2003 and in Marquis Who's Who in the World (2002-2016). He is a Prof. in the Dept. of MIS, Yuhan Univ., Puchon, South Korea. He has a national certificate of Professional Engineer for information processing systems. He was an Associate Editor of the Transactions of the Korea Information Processing Society (Part B) and was on the Editorial Board of the International Journal of Information Processing Systems (IJIPS). His research interests include mobile Web search engines, business intelligence, fuzzy systems, and group decision support system (GDSS), Web personalization.