Yen, B., Ni, K. & Chu, L.K. (2017). Graph transformation based guidance for Web navigation. In *Proceedings of The 17th International Conference on Electronic Business* (pp. 125-131). ICEB, Dubai, UAE, December 4-8.

Graph Transformation Based Guidance for Web Navigation

Benjamin P.-C. Yen*, The University of Hong Kong, Hong Kong, China, benyen@business.hku.hk
Kevin Ni, Southwestern University of Finance and Economics, China, kevinni0326@outlook.com
L.K. Chu, The University of Hong Kong, Hong Kong, China, lkchu@hkucc.hku.hk

ABSTRACT

With growing information volume and diverse user preferences on the web, the performance of web information retrieval has become a critical issue. Web navigation is dramatically influenced by the organizations of web contents. Hence, useful navigation guidance can considerably accelerate the information retrieval process. In this paper, web navigation is formulated as a Directed Group Steiner Forest (DGSF) problem in line graph representation of the website. A heuristic algorithm is proposed to tackle the DGSF problem and attain the suboptimal solution in polynomial time. Simulations are conducted to compare the mean searching time for the proposed DGSF-based navigation guidance and other approaches. The results suggest that the DGSF-based navigation guidance can significantly reduce the mean searching time, especially when the number of web pages is large while the number of destination pages is moderate. The discussion is also made for extending the model to take into account the websites owner's interests and other concerns as well.

Keywords: Information retrieval, Web navigation, guidance, line graph.

*Corresponding author

INTRODUCTION

The abundance of web pages of information that is featured by the Internet has two implications. On the one hand, the abundant information facilitates users to retrieve comprehensive information of interest; on the other hand, it increases the navigation time for users due to the excessive amount of irrelevant information, reducing web usability. Therefore, the accessibility of web pages should be adequately assessed and effectively enhanced. To accelerate information retrieval, navigation guidance can be provided to avoid irrelevant pages. Effective navigation guidance can benefit both surfers and website owners by reducing the searching time to retrieve information, and enhancing the advertisement and transaction functions of websites respectively.

Providing effective navigation guidance is not a trivial task due to the scale and complexity of web-sites (Wang & Yen, 2007). In addition, such a task can be further complicated by the diversity and multiplicity of web pages of interest. Basically, randomly exploring for the destination pages could be time-consuming and tiring. Besides, it is unrealistic to design a web site by having all the pages fully connected. Hence there would be a number of paths for users to follow in reaching a destination page. The major problem then is how to devise a mechanism to help the users to find optimal paths for retrieving the destination pages with minimum searching time. Therefore, this paper focuses on an approach to analyze the problem as the Directed Group Steiner Forest (DGSF) problem (Feldman *et al*., 2012). Line graph conversion (Gross & Yellen, 2006) of the web-site structure is applied for a systematic and intuitive analysis. The DGSF problem, which is essential NP-hard (Alon *et al*., 2006), can be solved approximately in polynomial time using a shortest path based heuristic algorithm proposed in this paper. It is verified by numerical experiments that this heuristic algorithm can efficiently generate navigation guidance for users.

Finally, we would like to comment on the applicability of the proposed web navigation guidance. The DGSF-based navigation guidance is not the only approach to improve the web navigation efficiency. In fact, existing studies have suggested several alternative approaches, such as grammatical inference approach (Korfiatis & Paliouras, 2006), web structure reorganization and optimization (Lin & Liu, 2008), etc. However, these approaches typically require the specialized technique of grammatical inference and the costly reorganization of web structure, provided that web contents are highly standard and the web structure is flexible. By contrast, the proposed web navigation guidance imposes little restriction on website contents and structure. Moreover, the proposed navigation guidance can also be applied in synergy with other approaches, for instance, it is possible to employ the grammatical inference approach to predict the preferred web pages that contain the desired information of a user; then the proposed web navigation guidance can then be generated with these web pages serving as the destination pages. Therefore, we believe this paper is a valuable development and complement to the research on web navigation guidance.

The remainder of this paper is organized as follows. A brief survey of related work is given in next section. A problem modeling is introduced afterwards, covering the web structure represented by the directed graph and its line graph, and the process of formulating the navigation as a Directed Group Steiner Forest (DGSF) problem. Subsequently a shortest path based heuristic algorithm for solving the DGSF problem is described. The numerical experiments for evaluating the performance of the DGSF-based navigation guide are presented. Finally, the paper concludes with summary and future directions.

LITERATURE REVIEW

Designing and maintaining a user-friendly website can be a challenging task. It has aroused awareness and attentions from both practitioners and researchers to look into this issue. Numerous research works have therefore been made on how to improve the websites. There are research projects investigates the general principles of web usability. Practitioners expressed the desire for high web usability, which has catalyzed many research works to address such needs. W3C (World Wide Web Consortium) provided various guidelines to make web content and web applications more accessible to people with disabilities (W3C, 2015). Matera *et al*. (2006) introduced several principles for promoting web usability during the web application lifecycle. They also discussed three classes of usability evaluation methods, namely, user testing, usability inspection, and web usage analysis. Conte *et al*. (2007) further proposed a usability evaluation technique based on the combination of web design perspective and heuristic. Spiliotopoulos *et al*. (2010) compiled experts' insights on applications to enable or facilitate the development of usable web systems. Schaik and Lin (2012) adopted cognitive-experiential approach to model interaction experience and measure nine dimensions of flow experience in human–computer interaction.

Another area of research focuses on the accuracy of user preference prediction. With increasing quantity and diversity of web users, the personalization of web-sites relies heavily on the performance of data mining techniques that underlies the prediction of user preferences (More, 2014). Sarukkai (2000) investigated the use of Markov Chains in link prediction and path analysis. Anderson *et al*. (2002) defined Relational Markov models (RMMs), a generalization of Markov models, make effective learning possible in domains with very large and heterogeneous state spaces, given only sparse data. RMMs were applied to modeling the behavior of web site users, improving prediction in our PROTEUS architecture for personalizing web sites. Perkowitz and Etzioni (2000) described two cluster-mining algorithms, namely PagerGather and IndexFinder, that could automatically gather web pages that were not linked together but were related to some topic in users' mind. Liu and Keselj (2007) then investigated the automatic classification of web user navigation patterns also by the integrated use of both web log data and web contents. The numerical results showed that their experimental system could achieve a classification accuracy of nearly 70% and a prediction accuracy of around 65%, which was much better than the case of using web server logs only.

Some research papers are concerned with the web personalization and recommendation system that improves the web usability. As the information technology evolves, various web personalization and recommendation systems have been proposed by researchers (Perkowitz & Etzioni, 2015). Verma *et al*. (2011) presented a comprehensive survey of over 100 research papers dealing with Web Mining framework for web personalization. Cho *et al*. (2002) then studied a personalized recommendation system for shopping in an on-line marketplace. The novelty of their method was the use of a decision tree induction to avoid poor recommendations that may lead to dissatisfaction of customers. Xuan *et al*. (2016) proposed a framework to identify the various levels of semantic uncertainty in terms of Web event so improve the satisfactions of Web page recommendations. Ho and Bodoff (2014) extended the theory on web personalization by providing a more complete picture of how sampling and processing of personalized recommendations influence a user's attitude and behavior toward the personalization agent. Drawing on social influence and similarity attraction theories, Li and Karahanna (2012) studied and compared social network-based personalization with the traditional peer-based personalization approach of collaborative personalization. Their findings indicated that social network-based personalization can provide accurate personalized offerings

Some research concentrates on improvement of information retrieval for web users. Levene (2010) provided a very clear overall picture about the technical issues of web navigations and search engines. He also extended the view to domains of mobile web and social network. Yen (2006) investigated the problem and suggested that the effectiveness and efficiency of information retrieval can be measured in terms of accessibility and popularity of web pages. By building four accessibility models, an A-P (accessibility and popularity) plot was deployed to guide web designers to improve the structure of web-sites. Aggarwal *et al*. (2014) presented an automated navigation-support tool, CoLiDeS++Pic, suggesting relevant links to click when a user visits a website for searching information to compute semantic similarity between the user goal and the website information using latent semantic analysis technique. Chen and Ryu (2013) proposed a mathematical programming model to improve the user navigation on a website while minimizing alterations to its current structure. Yen and Wan (2010) addressed the importance of efficient information retrieval and discuss the possible methods to improve the efficiency. By assuming one destination page, they analyzed four models of information retrieval with different degree of navigation guidance.

This study aims at investigating the navigation guidance to accelerate the online information retrieval. From the user's perspective, the information of interest was often scattered in several web pages. It is therefore desirable that the web navigation guidance could handle the diversity and multiplicity of destination pages, such that the efficiency of information retrieval could be generally improved. In this paper, the web navigation problem is formulated as a Directed Group Steiner Forest problem. A polynomial-time heuristic algorithm is proposed and simulations are conducted for performance evaluation.

PROBLEM MODELING

The structure of a website is represented as a directed graph, in which each node denotes a web page and each arc denotes a hyperlink. The node-weighted directed graph is then converted to its arc-weighted line graph, whereby a Directed Group Steiner Forest problem is formulated and a heuristics of navigation guidance is proposed for DGSF problems.

Structure of a Web-site

A web-site contains a number of Web pages, and each page may have hyperlinks connecting to other pages. A direct way to describe the web site is using a directed graph with weights assigned on nodes, $G=(V,E,w)$, as follows.

Nodes: $V=\{v_1,v_2,...v_n\}$, where v_i denotes a page, i=1,2,…,n;

Arcs: $E=\{e_{ij}=[v_i,v_j]\ or\ (i,j)\,|$ A hyperlink exists from page i to page $j\}$;

Weights: $w:V\rightarrow R^+$ is the node-associated weight that reflects the loading time.

The loading time would vary from page to page due to the page content and network condition. The page size is the most dominating factor for estimating the loading time. Thus the page size can be defined as the weight of a node to reflect the download time, if exact time estimation is not attainable.

$w:V\rightarrow R^+$

$w(v_i)=x_i$

where x_i is the page size and w is the weight.

As it is more natural to study a graph by estimating the "length" of arcs, or the "distance" between two vertices, using the line-graph representation that facilitates the modeling of the web navigation problem. A conversion of the node-weighted directed graph $G=(V,E,w)$ is made to encode the website structure into the corresponding arc-weighted line graph (Gross & Yellen, 2006). Essentially, the conversion process of $G=(V,E,w)$ to its line graph $L(G)=(D,A,h)$ can be summarized as follows:

$$D=\left\{g_i\,\middle|\,\exists e_{jk}\in E, g_i=e_{jk}\right\};$$

$$A=\left\{(g_i,g_j)\,\middle|\,g_i,g_j\in D,\ \text{and}\ \exists e_{kp},e_{pq}\in E,\ g_i=e_{kp},g_j=e_{pq}\right\};$$

$\forall\,(g_i,g_j)\in A$, $h(g_i,g_j)=w(v_k)$, where $v_k\in V$ is the head of g_i in G;

It is worthwhile to note that the line graph $L(G)$ is not necessarily connected, despite the connectivity of G by definition. With the line graph conversion, the navigation problem can be formulated as selecting the jointly optimal paths to minimize the total length. Another concern for adopting line graph conversion is whether it will incur too much overhead in implementation. The line graph is kept for subsequent calculations once the conversion is done. In principle, it is not necessary to do the line graph conversion for each round of navigation guidance unless the structure of the web-site changes. The overhead for line graph conversion might become considerable in the case that web-site structure is exposed to highly frequent changes.

Directed Group Steiner Forest (DGSF) Problem and A Shortest Path based Heuristic

The preferences of web navigation can be much diversified. It is therefore necessary to provide the web navigation guidance without any restriction on the number or type of the preferred web pages. Assume the user needs to access all the destination pages from root and the sum of the total searching time for destination pages must be minimized which constitutes an optimal web navigation problem. The optimal web navigation problem is formulated as a directed Group Steiner Forest (DGSF) problem, the extension of Group Steiner Forest (GSF) problem (Feldman *et al*., 2012; Korfiatis & Paliouras, 2008).

Directed Group Steiner Forest (DGSF) problem:

Instance: A directed graph $H=(D,A,h)$, where $\left\{h(x_i,x_j):(x_i,x_j)\in A\right\}$ are lengths of arcs, and a set S of setpairs in D.

Objective: Find a min-length subgraph (a forest) X of H that contains a (P,Q)-path (a directed path in H that originates in P while ends in Q) for every setpair $(P,Q)\in S$. Here the overall length of X is defined as: $|X|=\sum_{(x_i,x_j)\in X}h(x_i,x_j)$.

The DGSF problem is essentially NP-hard (Alon *et al*., 2006), so that heuristic algorithms are commonly applied to solve the problem approximately within an acceptable running time. The design of the heuristic is based on the idea of "finding shortest paths to link every (P,Q) setpair in S". It is common to use Dijkstra's algorithm (Dijkstra, 1959) to compute the shortest paths between vertices. However, finding a shortest path for a setpair would be a little bit different. Thus, a revised version of the Dijkstra's algorithm is used to find shortest paths for any setpairs. Then the heuristic can be developed to minimize the overall lengths of the paths for setpairs in *S*.

The shortest path based heuristic algorithm can be described as follows:

1. Initialization. $\tilde{S}\leftarrow S$; $X\leftarrow\phi$.

2. Iteration. While $\tilde{S}\neq\phi$, do:

- Solve: $\min_{(P,Q)\in \tilde{S}} \left\{ \text{dist}(P,Q) \middle| (P,Q) \in \tilde{S} \right\}$ using the revised Dijkstra's algorithm. Let (P_i, Q_i) be the optimum and T_i^* be the corresponding shortest distance path.
- $X \leftarrow X \cup T_i^*$, $\tilde{S} \leftarrow \tilde{S} \setminus \{(P_i, Q_i)\}$.
- $\forall (x,y) \in T_i^*$, $h(x,y) \leftarrow 0$.

End While

3. *Output*. Return X

Once the optimal solution of the DGSF problem is attained, the solution can be employed to generate the navigation guidance. For instance, the hyperlinks in the optimal navigation paths can be highlighted on the web-site.

NUMERICAL EXPERIMENTS

Numerical experiments are conducted in this section to assess the performance of the proposed navigation guidance. The assessment should be made for various structures of web-sites to evaluate the general effectiveness of the navigation guidance. Therefore, simulation approach is firstly applied to assess the performance of the proposed navigation guidance and then complemented with a case study on a real-world website.

Simulation Design

Comparisons are made mainly on the mean searching time of three navigation approaches:

(1) a surfer with DGSF-based navigation guidance;
(2) an experienced surfer without any guidance (e.g. without repeating visits to the same pages); and
(3) a surfer with simple navigation guidance.

To simulate web-sites, the varieties of web site structure should be adequately addressed. Here, the web site structure is randomly generated based on the following procedure:

(1) Set the number of pages n=|V| and number of destination pages k=|N|. Randomly select k destinations from the n nodes.
(2) Given a root, randomly generate an arborescence (or a directed tree) that can spans all the nodes in V. This ensures that every node can be reached from the root.
(3) Randomly generate additional outlinks for each node. According to the distribution of the number of outlinks of a web page described by Kumar *et al*. (Kumar *et al*., 1999), the total number of outlinks in a page varies but is approximately 8 on average.

Apart from the web-site structure, the loading time for each web page is also randomly assigned (uniformly distributed between 0.5 and 1.5). This random assignment of the loading time reflects the moderate differences in the sizes of web pages. As shown by a survey (All Things Web, 2015), the median page size is 44KB; nearly 50% of the pages are less than 30KB and 20% between 31KB and 60KB. Thus the sizes of web pages only exhibit moderate differences, which could be approximately reflected by simulating the loading time between 0.5 and 1.5.

Two parameters, n and k, are set before the simulation. It is estimated that in December 2011, there are roughly 50 billion Google indexed web pages (WorldWideWebSize, 2012) and 555 million web-sites globally (Netcraft, 2011). As a result, the average number of Google indexed web pages per site is 90 (50/0.555=90). Besides, a web user may be interested in some pages of a web-site. For a certain user, these pages of interest commonly just constitute a small portion of the entire web-site. As a result, the initial simulation is conducted by setting n=90 and k=5 (roughly 5% of the entire 90 web pages). Subsequently, simulations with varying n and k are also conducted for the further assessment of the navigation performances. For each group of parameters, we repeated the simulation with different number of runs (from 2000 to 8000), and found that it is sufficient to use 5000 runs to get stabilized simulation results.

Comparison

Reducing the mean searching time is the major concern in providing the web navigation guidance. By comparing with the case for experienced surfers without navigation guidance, the DGSF-based web navigation guidance is illustrated as effective in Table 1. It is also clear that the DGSF-based guidance is more effective in reducing searching time than the simple navigation guidance: 23.4223 vs. 29.3126. Further, the effectiveness of web navigation guidance for various numbers of pages is shown in Figure 1 that DGSF-based navigation guidance significantly outperforms the simple navigation guidance when the number of web pages is not too small.

The number of destination pages is another crucial factor for the performance of the proposed navigation guidance. Each surfer may expect different a number of destination pages. Figure 2 shows that the DGSF-based navigation guidance performs generally better than others, except for the case $k=1$ that it performs similarly as the simple navigation guidance, but much better than experienced surfer without guidance. Figure 2 also suggests that as k grows larger, there is a trend of convergence for the mean searching time of the three navigation approaches.

Table 1: Searching performance with and without web navigation guidance

n=90 k=5	Mean searching time	Average number of visited pages	Mean loading time per page
DGSF-based guidance	23.4223	24.5568	0.9538
Experienced surfer	76.2451	76.1796	1.0012
Simple navigation guidance	29.3126	30.5626	0.9591

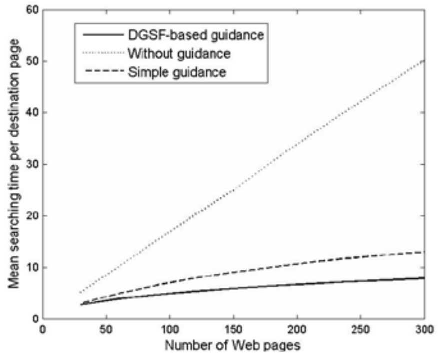

Figure 1: Mean searching time for varying number of web pages (k=5)

Next, it appears from previous analysis that the performance of the proposed navigation is largely determined by the amount of destinations as a proportion of overall web pages in the website. So, use $\eta = k / n$ to denote the proportion of the number of destinations to the overall number of web pages, which measures the relative size of information the surfer wish to extract from the website. Define θ_D and θ_S (the improvements of navigation by guidance) as follows:

$$\theta_D = 1 - \frac{\text{Navigation Time with DGSF-based Guidance}}{\text{Navigation Time without Guidance}}$$

$$\theta_S = 1 - \frac{\text{Navigation Time with Simple Guidance}}{\text{Navigation Time without Guidance}}$$

Then, additional simulations are performed to analyze how θ_D and θ_S should change with η as shown in Figure 3 that the DGSF-based navigation guidance can be 10% better than the simple guidance when the surfer wants to extract a moderate amount (less than 6%) of information from the website.

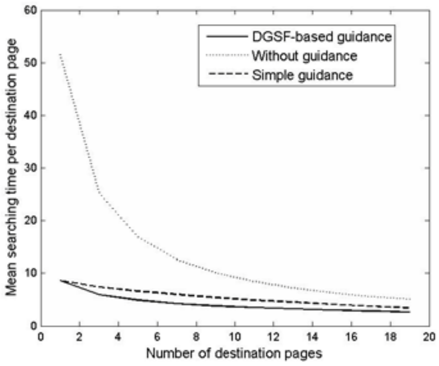

Figure 2: Mean searching time for varying number of destinations (n=100)

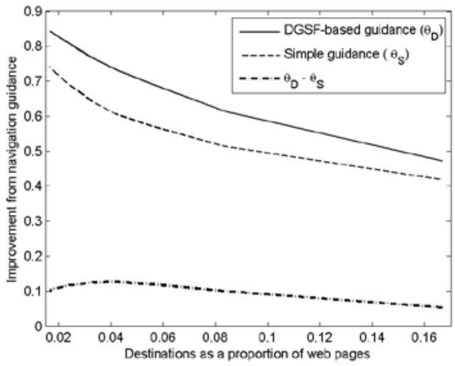

Figure 3: Improvements of navigation by guidance (θ_D, θ_S) as functions of the proportion (η) of the number of destinations to the number of web pages

A Case Study with Real-World Web-Site

To complement the numerical analysis, we conduct a case study using the web-site of an academic institute, the Faculty of Business and Economics of the University of Hong Kong. The website has 29,645 web objects, including 7087 web pages and 22,558 files (e.g. jpg, pdf, docs, pptx or xlsx) and approximately 900,000 links among web objects. The page loading time varies from 0.015 to 13.828 seconds and average is 0.3973 seconds. Figure 4 shows the histogram of page loading time that the majority falls within the interval of [0.2, 0.6] (seconds).

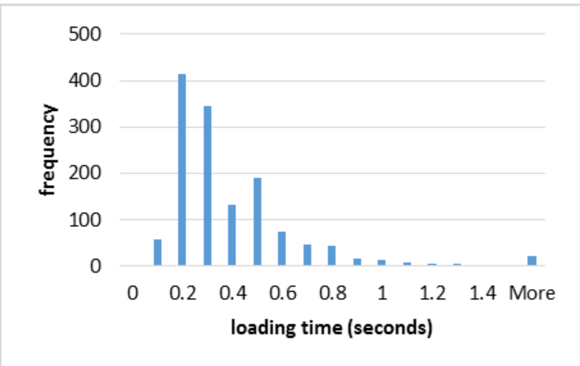

Figure 4: Histogram of the loading time in case

Table 2: Case study of the searching performance with or without navigation guidance

n=1372 k=10	Searching time (seconds)	Number of visited pages	Mean loading time per page (seconds)
DGSF-based guidance	22.5370	62	0.3635
Experienced surfer	74.9361	157	0.4773
Simple navigation guidance	32.3451	83	0.3897

The log file includes 51,995 entries during the period March 20th to March 27th 2016. To analyze the searching performance of navigation guidance, we select the more popular web pages in the HKU business school website as the proxy of destination pages. To do so, we firstly exclude those front pages that are the entry point of web-site, and then sort the remaining web pages according their average times of visit per minute. The top 10 web pages are chosen as the destination pages in the case study. The results are summarized in Table 2. As can be seen, the searching time of the DGSF-based guidance is 22.5370 seconds, which is significantly shorter than those of the simple navigation guidance (32.3451 seconds) and the experienced surfer (74.9361 seconds). Moreover, the mean loading time per page when applying the DGSF-based guidance is also the shortest (0.3635 seconds). Recall that the average loading time of web pages for the entire website is 0.3973 seconds. The result suggests that the DGSF-based guidance can automatically select web pages of shorter loading times along the navigation. Therefore, the searching performance of the DGSF-based guidance in the case study corroborates the previous numerical analysis.

CONCLUSIONS

The analysis mainly focuses on reducing the surfer's time of retrieving web pages. In many cases, however, it is also imperative to take into account the interests of the website owner. A common scenario is that there are advertisements placed on the website; so the website owner would benefit more if a surfer visits more web pages during the course of navigation. The numerical results (Figures 3 and 4) show that the DGSF-based navigation can automatically guide the surfer to intermediary web pages with fast loading, thus allowing for more web pages being visited per unit time. In this way, the website owner's benefits would be implicitly improved with the DGSF-based navigation guidance.

In this paper, we propose an effective approach for web navigation guidance to accelerate the information retrieval. It is assumed that the web structure is represented as a node-weighted directed graph which is then converted to the arc-weighted line graph. Thereby the web navigation problem formulated as a DGSF problem is to find paths with the smallest overall length. Since the DGSF problem is NP-hard, a shortest path based heuristic is used to solve the DGSF problem approximately and generate the DGSF-based navigation guidance accordingly. The proposed heuristic is essentially a polynomial time algorithm which is highly applicable in practice. To assess the effectiveness of the DGSF-based navigation guidance, simulations have been conducted to evaluate its performance for the mean searching time per destination page. The performance is then compared with two other approaches - navigation without guidance and simple navigation guidance. The comparison results show that the DGSF-based navigation guidance can significantly accelerate the online information retrieval.

The research contributions of this paper are as follows. First, the navigation problem is formulated as a DGSF problem in line graph representation of web-site structure, which can be regarded as finding shortest paths to a group of destination pages. Although this modeling approach seems not intuitive, the structure of the node-weighted network is precisely represented by the line graph conversion. Accordingly, a shortest path based heuristic algorithm is proposed to approximately solve the DGSF problem. As a result, the web navigation guidance is developed as an effective way to expedite the retrieval of web pages. The numerical results show that the web navigation guidance is effective in reducing searching time for the surfers, especially when the number of destination pages is moderately large.

Finally, this paper does not take into account all potential issues of the optimal web navigation problem. The algorithm can be extended to solve the general DGSF problem; but questions such as how to take into account the website owner's interests, how to incorporate the possible values that surfers may gain when coming across web pages of interests, and how to balance the interests of different parties, are still important subjects that may motivate future research. To further improve the path selection as guidance, the proposed approach can be extended by considering relationships among webpages (e.g. indexing network) (Jiang *et al*., 2014). The research can also be directed to personalize the navigation by reorganizing the web structure for each user as to provide personalized and adaptive services in intelligent environments (Aztiria *et al*., 2013). In the experiment, mean loading time for visited pages can be also considered for comparison.

ACKNOWLEDGEMENT

This work was supported in part by Hong Kong RGC under grant GRF 719409E and Joint Research Fund (JRF) 15-18. We would like to thank Mr. Chris Cheung and Mr. Culture Wang for their help with the case study of real world website.

REFERENCES

[1] Aggarwal, S., Van Oostendorp, H., & Indurkhya, B. (2014, June). Automating web-navigation support using a cognitive model. In *Proceedings of the 4th International Conference on Web Intelligence, Mining and Semantics* (WIMS14) (p. 19). ACM, Thessaloniki, Greece, June 2-4, 19.

[2] All Things Web (2015). How much is too much?. Retrieved from http://www.pantos.org/atw/35654.html (June 10, 2015)

[3] Alon, N., Awerbuch, B., Azar, Y., Buchbinder, N., & Naor, J. (2006). A general approach to online network optimization problems. *ACM Transactions on Algorithms*, 2(4), 640-660.

[4] Anderson, C.R., Domingos, P., & Weld, D.S. (2002). Relational Markov models and their application to adaptive web navigation. In *Proceedings of the 8th ACM SIGKDD international conference on Knowledge Discovery and Data Mining* (pp. 143-152). ACM, AB, Canada, July 23-25.

[5] Aztiria, A., Augusto, J.C., Basagoiti, R., Izaguirre, A., & Cook, D.J. (2013). Learning frequent behaviors of the users in intelligent environments. *IEEE Transactions on System, Man, and Cybernetics: System*, 43(6), 1265-1278.

[6] Berkovsky S. & Freyne, J. (2015). Web personalization and recommender systems. In *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2015)* (pp. 2307-2308). ACM, Sydney, Australia, August 10-13.

[7] Chen M. & Ryu, Y.U. (2013). Facilitating effective user navigation through website structure improvement. *IEEE Transactions on Knowledge and Data Engineering*, 25(3), 571-588.

[8] Cho, Y.H., Kim, J.K., & Kim, S.H. (2002). An intelligent web recommendation system: A web usage mining approach. *Expert Systems with Applications*, *23*, 329-342.

[9] Conte, T., Massollar, J., Mendes, E., & Travassos, G.H. (2007). Usability evaluation based on web design perspectives. In *Proceedings of the First International Symposium on Empirical Software Engineering and Measurement*, (pp. 146-155). ESEM, Madrid, Spain, September 20-21.

[10] Dijkstra, E.W. (1959). A note on two problems in connection with graphs. *Numerische Mathematik*, *1*, 269-271.

[11] Feldman, M., Kortsarz, G., & Nutov, Z. (2012). Improved approximating algorithms for directed Steiner forest. *Journal of Computer and System Sciences, 78*(1), 279-292.

[12] Gross J.T. & Yellen, J. (2006) *Graph Theory and Its Applications* (2nd ed., p. 265). Boca Raton, FL: CRC Press.

[13] Ho, S.Y. & Bodoff, D. (2014). the effects of web personalization on user attitude and behavior: an integration of the elaboration likelihood model and consumer search theory. *MIS Quarterly*, 38(2), 497-520.

[14] Jiang, C., Sun, H., Ding, Z., Wang, P., & Zhou, M.C. (2014). An indexing network: Model and applications. *IEEE Transactions on System, Man, and Cybernetics: System*, 44(12), 1633-1648.

[15] Klein, P. & Ravi, R. (1995). A nearly best-possible approximation algorithm for node-weighted Steiner trees. *Journal of Algorithms*, 19(1), 104-115.

[16] Korfiatis, G. & Paliouras, G. (2008). Modeling web navigation using grammatical inference. *Applied Artificial Intelligence*, 22(1-2), 116-138.

[17] Kumar, S.R., Raghavan, P., Rajagopalan, S., & Tomkins, A. (1999). Extracting large-scale knowledge bases from the web. In *Proceedings of the 25th VLDB Conference* (pp. 639-650). ACM, Edinburgh, Scotland99.

[18] Levene, M. (2010) *An Introduction to Search Engines and Web Navigation*, John Wiley & Sons.

[19] Li, S. & Karahanna, E. (2012). Peer-based Recommendations in Online B2C e-Commerce: Comparing Collaborative Personalization and Social Network-based Personalization. *Proceedings of IEEE 45th Hawaii International Conference on System Sciences (HICSS)*, Hawaii, USA), 733-742.

[20] Lin, W. & Liu, Y.A. A (2008). Novel Website Structure Optimization Model for More Effective Web Navigation. *Proceedings of the 1st International Workshop on Knowledge Discovery and Data Mining (WKDD)*, Adelaide, Australia), 36-41.

[21] Liu, H. & Keselj, V. (2007)' Combined Mining of Web Server Logs and Web Contents for Classifying User Navigation Patterns and Predicting Users' Future Requests. *Data & Knowledge Engineering*, 61), 304-330.

[22] Matera, F., Rizzo, F., & Carughi, G.T. (2006) Web usability: Principles and evaluation methods, In E. Mendes & N. Mosley (Eds.), *Web Engineering* (pp. 143-180). Springer, Berlin Heidelberg.

[23] More, S. (2014). Modified path traversal for an efficient web navigation mining. In *Proceedings of International Conference on Advanced Communication Control and Computing Technologies* (pp. 940-945). IEEE, Ramanathapuram, India.

[24] Netcraft. (2011) December 2011 web server survey. Retrieved from http://news.netcraft.com/archives/2011/12/09/december-2011-web-server-survey.html (September 20, 2017).

[25] Perkowitz, M. & Etzioni, O. (2000). Towards adaptive web sites: Conceptual framework and case study. *Artificial Intelligence*, *118*, 245-275.

[26] Sarukkai, R.R. (2000). Link prediction and path analysis using Markov chains. *Computer Networks*, *33*, 377-386.

[27] Schaik, P. & Ling, J. (2012). An experimental analysis of experiential and cognitive variables in web navigation. *Human–Computer Interaction*, 27(3), 199-234.

[28] Spiliotopoulos, T., Papadopoulou, P., Martakos, D., & Kouroupetroglou, G. (2010). *Integrating Usability Engineering for Designing the Web Experience: Methodologies and Principles.* Hershey, PA: IGI Global.

[29] Verma, V., Verma, A.K., & Bhatia, S.S. (2011). Comprehensive survey of framework for web personalization using web mining. *International Journal of Computer Applications*, 35(3), 23-28.

[30] W3C (2015) Web accessibility initiative (WAI). Retrieved from http://www.w3.org/WAI/ (September 20, 2017).

[31] Wang, M. & Yen, B.P.-C. (2007). Web structure reorganization to improve web navigation efficiency. In *Proceedings of the 11th Pacific Asia Conference on Information Systems (*pp. 411-422). PACIS, Auckland, New Zealand, July 4-7.

[32] WorldWideWebSize. (2012) The size of the World Wide Web (The Internet). Retrieved from http://www.worldwidewebsize.com/ (September 20, 2017).

[33] Xuan, J., Luo, X., Zhang, G., Lu, J., & Xu, Z. (2016). Uncertainty analysis for the keyword system of web events. *IEEE Transactions on System, Man, and Cybernetics: System*, 46(6), 829-842.

[34] Yen, B.P.C. (2006). The design and evaluation of accessibility on web navigation. *Decision Support Systems*, *42*, 2219-2235.

[35] Yen, B.P.-C. & Wan, Y.-W. (2010). Design and evaluation of improvement method on the web information navigation – A stochastic search approach. *Decision Support Systems*, *49*, 14-23.