Suggesting Relevant Queries Based on Transition Probability of Information in Web Graphs

Anoop V S
P.G Scholar
Dept of Computer Science and Engg
College of Engineering Perumon
Kollam-691601

Deepa K Daniel
Assistant Professor
Dept of Information Technology
College of Engineering Perumon
Kollam-691601

Abstract

Query suggestion is an important process in the case of a search engine to predict the user's information needs. In many cases, we can generate the relevant prediction from large scale Web graphs containing queries and other related information like clickthrough data generated by the search engine. However generating suggestion based on the semantic relevancy with the user's information need is a challenging problem.

In this paper, a modification of the general query suggestion technique is proposed which is based on query-URL graph based on the clickthrough data generated by the search engine. Here the transition probability of information between nodes in the graph is taken as the parameter to suggest relevant queries. Based on this, a sub graph is constructed by short random walk on the graph and the basic heat diffusion equation is applied in the sub graph to suggest the relevant queries. Also the complexity of the existing query suggestion algorithm based on the heat diffusion equation is reduced by this approach.

Keywords- Query suggestion, transition probability, diffusion

1. Introduction

1.1. Overview

Query suggestion techniques focus on suggesting a list of relevant queries to user's input, by mining related queries from previous knowledge, e.g., search engine logs. Due to the exponential nature of information on the web, the organization and presentation of this information becomes extremely difficult. Web search engines are the most to suffer from this problem. In the case of query suggestion, it has become more difficult for a search engine to predict what the user is actually searching for. When a user types a query "msg" to the search engines, he will be provided with quite a few alternative potential queries. For example, he will be suggested "msg Chinese food," "msg health," and "other names for msg" by Google, and "msg error," "msg network," and "msg seating chart" by Yahoo. There are also other query suggestion mechanisms which could automatically complete a query [1], and automatically correct spelling mistakes [2].Depending on the underlying algorithm, query suggestion algorithms can be classified in to graph based models [4,5] and probabilistic models[20,11]. Moreover, depending on the length of suggested queries, we can further divide existing query suggestion techniques into query expansion [12, 13], query reduction [14] and query reformulation [15, 16].

In most cases, such query suggestion mechanisms are formed based on morphological information of queries, or existence of one query word with other queries. The disadvantage of these approaches is that the generated query suggestions do not satisfy the semantic relationship with the original query. For example, people searching for "pop music" maybe interested in "Michael Jackson" but not "POP(Post Office Protocol)." A fine query suggestion system must consider all the possible features, and to ensure that the semantics of the suggested query is in close relation with the original query

Another challenge is the personalization feature of queries. Personalization is nothing but different users have different information needs. For example, consider a query "msd", it could mean "Microsoft Development" or it could be "Mahendra Singh Dhoni".It depends on the person who issues the query. A general query suggestion which does not take semantics of queries in to consideration fails in this case. In most cases, semantics of the queries is hard to define and find out and it is a challenging problem.

In this paper, a unified approach to query suggestion based on transition probability of information in large scale bipartite graph of queries and clickthrough data is proposed. Here the transition probability of information is compared with heat diffusion phenomenon in Web graphs. This method has got several advantages.1) The suggestions

generated do not have to occur with the original query. 2) It can give relevant results which satisfy semantic relationship with original query. 3) This technique can also be used in personalization technique. 4) This method can be used in many recommendation tasks. This model is based on the diffusion of information on both undirected graphs and directed graphs.

The rest of the paper is organized as follows. Section 2 presents the related works on query suggestion. Section 3 presents the heat diffusion model and the modification of this heat diffusion model used for query suggestion.

2 . Related works

Query recommendation.

Query recommendation is one of the main operations in commercial search engines. Most of the work on query recommendation is concentrated on measures of query similarity [21, 22] that can be used for query expansion [23] or query clustering [23, 24]. A first attempt to model the users' sequential search behavior is presented by Zhang and Nasraoui [25]: a dumping factor d is used as a weight for the arcs between consecutive queries in the same session, and the similarity values for non consecutive queries are calculated by multiplying the values of arcs that join them. Instead, Fonseca et al. [22] used a method which relates queries based on association rules.

Baeza-Yates et al. [23] concentrated on the problem of suggesting related queries issued by other users and query expansion methods to construct artificial queries. They tried to recommend queries that are related to the input query but may not be used for the same issue as the input query. Here the term-weight vector representation is obtained from the aggregation of the term-weight vectors representation of the URLs clicked after the query. These term-weight vectors are used for clustering. Wen et al. [24] also used a method for query suggestion which is based on a clustering method that is concentrated on four notions of query distance: the first notion is based on keywords or phrases of the query; the second on string matching of keywords; the third on common clicked URLs; and the fourth on the distance of the clicked documents in some pre-defined hierarchy.

Jones et al. introduced the concept of query substitution. They obtained similar queries by replacing the query as a whole, or by substituting constituent phrases [11]. Similar queries and phrases are derived from user query sessions, and they proposed models for query re-ranking based on the similarity of the new query to the original query.

Antonellis used query-click graph to rewrite the actual query. They used the concept similar to cocitation - two queries are similar if they refer the same document. To do the query rewrites they used a method similar to SimRank [29], which is a generalized measure of co-citation.

White et al. [27, 28] used the query rewrites observed in a query log to generate query recommendation. Given an input query, they generate two lists - (a) the top 100 queries that contain the original query as a sub-string, and (b) the top 100 queries which followed the input query. Then, each candidate query is then scored by multiplying its smoothed overall frequency of following the target query in the past sessions, using Laplacian smoothing.

3. Proposed solution

In this section, a novel graph diffusion model for query suggestion based on transition probability of similarity information is proposed. It is the modification of the query suggestion model which is based on heat diffusion. Here heat diffusion in graph is used to model the transition probability of similarity information in Web graphs. This model can be applied to both undirected graphs and directed graphs. Here we are concentrating only in directed query – URL graphs since we cannot employ the undirected query-URL graphs since these graphs do not interpret the correct relationship between queries and URLs. But for better understanding, we are presenting heat diffusion models on both undirected and directed graphs.

3.1. Heat diffusion

Heat diffusion is a physical phenomenon. It can be defined as in a medium, heat always flows from a position with high temperature to a position with low temperature.

3.2. Diffusion on undirected graphs

Consider an undirected graph G=(V,E), where V is the vertex set, and $V=\{v_1,v_2,\ldots v_n\}$. $E= \{(v_i,v_j)|$ there is an edge between v_i to $v_j\}$ is the set of all edges. The edge (v_i,v_j) is considered as a pipe that connects nodes vi and v_j. The value $f_i(t)$ describes the heat at node v_i at time t, beginning from an initial distribution of heat given by $f_i(0)$ at time zero. f(t) denotes the vector consisting of $f_i(t)$.,

The construction of model as follows: suppose, at time t, each node i receives an amount $M(i,j,t,\Delta t)$ of heat from its neighbor j during a time period Δt. The heat $M (i,j,t,\Delta t)$ should be proportional to the time period Δt and the heat

difference $f_j(t)$-$f_i(t)$. Moreover, the heat flows from node j to node i through the pipe that connects nodes i and j. Based on this consideration, it is assumed that

$$M(i, j, t, \Delta t) = \alpha(f_j(t) - f_i(t))\Delta t,$$

where α is the thermal conductivity—the heat diffusion coefficient. As a result, the difference in heat at node i between time t + Δt and time t will be equal to the sum of the heat that it receives from all its neighbors. This can be formulated as

$$\frac{f_i(t+\Delta t) - f_i(t)}{\Delta t} = \alpha \sum_{j:(v_i, v_j)\in E)} (f_i(t) - f_j(t)) \quad (1)$$

where E is the set of edges. To find a closed form solution to
(1), it is expressed it in a matrix form

$$\frac{f_i(t+\Delta t) - f_i(t)}{\Delta t} = \alpha(H - D)\, f(t) \quad (2)$$

Where

$$H_{ij} = \begin{cases} 1, & (v_i, v_i) \in E \text{ or } (v_j, v_i) \in E \\ 0, & i = j, \\ 0, & otherwise \end{cases} \quad (3)$$

And

$$D_{ij} = \begin{cases} d(v_i), & i = j, \\ 0, & otherwise \end{cases} \quad (4)$$

where $d(v_i)$ is the degree of node vi. From the definition, the matrix **D** is a diagonal matrix.

In order to generate a more generalized representation, all the entries in matrices **H** and **D** is normalized by the degree of each node. The matrices **H** and **D** can be modified to

$$H_{ij} = \begin{cases} 1/d(v_i), & (v_i, v_i) \in E \text{ or } (v_j, v_i) \in E \\ 0, & i = j, \\ 0, & otherwise \end{cases} \quad (5)$$

and

$$D_{ij} = \begin{cases} 1, & i = j, \\ 0, & otherwise \end{cases} \quad (6)$$

In the limit $\Delta t \to 0$ this becomes

$$\frac{d}{dt} f(t) = \alpha t(\mathbf{H} - \mathbf{D}) f(t) \quad (7)$$

Solving this differential equation,

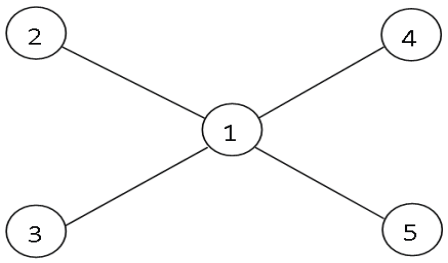
$$f(1) = e^{\alpha(\mathbf{H}-\mathbf{D})} f(0) \quad (8)$$

where $d(v)$ denotes the degree of the node v, and $e^{\alpha(\mathbf{H}-\mathbf{D})}$ could be extended as

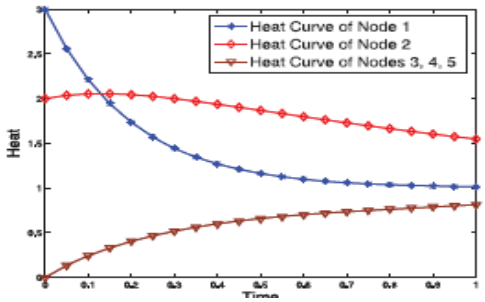
$$e^{\alpha(\mathbf{H}-\mathbf{D})} = \mathbf{I} + \alpha(\mathbf{H} - \mathbf{D}) + \frac{\alpha^2}{2!}(\mathbf{H} - \mathbf{D})^2 + \frac{\alpha^3}{3!}(\mathbf{H} - \mathbf{D})^3 + \ldots\ldots \quad (9)$$

The matrix $e^{\alpha(\mathbf{H}-\mathbf{D})}$ is called the diffusion kernel in the sense that the heat diffusion process continues infinitely many times from the initial heat diffusion.

In order to interpret (8) and the heat diffusion process more intuitively, we construct a small undirected graph with only five nodes as showed in Figure 1.



(a)



(b)

Figure 1.Simple heat diffusion example on an undirected graph. (a)Example (b) Curve of heat change with time

Initially, at time zero, suppose node 1 is given 3 units of heat, and node 2 is given 2 units of heat; then the vector f (0) equals $[3,2,0,0,0]^T$. The entries in matrices H and D will be

$$H = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 \\ \frac{1}{4} & 0 & 0 & 0 & 0 \\ \frac{1}{4} & 0 & 0 & 0 & 0 \\ \frac{1}{4} & 0 & 0 & 0 & 0 \\ \frac{1}{4} & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$D = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The entries in the matrix H – D will be

$$H - D = \begin{bmatrix} -1 & 1 & 1 & 1 & 1 \\ \frac{1}{4} & -1 & 0 & 0 & 0 \\ \frac{1}{4} & 0 & -1 & 0 & 0 \\ \frac{1}{4} & 0 & 0 & -1 & 0 \\ \frac{1}{4} & 0 & 0 & 0 & -1 \end{bmatrix}$$

Without loss of generality, we set the thermal conductivity α= 1, and vary time t from 0 to 1 with a step of 0.05. The curve for the amount of heat at each node with time is shown in Figure 1b. We can see that, as time passes, the heat sources nodes 1 and 2 will diffuse their heat to nodes 3, 4, and 5. The heat of nodes 3, 4, and 5 will increase respectively, and the trends of their heat curves are the same since these three nodes are symmetric in this graph. This indicates that if a node has more paths connected to the heat source, it will potentially obtain more heat. This is a perfect property for recommending relevant nodes on a graph.

3.3. Diffusion on directed graphs

Consider a directed graph G=(V,E,W), where V is the vertex set, $V=\{v_1,v_2,\ldots v_n\}$. W ={ w_{ij} } where w_{ij} is the probability that edge (v_i,v_j) exists} or the weight that is associated with this edge. E={v_i,v_j} there is an edge from v_i to v_j and $w_{i,j} > 0$} is the set of all edges

On a directed graph G=(V,E,W), in the pipe (v_i,v_j), heat flows only from v_i to v_j. Suppose at time t, each node vi receives RH =RH (i,j,t, Δt) amount of heat from vj during a period of Δt. We make three assumptions: 1) RH should be proportional to the time period Δt; 2) RH should be proportional to the heat at node v_j; and 3) RH is zero if there is no link from v_j to vi. As a result, v_i will receive $\sum_{j:(vj,vi)\in E} \sigma_j f_j(t)\Delta t$ amount of heat from all its neighbors that point to it.

At the same time, node vi diffuses DH(i,t, Δ t) amount of heat to its subsequent nodes. We assume that

- The heat DH(i,t, Δ t) should be proportional to the time period Δ t.
- The heat DH(i,t, Δ t) should be proportional to the heat at node vi.
- Each node has the same ability to diffuse heat.
- The DH(i,t, Δ t) should be proportional to the weight assigned between node vi and its subsequent nodes. As a result, node vi will diffuse $\alpha w_{ij} f_i(t)\Delta t \Big/ \sum_{k:(i,k)\in E} w_{ik}$ amount of heat to each of its subsequent nodes v_j, and each v_j should receive $\alpha w_{ij} f_i(t)\Delta t \Big/ \sum_{k:(i,k)\in E} w_{ik}$ amount of heat from node v_i.

Therefore, $\sigma_j = \alpha w_{ji} \Big/ \sum_{k:(j,k)\in E1} w_{jk}$

In the case that the outdegree of node v_i equals zero, we assume that this node will not diffuse heat to others. To sum up, the heat difference at node v_i between time $t + \Delta t$ and t will be equal to the sum of the heat that it receives, deducted by what it diffuses. This is formulated as

$$\frac{f_i(t+\Delta t) - f_i(t)}{\Delta t} = \alpha \left(-\tau_i f_i(t) + \sum_{j:(v_j,v_i)\in E)} \frac{w_{ij}}{\sum_{k:(j,k)\in E} w_{jk}} f_j(t) \right)$$

where τ_i is a flag used to identify whether the node v_i has any outlinks. Solving it, we obtain

$$f(1) = e^{\alpha(\mathbf{H}-\mathbf{D})} f(0) \qquad (10)$$

where

$$H_{ij} = \begin{cases} \dfrac{w_{ij}}{\sum_{k:(j,k)\in E} w_{jk}}, & (v_i,v_i)\in E \text{ or } (v_j,v_i)\in E \\ 0, & i = j, \\ 0, & otherwise \end{cases} \qquad (11)$$

$$D_{ij} = \begin{cases} \tau_i, & i = j, \\ 0, & otherwise \end{cases} \qquad (12)$$

This heat diffusion model in directed graph is used in the existing system for query suggestion. We are proposing a modification in the diffusion kernel of the basic heat diffusion equation based on the transition probability of similarity information on Web graphs.

3.4. Modification of the heat diffusion model in directed graphs

We propose a modification of the existing heat diffusion model using transition probability of information diffusion. In order to interpret the heat diffusion and the modification more intuitively, we construct an undirected query- URL bipartite graph, Bql = (V_{ql},E_{ql}), where V_{ql} = Q U L , Q = {q_1,q_2,q_3….q_n} and L = {l_1,l_2,,l_3,……l_p}. Eql = {(q_i,l_j)} | there is an edge from q_i to l_j} is the set of all edges. The edge (q_j,l_k) exists if and only if a user u_i clicked a URL l_k after issuing a query q_j. . See Fig. 1a for an example. The values on the edges in Fig. 1a specify how many times a query is clicked on a URL.

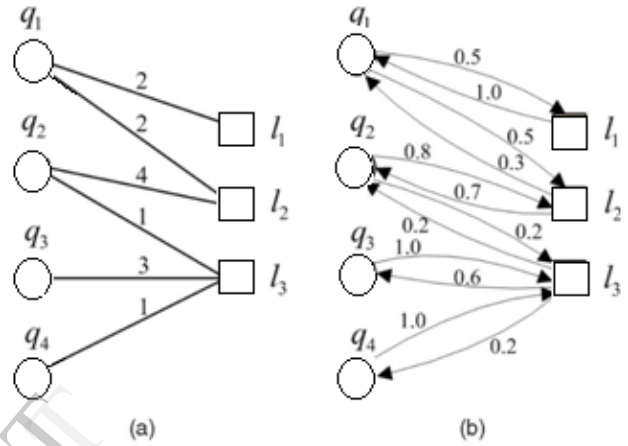


Figure 2. Graph construction for query suggestion. (a) Query-URL bipartite graph. (b) Converted query-URL bipartite graph.

This bipartite graph is converted into Fig. 1b. In this converted graph, every undirected edge in the original bipartite graph is converted into two directed edges. The weight on a directed query-URL edge is normalized by the number of times that the query is issued, while the weight on a directed URL-query edge is normalized by the number of times that the URL is clicked. From Figure 1, we see that every query is connected with a number of URLs, on which the users clicked when submitting this query to the search engine.

In the existing approach, in order to generate query suggestion, the heat flow between every node in the graph is considered. That is the heat values of the query nodes and URL nodes is calculated and is used in the heat diffusion process. This means the matrices H and D contains the values of all the nodes in the graph. But in the case of query suggestion, we don't have to calculate the heat values of all the nodes in the query – URL bipartite graph. Instead we

have to concentrate only in the heat flow between query nodes. In order to do this, the heat diffusion kernel of the heat diffusion equation is modified. The modification is on the calculation of the matrix H and it is as follows.

$$H_{ij} = \begin{cases} \sum_{k \in URL} \left(w_{jk} \Big/ \sum_{l:(j,l)\in E} w_{jl} \right) * \left(w_{ki} \Big/ \sum_{l:(k,l)\in E} w_{kl} \right), & v_i, v_j \in Q, \\ 0, \quad i = j, \\ 0, \quad otherwise \end{cases}$$

In the modified equation, we are considering only the heat flow between two query nodes and not on the URL nodes.

From the figure 2(b), the matrix H formed by the original equation is a 7×7 matrix. But according to our modified equation it will be 4×4 matrix because the graph contains four query nodes.

3.5. Query suggestion algorithm.

1. A bipartite graph $G = (V+ \cup V*, E)$ consists of query set $V+$ and URL set $V*$.
2: Given a query q in $V+$, a sub graph is constructed by using depth-first search in G. The search ends when the number of queries is larger than a predefined number.
3: As described above, set $\alpha = 1$, and without loss of generality, set the initial heat value of query equal to 1 (the choice of initial heat value will not affect the suggestion results). Start the diffusion process using

$f(1) = e^{\alpha(\mathbf{H}-\mathbf{D})} f(0)$.

4: Suggest the Top-K queries with the largest heat values in vector f(1) as the suggestions.

3.6. Complexity analysis

When the graph is very large, the computation of $e^{\alpha(\mathbf{H}-\mathbf{D})}$ is very difficult. We choose its discrete approximation to compute the heat diffusion equation

$f(1) = \left(I + \frac{\alpha}{p}(H - D) \right)^P f(0)$, where p is a positive integer. Thus supposing a graph is connected by M edges (relationships between nodes), the complexity of executing the heat diffusion process is *O(PM)*, which represents the number of iterations *P* multiplied by the number of edges *M* in a graph. In most cases, *P*=10 is enough for approximating the heat diffusion equation. The complexity *O(PM)* shows that our heat diffusion algorithm enjoys very good performance in scalability since it is linear with respect to the number of edges in the graph.

However, since the size of Web information is very large, the graph built upon the Web information can become extremely large. Then, the complexity *O(PM)* is also too high, and the algorithm becomes time consuming and inefficient to get a solution. To overcome this difficulty, we first extract a subgraph starting from the heat sources. Given the heat sources, the subgraph is constructed by using depth-first search in the original graph. The search stops when the number of nodes is larger than a predefined number. Then, the diffusion processes will be performed on this subgraph efficiently and effectively. Generally, it will not decrease the qualities of the heat diffusion processes since the nodes too far away from the heat sources are normally not related to the sources.

4. Conclusion and future scope

In this paper, a general algorithm which is based on heat diffusion is presented. This is a general framework which can be used in many recommendation tasks, such as query suggestions, image recommendations etc. The various contents on the Web is modeled in to graphs and after generating the graph we are applying the heat diffusion equation on these graphs to get suggestions.

The suggestions made by the heat diffusion model can also be used in advertisement field when customers tender for query terms. Since we are generating heat values for all the URLs, it is easy to understand that, for a given input query, after the diffusion process, the heat values of URLs stand for the relatedness to the original query, which can also be used as the ranking of these URLs. In the future, it is planned to compare this ranking method with other previous Web search results ranking approaches.

Since this model is quite general, it can be applied it to more complicated graphs and applications, such as Social Recommendation problem. In order to do that, a new algorithm must be developed. Nowadays, as the explosive growth of Web 2.0 applications, social-based applications are common on the Web. Social recommendation, which produces recommendations by using users' social network information, is becoming to be an crucial feature for the next generation of Web applications.

5. References

[1] K. Church and B. Thiesson. The wild thing! In Proceedings of the ACL 2005 on Interactive poster and demonstration sessions, pages 93–96, 2005.

[2] S. Cucerzan and E. Brill. Spelling correction as an iterative process that exploits the collective knowledgeof web users. In Proceedings of EMNLP 2004, pages 293–300, 2004.

[3] P.A. Chirita, C.S. Firan, and W. Nejdl, "Personalized Query Expansion for the Web," SIGIR '07: Proc. 30th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 7- 14, 2007.

[4] N. Craswell and M. Szummer. Random walks on the click graph. In SIGIR '07, pages 239–246,2007.

[5] Q. Mei, D. Zhou, and K. Church. Query suggestion using hitting time. In CIKM '08, pages 469–478, 2008.

[6] Y.-H. Yang, P.-T. Wu, C.-W. Lee, K.-H. Lin, W.H. Hsu, and H. Chen, "ContextSeer: Context Search and Recommendation at Query Time for Shared Consumer Photos," Proc. 16th ACM Int'l Conf. Multimedia, pp. 199-208, 2008."

[7] K. Church and B. Thiesson. The wild thing! In Proceedings of the ACL 2005 on Interactive poster and demonstration sessions, pages 93–96, 2005.

[5] Q. Mei, D. Zhou, and K. Church. Query suggestion using hitting time. In CIKM '08, pages 469–478, 2008.

[9] J.-T. Sun, D. Shen, H.-J. Zeng, Q. Yang, Y. Lu, and Z. Chen, "Web- Page Summarization Using Clickthrough Data," SIGIR '05: Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 194-201, 2005.

[11] R. Jones, B. Rey, O. Madani, and W. Greiner. Generating query substitutions. In WWW '06 pages 387–396, 2006.

[12] M. Mitra, A. Singhal, and C. Buckley. Improving automatic query expansion. In SIGIR '98, pages 206–214, 1998.

[13] K. Collins-Thompson and J. Callan. Query expansion using random walk models. In CIKM '05, pages 704–711.

[14] G. Kumaran and J. Allan. A case for shorter queries, and helping users create them. In Human Language Technologies 2007, pages 220–227, April 2007.

[15] Y. Song and L. He. Optimal rare query suggestion with implicit user feedback. In WWW '10, pages 901–910.

[16] P. Boldi, F. Bonchi, C. Castillo, D. Donato, A. Gionis, and S. Vigna. The query-flow graph: model and applications. In Procs of CIKM, 2008.

[20] X. Wang and C. Zhai. Mining term association patterns from search logs for effective query reformulation. In CIKM'08, pages 479–488, 2008.

[21] Zhang, Z., and Nasraoui, O. Mining search engine query logs for query recommendation. In WWW '06:Proceedings of the 15th international conference on World Wide Web (New York, NY, USA, 2006), ACM, pp. 1039-1040.

[22] Fonseca, B. M., Golgher, P. B., de Moura, E. S., and Ziviani, N. Using association rules to discover search engines related queries. In LA-WEB '03: Proceedings of the First Latin American Web (Washigton, DC, USA, 2003), IEEE Computer Society.

[23] Baeza-Yates, R. A., Hurtado, C. A., and Mendoza, M. Query recommendation using query logs in search engines. In EDBT Workshops (2004), vol. 3268 of LNCS, Springer, pp. 588-596.

[24] Wen, J.-R., Nie, J.-Y., and Zhang, H.-J. Clustering user queries of a search engine. In WWW '01: Proceedings of the 10th international conference on World Wide Web (New York, NY, USA, 2001), ACM, pp. 162-168.

[25] Zhang, Z., and Nasraoui, O. Mining search engine query logs for query recommendation. In WWW '06:Proceedings of the 15th international conference on World Wide Web (New York, NY, USA, 2006), ACM, pp. 1039-1040.

[26] Antonellis, I., Garcia-Molina, H., and Chang, C.-C. Simrank++: Query rewriting through link analysis of the click graph. In Proceedings of VLDB (Dec 2008), pp. 408-421.

[27] White, R. W., Bilenko, M., and Cucerzan, S. Studying the use of popular destinations to enhance web search interaction. In SIGIR '07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval (New York, NY, USA, 2007).

[28] White, R. W., Bilenko, M., and Cucerzan, S. Leveraging popular destinations to enhance web search interaction. ACM Trans. Web 2, 3 (July 2008), 1-30.

[29] Jeh, G., and Widom, J. Simrank: a measure of structural-context similarity. In KDD '02: Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining (New York, NY, USA, 2002), ACM Press, pp. 538-543.