Leader-page Resources in World Wide Web

D.Ravi Shankar*
International Institute of Information Technology
Hyderabad, India
d.ravishankar@gmail.com

Pradeep Beerla*
International Institute of Information Technology
Hyderabad, India
pradeep.beerla@gmail.com

P.Krishna Reddy
International Institute of Information Technology
Hyderabad, India
pkreddy@iiit.ac.in

Abstract

Ranking the search results is an important research problem in WWW. HITS, PageRank and variations of these algorithms are widely used approaches for ranking. In this paper we proposed a new ranking algorithm to rank the search results by introducing the concept of "leader-page". The notion of "leader-page" is defined by extending the concept of "leader" from leadership theory. In leadership theory, a leader is defined as a person, who has more contacts with the other members of the group, both initiates and receives communication, and whose characteristics are the most similar to the group's own characteristics. Similarly, in WWW, the proposed approach identifies leader-pages and assigns "leadership score" to them based on several kinds of cyclic and similarity relationships it establishes with other web pages. Given a set of key words (search query), the proposed approach ranks the related pages based on the corresponding "leadership score". The experiment results show that the proposed approach gives high leadership score to resourceful pages as compared to the results of the HITS algorithm and the Google search engine.

1. Introduction

The WWW is the single largest global repository of information and human knowledge. It continues to grow at a remarkable pace with contributions from allover the world. The knowledge discovered through navigation of this complex heterogeneous collection of text (content) and hyperlinks (that lend it a structure) [15] is enormously benefiting the mankind. However, owing to the hugeness and the diversity of the web, users are drowning in information and are facing the problem of information overload. Several approaches have been proposed in the literature to find relevant information on the web. Normally, a user tries to find the relevant information in WWW by giving a query (set of keywords) to the search engine. The number of results listed by the available search engines is usually very high. It is very difficult for the user to scan through all the listed results and identify the relevant information. So, the ranking of searched results is very important. Several efforts are being made to improve the ranking methodology by extending the concepts from social networks [25], citation analysis [30], graph theory [11, 12] and bibliometrics [26]. HITS [1], PageRank [3, 4] and variations of these algorithms are being widely used approaches for ranking the search results.

In this paper we proposed a new ranking algorithm for ranking the search results by introducing the concept of "leader-page". The notion of "leader-page" is defined by extending the concept of "leader" from the leadership theory [21]. In leadership theory, a leader is defined as a person, who establishes several kinds of relationships with many other members of the community and becomes a representative of the whole community. Similarly, in WWW, a notion of leader-page is defined and is assigned a corresponding "leadership score" based on several kinds of cyclic and similarity relationships it establishes with the other web pages in the corresponding web community [2, 9, 10, 11, 14]. Given a set of key words (search query – broad topic queries [16]), the proposed approach gives "leadership score" to the related pages. We have conducted experiments on various queries. The experiment results on the queries "Child Labor", "Aids", "Globalization" and "Jaguar" show that the proposed approach gives high leadership score to resourceful pages.

We now explain the proposed approach through an example. When the search query "Child Labor" was given to the Google's search engine [32], the top result displayed was

International Conference on Management of Data
COMAD 2005b, Hyderabad, India, December 20–22, 2005

**Currently with Tata Consultancy Services, Hyderabad(India).*

www.historyplace.com/unitedstates/ChildLabor. This site is a collection of photographs regarding "Child Labor". By applying HITS algorithm the top authority is *www.ChildLaborlaws.org*. This URL contains some related information about Child Labor. The proposed approach has identified w*ww.stopchildLabor.org* as the top leader-page. This web-page contains a lot of information regarding Child Labor. It also contains links to several related resources. By comparing these results with the results of Google [5, 32] and HITS algorithm, we found that the proposed approach assigns high leadership score to the pages, which are more resourceful. This can be attributed to many cyclic and similarity relationships the leader page established with other pages in the web.

The paper is organized as follows. In section 2, we discuss the related work. In section 3, we explain the notion of "leader" in the leadership theory. In section 4, we explain how we extended the concept of leader to define a "leader-page" in WWW. In section 5, for a given set of key words, we present the algorithm to extract "leader-pages" in WWW. In section 6, we present the experiment results. In the last section, we present the summary and conclusions.

2. Related Work

Most of the search engines perform both link and text [7, 8] analysis to improve the quality of search results. Many researchers have proposed schemes based on citation analysis, bibliometrics [1, 10], social networks and graph theory [9, 11, 12, 14] to improve the ranking methodology and the search performance.

The Hyper-link-Induced Topic Search (HITS) [1] and PageRank [3, 4] algorithms are the widely used algorithms in search engines to rank search results, which exploit the connectivity information among web pages.

The HITS algorithm is based on the following intuition: a page that many pages point to is a good authority and the page that points to many others is a good hub. In HITS mutual reinforcement occurs between hubs and authorities: a good hubs point to good authorities and a good authority is pointed to by good hubs. The HITS algorithm repeatedly updates authority and hub scores so that pages with high authority scores are expected to have relevant content and pages with high hub scores are expected to contain-links to pages with relevant content. Projection and downsizing methods [17] are some of the recent improvements regarding HITS algorithm.

PageRank [3, 4] algorithm relies on the link structure of the web as an indicator of an individual page's value. PageRank is a numeric value that represents how important a page is on the web. When one page has a URL to another page, it is effectively casting a vote for the other page. The in-links to a page indicate the importance of the page. PageRank algorithm calculates a page's importance from the number of in-links it has. The importance of each in-link is also taken into account. Extrapolation methods, distributed PageRank and topic sensitive PageRank [18] are some of the recent improvements regarding PageRank algorithm.

The proposed approach differs from the preceding approaches as we have extended the notion of "leader" from the leadership theory and made an effort to propose a new ranking methodology.

3. Leader and leadership theory

In this section we discuss the notion of "leader" according to the leadership theory.

In any society, the phenomenon of leadership has a great prominence. A leader is interpreted as a person who sets direction in an effort and influences other members of the society to follow that direction. The influence on the other members of the society depends on a variety of factors. The leaders have strong mutual relationships with other members of the society. The leaders are the role models of the society. A society can be analyzed by studying the leadership phenomenon in the society.

A scan of various theories of leadership can help to comprehend the leadership phenomenon. The phenomenon of leadership has been studied since Aristotle [21]. In the beginning, leadership theories focused on the personality traits, the leader behavior, the group process and the context of leadership. But the recent studies have identified leadership as a powerful mutual relationship, influence on others and initiation of a structure. Most of the general definitions of leadership use combination of these concepts. In the leadership literature [27] several theories have been proposed to understand the leadership phenomenon in a society.

Trait theory [19] is one of the earliest theories on leadership. This theory of leadership focuses on the traits of the leader that make him a leader. This theory assumes that leaders are born naturally. Later on, the focus has shifted towards the behavior of the leader. Studies have led to the notion of "Charismatic leadership". A charismatic leader continually assesses the environment. He/she communicates with other people, and builds trust and commitment. Finally he/she is the role model of the whole community. Although no universally accepted set of features define charismatic leader, the charisma component is considered as important factor. However, features of an optimal leadership behavior are provided by the Managerial grid theory.

The leadership studies shifted from "what a leader has" to "what a leader does". The leader-member exchange theory [20] emphasizes the dyadic relationship between the leaders and the other members of the

community. The notion of community is inherent in the concept of leadership. This theory divides the society into in-group and out-group. The in-group members have a common bond and value system. They are similar and have more mutual communications with the leader. Every collection of people does not form an in-group or a community. An in-group is formed only when people have enough shared features and strong mutual relationships among themselves. The out-group is the exterior of the in-group. The out-group has fewer relations with the leader. The transactional and the transformational approaches of leadership stress on the mutual relationship a leader has with other members of the community. The leaders become representative of the whole community. The servant leadership theory, Fielder's Contingency model, Situational leadership, Path-goal leadership and Sashkin's leadership provide the insights for this model of a leader [21].

George C Homans [22] has defined leader as a person who interacts the most with other members of the community, both initiates and receives the communication, has more social contacts within the community and whose actions and characteristics are the most similar to the community's own actions and characteristics. Leader is representative of the community. Study of leadership phenomenon helps in analyzing the society.

4. Leader-page notion in WWW

By extending the notion of the leader from the leadership theory, in this section, we define the notion of a "Leader-page" in the WWW. In this paper P0, P1… denote the web pages.

By considering the web as a complex social network or society [13], where web pages are nodes and the relationship between the pages is conveyed by the existence of links and similarities [6], we have made an effort to propose a new ranking algorithm through the notion of "leader-page". In WWW, several kinds of relationships can exist between any two web pages. For example, the web pages Pi and Pj can be related based on the existence of hyperlinks, cocitation relationship, coupling relationship, and other similarities. We call a web page as a "leader-page" if it establishes several link-based cyclic and similarity relationships with other pages. A page is called a good leader-page if it establishes relatively large number of link-based cyclic and similarity-based relationships with other pages.

Every page Pi in the web can be assigned a leadership score L[Pi]. The leadership score is determined based on the direct and indirect-cyclic-based relationships and similarity based on coupling and cocitation-based relationships with other pages of the web. L(Pi) is the weighted sum of leadership scores of all the other pages which participate in preceding relationships with Pi. In WWW, these relationships are manifested in the following ways.

I. Direct link-based cyclic relationship (DLC)
II. Indirect link-based cyclic relationship (ILC)
III. Cocitation based similarity relationship(CociteS)
IV. Coupling based similarity relationship(CoupleS)

4.1 Direct link-based cyclic (DLC) relationship

For a pair of web pages Pi and Pj, they participate in a DLC relationship if the web page Pi establishes link to Pj and the web page Pj establishes link to Pi. If Pi establishes many such DLC relationships with other web pages, then Pi's potential of becoming a leader increases. In Figure 1, P0 establishes a DLC relationship with P1, P2 … Pn.

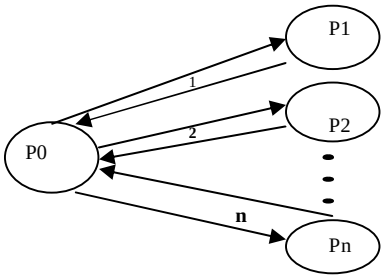

Figure 1. A DLC relationship

4.2 Indirect link-based cyclic (ILC) relationship

The ILC relationship is identified among the web pages Pi, Pj and Pk if Pi establishes link with Pj, Pj establishes link with Pk and Pk establishes link with Pi. If a web page Pi establishes many ILC relationships, Pi's potential of becoming a leader increases. Figure 2 depicts how P0 establishes several ILC relationships.

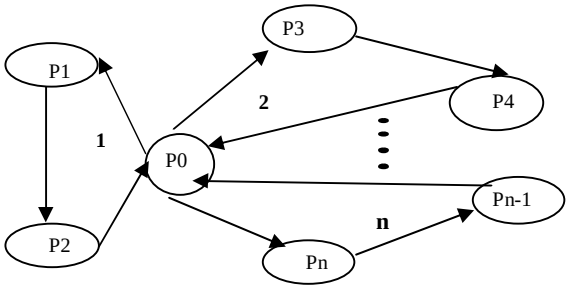

Fig. 2. Depiction of how web page P0 establishes ILC relationships with other pages.

Note that we can say DLC as 1-cycle relationship and ILC as 2-cycle relationship. Similarly, we can define more relationships by increasing the cycle length. However, as the cycle length increases the complexity to

extract such cycles also increases and at the same time the influence of such relationships on the leadership scores decreases. So, we considered link-based cyclic relationships of cycle length one and two only. We are going to investigate the effect of the relationships having cycle length greater then two as a part of future work. Also, both DLC and ILC can be presented as a single relationship. However, we considered them as different relationships as both have different degrees of influence on leadership score. They also have different degrees of complexity for extraction.

4.3 Cocitation-based Similarity (CociteS) relationship

The "Leader-page" has the property of being most similar to other pages of the web community. The CociteS relationship exists between two web pages Pi and Pj if a set of web pages have reasonable number of links to both Pi and Pj. This implies that the pages Pi and Pj are similar and this is attributed to cocitation by other web pages. If Pi participates in several CociteS relationships with other pages, Pi can be considered as potential leader. In Figure 3, P0 establishes a CociteS relationship with P1, P2 … Pn.

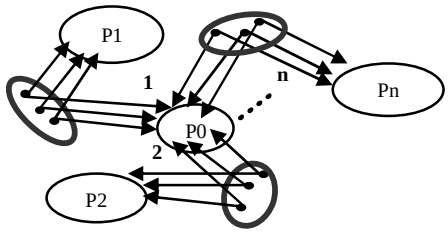

Figure 3: A CociteS relationship. The dots in thick circles indicate web pages.

4.4 Coupling-based similarity (CoupleS) relationship

Similar to CociteS relationship, the relationship between two web pages Pi and Pj can also be identified through CoupleS relationship. Two web pages are related through coupling if they have common out-links (or children). When the intersection of the set of out-links of Pi and Pj is above certain threshold, they exhibit the CoupleS relationship. If Pi participates in many CoupleS relationships with other pages, its potential of becoming a leader increases. In Fig. 4, P0 establishes a CoupleS relationship with P1, P2 … Pn.

By incorporating preceding four formulations, the leadership score L[Pi] for Pi is given by the following formula.

$L[Pi] = k_{dl}(DL(Pi)) + k_{indl}(INDL(Pi)) + k_{coct}(COCT(Pi)) + k_{coup}(COUP(Pi))$, where

DL(Pi)=sum of leadership scores of all pages having DLC relationship with Pi.

INDL(Pi)= sum of leadership scores of all pages having ILC relationship with Pi.

COCT(Pi)=sum of leadership scores of all pages having CociteS relationship with Pi.

COUP(Pi)=sum of leadership scores of all pages having CoupleS relationship with Pi.

Here, k_{dl}, k_{indl}, k_{coct}, and k_{coup} are the parameters that determine the weights of corresponding of DLC, ILC, CociteS, and CoupleS relationships.

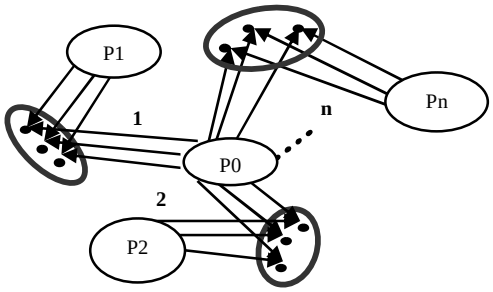

Figure 4. A CoupleS relationship. The dots in thick circles indicate web pages.

5. Leader-page extraction algorithm

For a given search query, the Leader-page extraction algorithm extracts the corresponding "leader-pages" from WWW. The process of extraction of leader-pages is similar to the extraction of Hub and Authority web pages in HITS [1]. For the specific search query, we build a focused sub-graph. Next, we apply the leader-page extraction algorithm to calculate leadership scores to all the pages in the focused sub-graph. The pages with high leadership score are considered as leaders for the search query. The implementation of the leader-page extraction algorithm is done in two phases. One is building focused sub-graph (S) and the other is calculation of leadership scores of all the pages in S.

5.1 Building the focused sub-graph

The process of building S for a given set of keywords is same as the algorithm followed in HITS. The search query is given to a search engine. By taking a reasonable number of top pages in the output list corresponding root-set is formed. For each web page in the root-set, corresponding parents and Children are extracted. The parents and children of all the pages and pages of root set form a base-set. Several pre-processing techniques are applied on the base-set. All the intrinsic links (links between pages with the same domain name) are removed

keeping only the traverse links. Also, URLs of top search engines and online repositories were removed. This is called the focused sub-graph of WWW corresponding to search query.

5.2 Calculating the leadership scores

The algorithm to calculate the leadership scores for the web pages in S is given below. L[Pi] denotes the leadership score of the page Pi and L denotes the leadership score vector for all the pages in S. The leadership scores of all the pages in S are initialized to one. For each web page Pi in S,

(i) If Pi forms a DLC relationship with Pj in S, then L[Pi]= L[Pi]+k_{dl}(L[Pj]).
(ii) If Pi forms an ILC relationship with Pj and Pk in S, then L[Pj]= L[Pi]+k_{indl}(L[Pj]+L[Pk]).
(iii) If Pi forms a CociteS relationship with another Pj in S, then L[Pi]= L[Pi]+k_{coct}(L[Pj]).
(iv)If Pi forms a CoupleS relationship with Pj, then L[Pi]= L[Pj]+k_{coup}(L[Pj]).

After updating L[Pi], the leadership score vector is normalized. The leader-page extraction algorithm repeatedly updates and normalizes the leadership scores. This whole process is iterated until leadership vector converges to a fixed value (a small difference in successive values). The web pages are sorted based on corresponding leadership scores. The values for parameters k_{dl}, k_{indl}, k_{coct} and k_{coup} should be selected based on the corresponding influence on the leadership score. The range for these values is (0,1) to prevent the high rise in the leadership scores. The pseudo-code for calculating the leadership scores is given in Figure 5. The complexity of algorithm comes to O(m*n*n) where m is number of iterations and n is number of pages.

Input: Leadership score vector L, Focused Sub-graph S
Output: Converged leadership score vector L
Algorithm: CALCULATE_LEADERSHIP_SCORE
/*Construct initial leadership score vector*/
1. Let S' be the set of pages in S. Suppose Pi, Pj and Pk ª S'.
2. For all pages Pi ª S', initialize L[Pi] to 1.
3. While L is not *converged* repeat the following steps.
Begin
4. For all Pi ª S' repeat the steps 4.1 to 4.4.
4.1 If (Pi establishes a DLC relationship with Pj) then
L[Pi] = L[Pi]+ k_{dl}(L[Pj])
4.2 If (Pi establishes an ILC relationship with Pj and Pk) then
L[Pi] = L[Pi]+ k_{indl}(L[Pj]+L[Pk])
4.3 If (Pi establishes a CociteS relationship with Pj) then
L[Pi] = L[Pi]+ k_{coct}(L[Pj])
4.4 If (Pi establishes a CoupleS relationship with Pj) then
L[Pi] = L[Pi]+ k_{coup}(L[Pj])
/*Normalization*/
5. For all Pi ª S'
L[Pi]= L[Pi] / c (c is a parameter) where $(L[Pi]/c)^2=1$.
End

Figure 5: Pseudo-code for calculating the Leadership Scores

6. Experiments and results

In this section, we compare the results of the proposed approach with the results of Google and HITS algorithm for a selected set of keywords. We have performed experiments on several keywords. Here we present the results for the queries "Child Labor", "Aids", "Globalization" and "Jaguar".

We have computed leadership scores as follows. We gave these queries as input to Yahoo [31] search engine and selected the top 50 pages as the root-set. We extracted all the pages in the root-set, expanded it to the base-set. We then applied preprocessing techniques to get the focused sub-graph. With leader-score calculation algorithm we calculated leadership scores for all the pages in S. It took about 20 iterations to for the leadership vector to converge. We have conducted experiments by selecting the parameter values as k_{dl} = 0.2, k_{indl} = 0.2, k_{coct} = 0.1 and k_{coup} = 0.1. We have given more weight to DLC and ILC relationships because we believe that the influence of these relationships on the formation of leader-page is high to other relationships. The leader-pages are ranked on the value of corresponding leadership scores.

With HITS algorithm the hub and authority scores for all the pages in S are calculated.

We have extracted the top Google's results by giving the same set of keywords to Google search engine.

Figure 6 displays the results for the query "Child Labor". The top URL identified using Google's search is *www.historyplace.com/unitedstates/ChildLabor*. This contains a collection of photographs related to "Child Labor". The HITS algorithm gave *www.ChildLaborlaws.org* as the top authority. This URL contains some related information about "Child Labor". The top leader-page identified is *www.stopChildLabor.org*. It contains a lot of information regarding Child Labor. It also contains links to several other related web pages of "Child Labor". It was observed that for the query Child Labor, the proposed approach identified URLs which are more resourceful over Google and HITS. Figure 7 shows top hub pages extracted by HITS. It can be observed that the hub pages do not find place in the results of Google and the proposed approach.

Figure 8 displays the results for the query "Aids". The top Google's result is *www.Aids.org*. This is the web-site for a non-profit private organization. The top authority is *www.unAids.org*. This is the official web-site for the United Nations Aids community. The top leader-page identified is *www.cdcnpin.org*. The top results from Google and HITS are also related to Aids, but the top leader-page is the US's largest collection of information and resources on aids and related diseases. This URL is

not listed in the top results of the Google's results for Aids. We have observed that several cyclic and similarity relationships exist between the URLs *www.cdcnpin.org*, *www.aegis.org*, *www.unAids.org*, *www.cdc.gov* etc. The leader-page extraction algorithm was able to use this information and identify more resourceful pages. Figure 9 shows the corresponding hub pages.

URL	Authority Rank	Google Rank	Leadership Rank
www.stopChildLabor.org	5	4	1
www.global-unions.org	-	-	2
www.ilo.org	3	8	3
www.iccle.org	-	-	4
www.sadashivan.com	15	17	5
www.unicef.org	4	2	6
www.globalmarch.org	-	5	7
www.icftu.org	11	-	8
www.historyplace.com/unitedstates/ChildLabor	21	1	-
www.ChildLaborlaws.org	1	-	-

Figure 6: Results for the query "CHILD LABOR"

URL	Hub Rank
www.yahoo.co.uk	1
www.bbc.co.uk	2
www.altavista.digital.com	3
www.standards.dfee.gov.uk	4
www.alltheweb.com	5
www.learningcurve.pro.gov.uk	6
www.northernlight.com	7
www.excite.co.uk	8
www.service.bfast.com	9
www.spartacus.schoolnet.co.uk	10

Figure 7: Hub-pages for the query "CHILD LABOR"

URL	Authority Rank	Google Rank	Leadership Rank
www.cdcnpin.org	-	-	1
www.aegis.org	7	6	2
www.Aidsinfonet.org	8	-	3
www.Aids.org	-	1	4
www.cdc.gov	19	18	5
www.theglobalfund.org	16	-	6
www.unAids.org	1	4	7
www.Aidsonline.com	13	3	8
www.Aidsaction.com	3	16	-
www.nnaapc.org	-	17	9

Figure 8: Results for the query "AIDS"

URL	Hub Rank
www.specialweb.com	1
www.refdesk.com	2
www.blackAids.org	3
www.library.uchc.edu	4
www.nurseweb.ucsf.edu	5
www.epibiostat.ucsf.edu	6
www.docguide.com	7
www.bmsvirology.com	8
www.store.yahoo.com	9
www.lando.co.za	10

Figure 9: Hub-pages for the query "AIDS"

Figure 10 displays the results for the query "Globalization". The top Google's result is *www.ifg.org* (International Forum on Globalization). The top authority is *www.polity.co.uk*. This is the web-site of a book publisher. The top leader-page identified is *www.Globalization.about.com*. This has some general information regarding Globalization. In this case top two results are the same for Google and the proposed algorithm, but in different order. However the URL *www.imf.org* is not identified in the Google's and HITS top results. International Monitory Fund (IMF) is an important organization regarding Globalization. The proposed approach was able to identify this resource. Figure 11 shows the corresponding hub results.

URL	Authority Rank	Google Rank	Leadership Rank
www.Globalization.about.com	-	2	1
www.un.org	4	-	2
www.ifg.org	7	1	3
www.globalpolicy.org	-	-	4
www.worldbank.org	10	4	5
www.Globalization.com	-	5	6
www.imf.org	-	-	7
www.polity.co.uk	1	-	-
www.prospect.org	5	-	-

Figure 10: Results for the query "GLOBALIZATION"

URL	Hub Rank
www.questia.com	1
www.en.wikipedia.org	2
www.sociology.emory.edu	3
www.polity.co.uk	4
www.questionsforthefuture.tv	5
www.globalenvision.org	6
www.prospect.org	7
www.iatp.irex.am	8
www.yahoo.com	9
www.globalgrn.org	10

Figure 11: Hub-pages for the query "GLOBALIZATION"

Figure 12 displays the results for the query "Jaguar". The top Google's result is *www.Jaguar.com*. This is the official site of Jaguar manufacturer. The top authority is *jag-lovers.com*. This has several discussion forms and archives regarding Jaguar cars. The top leader-page identified is *www.jagweb.com*. The propose approach was able to discover *www.jagweb.com*, the largest online Jaguar directory. Figure 13 shows the corresponding hub results.

We have conducted experiments for several other queries. It was found out that the proposed approach assigns high leadership score to the web pages which are more resourceful or contains lot of information. The reason can be attributed to the fact that the proposed approach uses the cyclic and similarity relationships among the pages in the web.

URL	Authority Rank	Google Rank	Leadership Rank
www.jagweb.com	-	-	1
www.Jaguar.com	-	1	2
www.apple.com/macosx/	-	3	3
www.autoseek.co.uk	7	-	4
www.mgcars.org.uk	9	-	5
www.britishcarlinks.com	12	-	6
www.jagads.com	3	-	7
www.jag-lovers.org	1	11	8
www.Jaguarcars.com	-	2	9

Figure 12: Results for the query "JAGUAR"

URL	Hub Rank
www.nature.ca	1
www.jags.org	2
www.nsrl.ttu.edu	3
www.en.wikipedia.org	4
www.netscape.com	5
www.thewildones.org	6
www.amazon.com	7
www.bbc.co.uk/nature	8
www.dir.yahoo.com	9
www.civilization.ca	10

Figure 13: Hub-pages for the query "JAGUAR"

7. Summary and Conclusions

In the literature several approaches have been proposed to improve the ranking performance of search results by extending the concepts of citation analysis, bibliometrics, graph theory and social networks. HITS, PageRank and variations of these approaches are being widely used for ranking. In this paper we proposed a new ranking algorithm by extending the notion of "leader" from the leadership theory to define a leader-page in the WWW. In WWW, we identified leader-pages based on the cyclic and similarity relationships they establish with other pages of the web community. The leadership extraction algorithm is applied on several search queries. The experiment results show that search results ranked by the proposed approach are more resourceful over the results of Google's search engine and HITS algorithm.

One of the methods to analyze communities in society is by understanding the evolution of leaders. The leaders in the community are formed by the mutual relationships with other members of the community. Based on the notion of leader, we investigated the existence of leader-pages in WWW. The characteristics of leader-page are that it establishes cyclic and similarity-based relationship with several other web pages and contains important information related to search query. The leadership phenomenon is not being exploited by the existing methods. We feel that a leadership phenomenon in WWW shows a promise to improve the search performance.

As a part of future work, we are planning to perform extensive experiments by applying the leader-page extraction algorithm on various kinds of search queries. We are building a search engine prototype based on the proposed approach. In this paper we have selected the fixed values for k_{dl}, k_{indl}, k_{coct} and k_{coup} and carried out experiments. As a part of future work we are going to investigate the how the different values of these parameters (with different combinations) influences the leadership scores. We are also planning to carry out experiments by incorporating the leadership factor in PageRank algorithm and investigate the differences. We also plan to look at the extent the text-based similarity can be used for the leadership resources in the WWW.

References

[1]Jon M. Kleinberg. Authoritative sources in a hyperlinked environment. *9th ACM-SIAM Symposium on Discrete Algorithms*, 1998.

[2]D. Gibson, J. Kleinberg, P. Raghavan. Inferring Web Communities from Link Topology. *9th ACM Conf on Hypertext and Hypermedia*, 1998.

[3]Junghoo Cho, Hector Garcia-Molina, Lawrence Page. Efficient crawling through URL ordering. *7th WWW Conference*, 1998.

[4]L. Page, S. Brin, R. Motwani, and T. Winograd. PageRank Citation Ranking: Bringing order to Web. *Stanford Digital Library Technologies Project*, 1998.

[5]Sergey Brin and Lawrence Page. The Anatomy of a Large-Scale Hypertextual Web Search Engine. *7th WWW Conference*, 1998.

[6]Sergey Brin. Extracting patterns and relations from the WWW. *WebDB workshop at 6th EDBT*, 1998.

[7]Soumen Chakrabarti, Byron Dom, Rakesh Agrawal, Prabhakar Raghavan. Scalable feature selection, Classification and signature generation for organizing large text databases into hierarchical topic taxonomies. *VLDB Journal*, 1998.

[8]K.Bharat and M. Henzinger. Improved algorithms for Topic Distillation in Hyperlinked environments. *21st ACM SIGIR Conference*, 1998.

[9]Ravi Kumar, Prabhakar Rhaghavan,Sridhar Rajagopalan, Andrew Tomkins. Trawling the web for emerging Cyber-cmunitites. *8th WWW Conference*, 1999.

[10]Jeffrey Dean and Monika R. Henzinger. Finding related pages in the WWW. *8th WWW Conference*, 1999.

[11]Gary William Flake, Steve Lawrence and C. Lee Giles. Efficient identification of Web Communities. *6th ACM SIGKDD*, 2000.

[12]P.Krishna Reddy, Masaru Kitsuregawa. An approach to relate the web communities through bipartite graphs. *2nd*

International Conf on Web Information Systems Engineering (WISE), IEEE Computer Society, 2001.
[13]Andrei Broder, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, Janet Wiener. Graph structure in the web. *9th WWW Conference*, 2001.
[14]P.Krishna Reddy and Masaru Kitsuregawa. An Approach to build Cyber Community Hierarchy. *Workshop on Web Analytics*, 2002
[15]Soumen Chakrabarthi. *Mining the Web: Discovering Knowledge from Hypertext Data*. Morgan Kaufmann Publishers, 2002.
[16]Soumen Chakrabarti. The Structure of Broad Topics on the Web. *11th WWW Conference*, 2002.
[17]Saeko Nomura, Satoshi Yamda, Tetsuo Hayamizu and Toru Ishida. Analysis and Improvement of HITS Algorithm for detecting web communities. *Symposium on Applications and the Internet*, 2002.
[18] Sepandar D.Kamvar, Tahel H.Pardiwala, Christoper D.Manning, Gene H.Golub. Extrapolation methods for accelerating PageRank Computation. *12th WWW Conference*, 2003.
[19]Bass, B. M. *Leadership performance beyond expectations*. New York: Academic Press, 1985.
[20]Graen G.G, Uhl-Bien. Relationship-based approach to leadership: Development of leader-member exchange (LMX) theory of leadership over 25 years: Applying a multi-level multi-domain perspective. *Leadership Quarterly*, 6(2), 219-247, 1995.
[21] Northouse, P.G. *Leadership theory and practice*. Thousand Oaks, CA: Sage Publications, 2001.
[22]George C Homans. *The Human Group*. New York Harcourt, Brace & World, 1950.
[23]M.M Kessler. Bibliographic coupling between scientific papers. *American Documentation* 14, 1963.
[24]Small H.G. Cocitation in scientific Literature: A new measure of relationship between two documents. *Journal of ASIA*, 24, no.4, 1973.
[25]Charles Kadushin. *Intro to Social Network Theory*. Brandeis University, 2004.
[26]R. Larson. Bibliometrics of the WWW: An exploratory analysis of the intellectual structure of cyberspace. *Meeting of the American Soc. Info. Sci*, 1996.
[27] Avolio, B. J. *Full leadership development: Building the vital forces in organizations*. Thousand Oaks, CA: Sage, 1999.
[28]T.H Cormen, C.E Leiserson, and R.L Rivest. *Introduction to Algorithms*. 2nd Edition, Prentice Hall Publishers, 2001.
[29]Robert J. Valenza. *Linear Algebra – An introduction to abstract Mathematics*. Springer Publications, 1993.
[30]Eugene Garfield. Citation analysis as a tool in journal evaluation. *Science* 178, 1972.
[31]Search engine *http://www.yahoo.com*, July 2005.
[32]Search engine *http://www,google.com*, July 2005.