The Importance of Link Evidence in Wikipedia

Jaap Kamps[1,2] and Marijn Koolen[1]

[1] Archives and Information Studies, University of Amsterdam,The Netherlands
[2] ISLA, University of Amsterdam,The Netherlands

Abstract. Wikipedia is one of the most popular information sources on the Web. The free encyclopedia is densely linked. The link structure in Wikipedia differs from the Web at large: internal links in Wikipedia are typically based on words naturally occurring in a page, and link to another semantically related entry. Our main aim is to find out if Wikipedia's link structure can be exploited to improve ad hoc information retrieval. We first analyse the relation between Wikipedia links and the relevance of pages. We then experiment with use of link evidence in the focused retrieval of Wikipedia content, based on the test collection of INEX 2006. Our main findings are: First, our analysis of the link structure reveals that the Wikipedia link structure is a (possibly weak) indicator of relevance. Second, our experiments on INEX ad hoc retrieval tasks reveal that if the link evidence is made sensitive to the local context we see a significant improvement of retrieval effectiveness. Hence, in contrast with earlier TREC experiments using crawled Web data, we have shown that Wikipedia's link structure can help improve the effectiveness of ad hoc retrieval.

1 Introduction

Wikipedia is a free Web-based encyclopedia, that is collaboratively edited by countless individuals around the globe [1]. As an encyclopedia, it consists of individual entries on a single subject (Wiki pages) that are densely hyperlinked to related content (using Wikilinks). Wikipedia's links are a special case of the general hyperlinks that connect the World Wide Web. On the Web, link structure has been exploited to improve information retrieval in algorithms like PageRank [2] and HITS [3]. In the simplest case of a link from a page A to a page B, we can count this as a vote by the author of page A for page B as being authoritative [3]. Pages with a high number of incoming links are considered important pages. Commercial Internet search engine companies have heralded the use of link structure as one of their key technologies.

Retrieval using Web data has been studied at TREC since TREC-8 in 1999. Despite high expectations, TREC experiments failed to establish the effectiveness of link evidence for general ad hoc retrieval [e.g., 4; 5]. As Hawking and Craswell [6, p.215] put it:

> Hyperlink and other web evidence is highly valuable for some types of search task, but not for others. Because binary judgements were employed and judges looked only at the text of the retrieved pages, the TREC-8 Small Web Task

> and the TREC-9 Main Web Task did not accurately model typical Web search.
> ...
> In prototypical TREC Ad Hoc methodology, the task presupposes a desire to read text relevant to a fairly precisely defined topic, and documents are judged on their own text content alone as either relevant or not relevant. By contrast, Web searchers typically prefer the entry page of a well-known topical site to an isolated piece of text, no matter how relevant. For example, the NASA home page would be considered a more valuable answer to the query "space exploration" than newswire articles about Jupiter probes or NASA funding cuts.

These observations led to the definition of a range of Web-centric tasks, like known-item (homepage, named-page) search and topic distillation. For these special tasks, the URL-type, anchor text and link indegree are effective to improve retrieval performance [7; 8; 9].

Our conjecture is that the links in Wikipedia are different from links between Web documents. Whereas in Web documents, an author can arbitrarily link his page to any other page, whether there is a topical relation or not, in Wikipedia, links tend to be semantic: a link from page A to page B shows that page B is semantically related to (part of) the content of page A. Arguably, there will be some fraction of links that do not denote an important topical relation between pages, and not all links will be equally meaningful in all search contexts. For example, Wikipedia "bots" may automatically insert links serving a particular purpose (think of the year links). However, there is a clear mechanism in place that results in links that are relevant to the context.[3] This immediately prompts the question: what is the importance of link evidence in such a semantically linked collection?

Given its encyclopedic content, Wikipedia is a particularly attractive resource for informational search requests.[4] Hence, in this paper, our main aim is to find out if Wikipedia's link structure can be exploited to improve the ad hoc retrieval of relevant information. Our experimental evidence is based on the INEX 2006 test collection consisting of an XML version of Wikipedia containing over 650,000 articles, and a set of 114 ad hoc topics with judgments on the passage-level [11]. Specifically, we want to know:

– Can the link degree structure of a semantically linked document collection be used as evidence for the relevance of ad hoc retrieval results?

To answer this question, we analyse the link structure of Wikipedia pages, and of Wikipedia pages relevant for a particular ad hoc retrieval topic. Furthermore, link structure can be considered in various ways: on a global level, i.e., the number of incoming links over the whole collection, or on a local level, i.e., the

[3] See, for example, `http://en.wikipedia.org/wiki/Wikipedia:Only_make_links_that_are_relevant_to_the_context`.

[4] That is, in terms of the Broder [10] taxonomy, Wikipedia seems a less suitable resource for navigational queries (with the intent to reach a particular site), or for transactional queries (with the intent to perform some web-mediated activity).

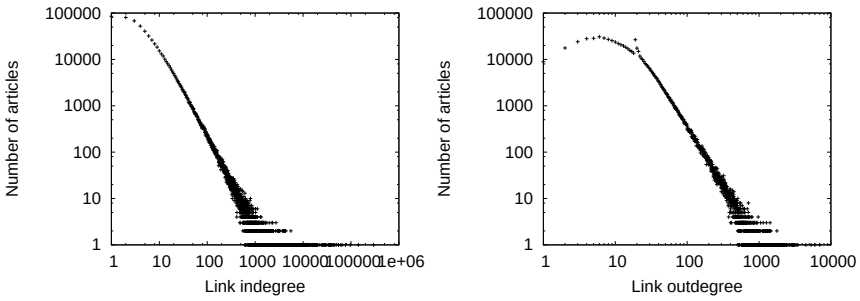

Fig. 1. Wikipedia collection link indegree (left) and outdegree (right) distribution over 659,304 pages

number of incoming links within the subset of articles retrieved as results for a given topic. So, more specifically, we want to know:

– Can global indegree, local indegree, or a combination of the two, be used effectively to rank results in general ad-hoc retrieval?

To answer this question, we investigate the effectiveness of a number of link degree priors for three different INEX ad hoc retrieval tasks.

The rest of this paper is structured as followed. Next, in Section 2, we analyse the link structure of Wikipedia, trying to establish its relation with relevance. Then, in Section 3, we detail on how link evidence can be incorporated into the scoring of a retrieval system. This is followed, in Section 4, by a range of retrieval experiments trying to establish the impact of link evidence on retrieval effectiveness. We end in Section 5 by drawing conclusions and discussing our findings.

2 Analysis of Wikipedia Link Structure

In this section, we analyse the link structure of Wikipedia. We use the XML'ified snapshot of the English Wikipedia (of early 2006) used at INEX [12]. We use the set of 114 ad hoc topics from INEX 2006, with their associated relevance judgments. In this section, we only consider Wikipedia pages and the internal links ignoring links to pages outside Wikpedia.

2.1 Degree Distribution

Is the link structure of Wikipedia different from the link structure of the Web? Recall from the above, links in Wikipedia are unlike generic Web links. Does the encyclopedic organization, where there is little redundant information, put a bound on the number of incoming links? Does the organization in mono-topical entries or lemmas restrict the number of outgoing links? We look at the number of different incoming links (indegree) and the number of outgoing links (outdegree). Figure 1 shows the degree distribution of Wikipedia. The indegree or number of

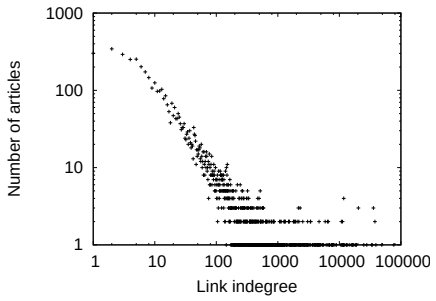

Fig. 2. Wikipedia collection link indegree distribution of 5,646 "relevant" pages

incoming links is shown on the left hand side, and the outdegree or number of outgoing links is shown on the right hand side. Both curves approximate straight lines on the log-log scale, suggesting a power-law distribution that is familiar for the Web at large [13]. In the rest of this paper, we focus on indegrees.

What is the degree distribution of relevant pages? The relevance assessments are at the passage level; we treat a Wikipedia page as relevant for a given topic if, and only if, it contains relevant information. Figure 2 plots the degree distribution for the subset of articles relevant for a INEX 2006 topic. Although there are far fewer data points, we see a similar distribution. There is no absolute evidence in the link indegree: both low indegree and high indegree pages can be relevant.

2.2 Local Degree Distribution

So far, we have looked at global evidence provided by the absolute number of links. We now zoom in on local evidence provided by the number of links among a subset of local pages. We used a standard retrieval system (discussed in detail in Section 4 below) to find the top 100 best matching Wikipedia articles for each of the INEX 2006 topics. We treat these pages as local context, and only consider links between pages in this subset and ignore all further links.

Recall again our conjecture that links in Wikipedia are semantic links. By restricting our view to the local context, a large fraction of these local links should relate to the topic at hand. Is this local structure different from the global link structure investigated above? The left hand side of Figure 3 shows the local degree distribution of pages in the local context of any of the INEX 2006 topics. Again, the plot suggests a power-law distribution. On the right hand side of Figure 3, we zoom in on only those articles which are relevant for any of the INEX topics. Also here we see a similar distribution. This also shows that local indegree is no absolute evidence of relevance: both low local indegree as well as high local indegree pages can be relevant.

2.3 Prior Probability of Relevance

Above, we saw that neither global nor local indegree provides absolute evidence of relevance. But can global or local indegree be used as a (possibly weak) indi-

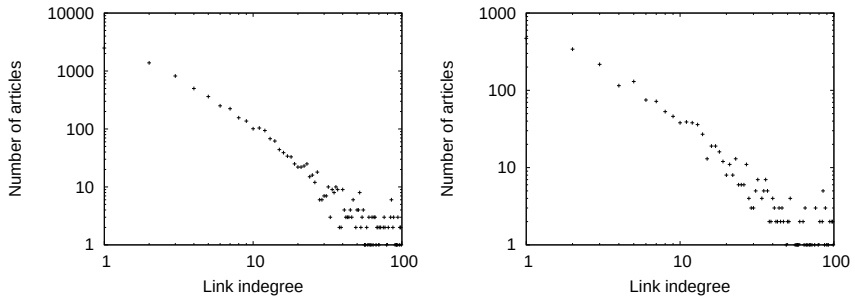

Fig. 3. Wikipedia local link indegree distribution of 11,339 local pages (left) and of 2,489 local relevant pages (right)

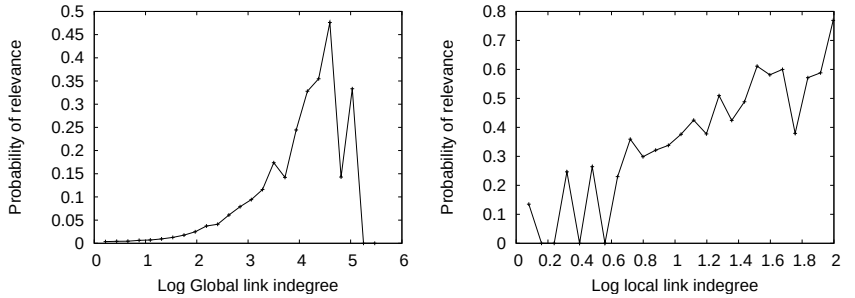

Fig. 4. Prior probability of relevance of Wikipedia global indegree (left) and local indegree (right)

cator of relevance? That is, if we would know nothing more of a page than its global or local indegree, can we make an educated guess about the relevance of the page?

For a page of a given indegree, we can calculate the prior probability that it is relevant (with respect to at least one of the INEX topics). Specifically, we calculate the fraction of pages with that indegree that is relevant to any of the topics. To overcome data sparseness, we group the indegrees in bins for which we use an exponential scale. The left hand side of Figure 4 shows the prior probability of relevance of global indegree. We see a clear increase in the prior probability of relevance with increasing global indegree. Although there are more relevant pages with a low indegree (as was shown in Figure 2), this number is dwarfed by the total number of pages with a low indegree (as shown in Figure 1), leading to a relatively low prior probability of relevance. Conversely, although the number of relevant pages with a high indegree is modest, this is still a substantial fraction of all the pages with a high indegree—up to a third of these pages is relevant for at least one of the INEX topics.

We do the same analysis for the local indegree, shown on the right hand side of Figure 4. The prior probability of relevance also clearly increases with local indegree. Again, although the absolute number of relevant pages with a low local

Table 1. Top 10 Wikipedia articles for topic 339 "Toy Story"

Title	Global indegree	Title	Local indegree
1990s	10,033	Toy Story	96
Screenwriter	762	Toy Story 2	34
Gnosticism	424	Pixar	34
Madeira Islands	339	Monsters, Inc.	25
Psychedelic music	339	Finding Nemo	22
1995 in film	314	Aladdin (1992 film)	14
Computer-generated imagery	310	Madeira Islands	12
Academy Award for Original Music Score	268	Computer-generated imagery	12
Tom Hanks	248	Buzz Lightyear	12
Debian	210	1990s	11

indegree is higher (as shown in Figure 3), a larger fraction of pages with a high local indegree is relevant. The prior probability of relevance rises to well above 0.5 for pages with a high local indegree.

2.4 Naive Reranking

We selected one topic to look in detail at what happens to the top results when naively reranked by indegree. Topic 339 has title *Toy Story*, and is about the computer animated movie from 1995. We took the top 100 articles from the baseline run described in Section 3 below, and list the 10 articles with the highest global indegree (on the left) and the 10 highest local indegree (on the right) in Table 1.

The articles with the highest global indegree are at best slightly related to the *Toy Story* movie, but there are also infiltrations like *Screenwriter* and *Gnosticism*—pages that might be important or authoritative, but outside the scope of our topic at hand. The articles with the highest local indegree look much more promising, but still there are a lot of articles that are only weakly related to *Toy Story*. The qualitative analysis suggests that global and local indegree are weak indicators of relevance. Therefore, in reranking, their weight should be small compared to the weight of the content-based retrieval score.

Summarizing, our analysis of the link structure reveals that the Wikipedia link structure is a (possibly weak) indicator of relevance of Wiki pages. A naive reranking based on only global or local indegree is not effective: it leads to the infiltration of important but off-topic pages.

3 Incorporating Link Evidence

In this section, we discuss how link evidence can be incorporated in our retrieval model. Results at INEX can be arbitrary XML elements (such as paragraphs, sections, or the whole article), and we simply index each XML element as a separate document. Link structure is used at the article or Wiki page level, we simply associate each element with the link structure of the article it is part of.

3.1 Retrieval Model

We use a language model extension of Lucene [14], i.e., for a collection D, document d and query q:

$$P(d|q) = P(d) \cdot \prod_{t \in q} \left((1 - \lambda) \cdot P(t|D) + \lambda \cdot P(t|d) \right), \tag{1}$$

where $P(t|d) = \frac{\text{freq}(t,d)}{|d|}$, $P(t|D) = \frac{\text{freq}(t,D)}{\sum_{d' \in D} |d|}$, and $P(d) = \frac{|d|}{\sum_{d' \in D} |d|}$. For our experiments we have used $\lambda = 0.15$ throughout. Our efficient implementation of the model calculates ranking-equivalent logs of the probabilities [15]. We take the exponent to get a score resembling a probability, and only then apply the length-prior.

3.2 Link Degree Priors

We incorporate link evidence by by multiplying the retrieval score with a further link degree prior:

$$Score = Score_{retrieved} \cdot Prior \tag{2}$$

We will, for convenience, refer to the link evidence as prior, even though we do not actually transform it into a probability distribution. Note that we can turn any prior into a probability distribution by multiplying it with a constant factor $\frac{1}{\Sigma_{d \in D} Prior(d)}$, leading to the same ranking.

Recall from above that we need to be careful when incorporating link evidence. We do not want to retrieve pages that only have a high link score, i.e., pages that may be important but unrelated to the topic of request. Hence, as a safe-guard, we apply the link priors only to the first 100 retrieved articles per topic. That is, we process the list of XML element results until we have encountered 100 articles, and only use the global or local indegree of these articles by simply treating the indegrees of lower ranked articles as zero. Note that elements deep down the result list may still be boosted if they belong to one of the top 100 articles.

3.3 Baseline

Our baseline is the retrieval model without using link evidence. To explain the impact of the link evidence, we look again in detail at Topic 339 and the effects of the priors on the top 10 articles. In the upper left corner of Table 2 the titles of the top 10 retrieved Wikipedia articles for the baseline run are given.

3.4 Global Indegree

The *global indegree prior* is proportional to the global degree of an article:

$$P_{\mathsf{Glob}} \propto 1 + global \tag{3}$$

Table 2. Top 10 Wikipedia articles for topic 339 "Toy Story"

Baseline run	**Global indegree prior**
List of Disney animated features' titles in various languages	1990s
Toy Story	Screenwriter
Toy Story 2	Gnosticism
Buzz Lightyear	1995 in film
Toy Story 3	Computer-generated imagery
List of computer-animated films	Toy Story
100 Greatest Cartoons	Academy Award for Original Music Score
C64 Direct-to-TV	Tom Hanks
List of Capcom games	Pixar
Pixar	Debian
Local indegree prior	**Local/Global indegree prior**
Toy Story	Toy Story
Toy Story 2	List of Disney animated features' titles in various languages
Pixar	Toy Story 2
Monsters, Inc.	Toy Story 3
Buzz Lightyear	Buzz Lightyear
Finding Nemo	List of computer-animated films
Aladdin (1992 film)	100 Greatest Cartoons
Toy Story 3	Sheriff Woody
Computer-generated imagery	Timeline of CGI in film and television
1990s	Andrew Stanton

Alternatively, we use a conservative *log global indegree prior*:

$$P_{\mathsf{LogGlob}} \propto 1 + \log(1 + \mathit{global}) \tag{4}$$

Inspecting our running example immediately confirms that we need to be careful when incorporating global link evidence. In the upper right corner of Table 2, the top 10 articles after reranking by global indegree are given. Although some top 10 results are retained, *Toy Story* at rank 6 and *Pixar* at rank 9, we see the infiltration of pages with a high global indegree—all of them are in Table 1—but with only a loose relation to the topic at hand.

3.5 Local Indegree

The *local indegree prior* is proportional to the local degree of an article:

$$P_{\mathsf{Loc}} \propto 1 + \mathit{local} \tag{5}$$

Alternatively, we use a conservative *log local indegree prior*:

$$P_{\mathsf{LogLoc}} \propto 1 + \log(1 + \mathit{local}) \tag{6}$$

The local indegree prior (shown in the lower left corner of Table 2) results in the *Toy Story* page at the top rank, thereby improving upon the baseline run. Also, *Pixar* moves up from rank 10 to 3. However, at rank 10 we see that *1990s*, the article with the highest local indegree in Table 1, has infiltrated the top results.

Table 3. Results of link evidence on three INEX 2006 ad hoc retrieval tasks. Best scores are in bold-face. Significance levels are 0.05 (*), 0.01 (**), and 0.001 (***).

Run ID	Thorough		Focused		Relevant in Context	
	MAep,off		nxCG@10,off		MAgP	
Baseline	0.0353		0.3364		0.1545	
Global Indegree	0.0267	-24.40***	0.1979	-41.16***	0.1073	-30.57***
Log Global Indegree	0.0335	-4.99	0.3066	-8.87**	0.1352	-12.50***
Local Indegree	0.0405	+14.75*	0.3218	-4.34	0.1467	-5.02*
Log Local Indegree	0.0418	+18.46***	0.3460	+2.85	0.1515	-1.96
Local/Global Indegree	**0.0463**	+31.08***	**0.3629**	+7.88**	**0.1576**	+1.99*

3.6 Local/Global Indegree

We can also weight the importance of observing local links by their number of global links—basically a $tf \cdot idf$ weighting of link evidence [16]. The combined *local/global indegree prior* is calculated as:

$$P_{\mathsf{LocGlob}} \propto 1 + \frac{local}{1 + global} \tag{7}$$

The combination prior (lower right corner of Table 2) improves further on the original top 10 by ranking *Toy Story* as the top articles and moving *Toy Story 3* from rank 5 to 4. Also, some unrelated articles like *C64 Direct-to-TV* and *List of Capcom games* are replaced by closer related articles, *Timeline of CGI in film and television* and *Andrew Stanton* (one of the writers of *Toy Story*).

Summarizing, we defined a number of ways—global, local, and combined local/global indegree—to incorporate link evidence into the retrieval model. The different indegree priors correspond to different levels of sensitivity to the local context of the topic of request.

4 Experimental Results

In this section, we discuss the results of applying the degree priors to three of the INEX 2006 Ad Hoc retrieval Tasks.

4.1 Baseline

Our baseline run is a standard language model run, using an index containing all the XML elements of the Wikipedia XML Collection [17]. The scores of the baseline run are in Table 3.

For the INEX 2006 *Thorough* task, where the aim is to estimate the relevance of individual XML elements, this run scores 0.0353 on the official MAep measure.[5] For the *Focused* task, no overlapping elements are allowed and we

[5] Mean average effort/precision (MAep) is a generalized MAP-like measure; normalized extended cumulative gain (nxCG@10) is resembling to precision at rank 10; mean average generalized precision (MAgP) is a version of MAP with partial scores per article. See [18] for details of the INEX 2006 measures.

post-process the Thorough run using a top-down list-based removal of elements partially overlapping with earlier seen results. That is, we traverse the list top-down, and simply remove any element that is an ancestor or descendant of an element seen earlier in the list. The resulting run scores 0.3364 on the official nxCG@10 measure. Finally, for the *Relevant in Context* task, there is a further restriction that articles may not be interleaved and, again, we post-process the Focused run using a top-down list-based clustering of results per retrieved article. The resulting run scores 0.1545 on the official MAgP measure.

4.2 Global Indegree

Now, we turn our attention to global indegree. The results are negative: both the global indegree prior and the log global indegree prior lead to loss of performance for all three tasks. The decrease in performance is significant (bootstrap test, one-tailed) for all cases except for the log global indegree prior and the Thorough task. A plausible explanation is suggested by looking at Table 2. The original top 10 articles of the baseline run are infiltrated by non-relevant documents with high global indegrees—important pages, but off-topic.

4.3 Local Indegree

Next, we try the local link evidence, and use the (log) local indegree prior. The results are mixed. The local indegree prior leads to a significant gain in performance for Thorough (15%), but a loss for Focused (-4%) and for Relevant in Context (-5%). The more conservative log local indegree prior fares better and leads to a gain in performance for Thorough (18%, significant at $p < 0.001$) and for Focused (3%), but still a loss for Relevant in Context (-2%). Although the scores are much better than for the global indegree, there is still no overall improvement. This may still be due to the infiltration of non-relevant documents with high local indegrees. Since the local indegrees are generally much lower than global indegrees—with $N = 100$ the maximal local indegree is 99—the infiltration effect is also much smaller.

4.4 Local/Global Indegree

Finally, we experiment with the combined prior. Here the situation is quite different. For the Thorough task, we see an improvement of 31%. For the Focused Task, we see and improvement of 8%. For the Relevant in Context Task, we get an improvement of 2%. For all three tasks, the improvement is significant. The combined prior seems to effectively take the local context into account, and is effective for improving ad hoc retrieval.

Summarizing, we experimented with the use of global, local, and combined local/global link evidence, and found that only the combined local/global prior leads to a significant improvement of retrieval effectiveness for all tasks.

5 Discussion and Conclusions

In this paper, we investigated the importance of link evidence in Wikipedia ad hoc retrieval. The link structure of Wikipedia is an interesting special case of hyperlinking on the Web at large: the links in Wikipedia are semantic—they link to other pages relevant to the local context. Our main aim is to find out if Wikipedia's link structure can be exploited to improve the ad hoc retrieval of relevant information.

Our first research question was:

– Can the degree structure of a semantically linked document collection be used as evidence for the relevance of ad hoc retrieval results?

We analysed the degree structure of Wikipedia pages, and of Wikipedia pages relevant for a particular set of ad hoc retrieval topics. Our findings are that pages with a high global indegree are more likely to be relevant than pages with a low global indegree. Since global link evidence may lead to the retrieval of important but off-topic pages, we also looked at local indegree considering only links between pages retrieved in response to a search request. Also here we saw that pages with a high local indegree are more likely to be relevant than pages with a low local indegree. So the answer to our first research question is yes: the Wikipedia link structure is a (possibly weak) indicator of relevance.

Our second research question was:

– Can global indegree, local indegree, or a combination of the two, be used effectively to rerank results in general ad hoc retrieval?

In order to answer this question, we have to operationalize how to incorporate link evidence into our retrieval model, and then conduct experiments that try to establish its utility. The link topology in itself is not sensitive to the query or local context. Put differently, if we browse following a sequence of links, the similarity to the source page will water down quickly. This is especially true for the global link structure, which may lead quickly to a loss of focus on the topic at hand and allow for the infiltration of authoritative but off-topic pages. The local link structure ensures that only links within the local context are awarded, but the ranking may still suffer from a similar bias on authoritativeness over topicality (although the effect will be less strong). Hence, this leads to a third way in which the number of local links is normalized by the number of global links (basically a $tf \cdot idf$ weighting of link evidence). As it turns out, the use of global link evidence leads to a loss of performance, the use of local link evidence leads to mixed results, but the combined local/global link evidence leads to significant improvement of retrieval effectiveness. So, the answer to our second research question is also yes: if the link evidence is made sensitive to the local context we see an improvement of ad hoc retrieval effectiveness.

Earlier experiments at TREC using crawled Web data have failed to establish the utility of link evidence for ad hoc retrieval. In contrast with these TREC experiments, Wikipedia's link structure can help improve the effectiveness of ad hoc retrieval.

References

[1] Wikipedia: The free encyclopedia (2008) http://en.wikipedia.org/.

[2] Page, L., Brin, S., Motwani, R., Winograd, T.: The pagerank citation ranking: Bringing order to the web. Technical report, Stanford Digital Library Technologies Project (1998)

[3] Kleinberg, J.M.: Authoritative structures in a hyperlinked environment. Journal of the ACM **46** (1999) 604–632

[4] Hawking, D.: Overview of the TREC-9 web track. In: The Ninth Text REtrieval Conference (TREC-9), National Institute for Standards and Technology. NIST Special Publication 500-249 (2001) 87–102

[5] Kraaij, W., Westerveld, T.: How different are web documents? In: Proceedings of the ninth Text Retrieval Conference, TREC-9, NIST Special Publication (May 2001)

[6] Hawking, D., Craswell, N.: Very large scale retrieval and web search. In: TREC: Experiment and Evaluation in Information Retrieval. MIT Press (2005) 199–231

[7] Kraaij, W., Westerveld, T., Hiemstra, D.: The importance of prior probabilities for entry page search. In: Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM Press, New York NY, USA (2002) 27–34

[8] Ogilvie, P., Callan, J.: Combining document representations for known-item search. In: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM Press, New York NY, USA (2003) 143–150

[9] Kamps, J.: Web-centric language models. In: CIKM'05: Proceedings of the 14th ACM International Conference on Information and Knowledge Management, ACM Press, New York NY, USA (2005) 307–308

[10] Broder, A.: A taxonomy of web search. SIGIR Forum **36**(2) (2002) 3–10

[11] INEX: INitiative for the Evaluation of XML retrieval (2007) http://inex.is.informatik.uni-duisburg.de/.

[12] Denoyer, L., Gallinari, P.: The Wikipedia XML Corpus. SIGIR Forum **40** (2006) 64–69

[13] Faloutsos, M., Faloutsos, P., Faloutsos, C.: On power-law relationships of the internet topology. In: SIGCOMM '99: Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communication, ACM Press, New York NY, USA (1999) 251–262

[14] ILPS: The ILPS extension of the Lucene search engine (2008) http://ilps.science.uva.nl/Resources/.

[15] Hiemstra, D.: Using Language Models for Information Retrieval. PhD thesis, Center for Telematics and Information Technology, University of Twente (2001)

[16] Sparck Jones, K.: A statistical interpretation of term specificity and its application in retrieval. Journal of Documentation **28** (1972) 11–21

[17] Sigurbjörnsson, B., Kamps, J., de Rijke, M.: An element-based approach to XML retrieval. In: INEX 2003 Workshop Proceedings. (2004) 19–26

[18] Lalmas, M., Kazai, G., Kamps, J., Pehcevski, J., Piwowarski, B., Robertson, S.: INEX 2006 evaluation measures. In: Comparative Evaluation of XML Information Retrieval Systems: Fifth Workshop of the INitiative for the Evaluation of XML Retrieval (INEX 2006). Volume 4518 of LNCS., Springer Verlag, Heidelberg (2007) 20–34