Overview of INEX 2007 Link the Wiki Track

Darren Wei Che Huang[1], Yue Xu[2], Andrew Trotman[3], Shlomo Geva[4]

Faculty of Information Technology, Queensland University of Technology, Brisbane
Queensland Australia
w2.huang@student.qut.edu.au [1]
yue.xu@qut.edu.au [2]
andrew@cs.otago.ac.nz [3]
s.geva@qut.edu.au [4]

Abstract. Wikipedia is becoming ever more popular. Linking between documents is typically provided in similar environments in order to achieve collaborative knowledge sharing. However, this functionality in Wikipedia is not integrated into the document creation process and the quality of automatically generated links has never been quantified. The Link the Wiki (LTW) track at INEX in 2007 aimed at producing a standard procedure, metrics and a discussion forum for the evaluation of link discovery. The tasks offered by the LTW track as well as its evaluation present considerable research challenges. This paper briefly described the LTW task and the procedure of evaluation used at LTW track in 2007. Automated link discovery methods used by participants are outlined. An overview of the evaluation results is concisely presented and further experiments are reported.

Keywords: Wikipedia, Link Discovery, Assessment, Evaluation

1 Introduction

In 2007, Geva and Trotman suggested the Link the Wiki track that aims to provide an evaluation forum for link discovery in Wikipedia and for objectively evaluating the performance of such algorithms. Wikipedia is composed of millions of articles in English and it offers many attractive features as a corpus for information retrieval tasks. The INEX Wikipedia collection has been converted from its original wiki-markup into XML [1]. This collection is composed of a set of XML files where each file corresponds to an online article in Wikipedia.

Links between pages are essential for navigation, but most systems require authors to manually identify each link. Authors must identify both the anchor and the target page in order to build a knowledge network. This creates a heavy and often unnecessary burden on content providers [2] who would prefer to focus on the content and let the system assist in discovering the relationship between the new content and content already in the collection. Without assistance, as the size of the collection increases, link creation and maintenance can become unmanageable. The maintenance cost of keeping the entire network up to date is huge – and Wikipedia has seen faster than linear growth for many years. Authors are typically unaware of all available links, and even if they are aware of the pre-existing content they are unlikely to be aware of newly created content to which they could link. Page maintenance, in particular linking to content added after a page is created, is a burden on content providers who often do not maintain their content (hence the collaborative nature of these information resources). Ellis et al. [3] have shown significant differences in the links assigned by different people. To eliminate the human effort required to build a highly accurate linking network, to reduce the chance of erroneous links, and to keep links up-to-date, automatic link discovery mechanisms are needed.

The user scenario for the Link the Wiki task is that of an end user who creates a new article in the Wikipedia. A Wikipedia link discovery system then automatically selects a number of prospective anchor texts, and multiple link destinations for each anchor. This is namely the discovery of *outgoing links*. A Wikipedia link discovery system also offers prospective updates to related links in other (e.g. older) wiki articles, which may point to a Best Entry Point (BEP) within this newly created article. In this way, *incoming links* are generated. Therefore, links on each article can always be up-to-date with the latest information existing within the wiki system.

At INEX 2007, the LTW task addressed only document-to-document links, in order to bootstrap the track. Systems were required to discover incoming and outgoing links for selected topics. Evaluation was based on existing Wikipedia links and performance was measured using standard IR metrics.

The remainder of this paper is organized as follows. In the next section, we review the link discovery literature, Wikipedia related research, assessment and evaluation. In section 3 we describe the LTW track at INEX, which comprises the description of task and submission. Section 4 investigates Wikipedia links while the procedure of assessment and evaluation including the result set generation and evaluation procedure is depicted in Section 5. Following that, the techniques used by all participants and the results of evaluation are depicted, and the better performing runs are discussed in some more detail. Finally, conclusions and future direction are presented.

2 Background of Link Discovery

As suggested by Wilkinson & Smeaton [2], navigation between linked documents is a great deal more than simply navigating multiple results of a single search query. Linking between digital resources is becoming an ever more important way to find information. Through hypertext navigation, users can easily understand context and realize the relationships of related information. However, since digital resources are distributed it has become a difficult task for users to maintain the quality and the consistency of links. Automatic techniques to detect the semantic structure (e.g. hierarchy) of the document collection and the relatedness and relationships of digital objects have been studied and developed [4]. Early works, in the 1990s, determined whether and when to insert links between documents by computing document similarity. Approaches such as term repetition, lexical chains, keyword weighting and so on were used to calculate the similarity between documents [5, 6, 7]. These approaches were based on a document-to-document linking scenario, rather than identifying which parts of which documents were interrelated.

Adafre and de Rijke [8] state that most links in Wikipedia are conceptual. The Wikipedia linking network offers hierarchical information and links aim to expand on the concepts in their anchors. The anchors imply the concept while the links are complementary to the concept. Since there is no strict standard of editing there are problems with *over linking* and *missing links*. They proposed a method of discovering *missing links* in Wikipedia pages by clustering topically related pages using *LTRank* and identified link candidates by matching anchor texts. Page ranks using the *LTRank* method are based on the co-citation and page title information. Experiment results show reasonable outcome.

Jenkins [9] developed a link suggestion tool, *Can We Link It*. This tool extracts a number of anchors which have not been linked within the current article and that might be linked to other pages in the Wikipedia**.** With this tool, the user can accept, reject, or click *"don't know"* to leave a link as undecided. Using this tool the user can add new anchors and corresponding links back to the Wikipedia article.

A collaborative knowledge management system, called *PlanetMath*, based on the Noosphere system has been developed for mathematics [10]. It is encyclopaedic, (like the Wikipedia), but mainly used for the sharing of mathematical knowledge. Since the content is considered to be a semantic network, entries should be cross-referenced (linked). An automatic linking system provided by Noosphere employs the concept of conceptual dependency to identify each entry for linking. A classification hierarchy used in online encyclopedias is used to improve the precision of automatic linking. In practice, the system looks for common anchors that are defined in multiple entries and creates links between them, once the page metadata is identified as related. Based on the Noosphere system, NNexus (Noosphere Networked Entry eXtension and Unification System) was developed to automate the process of the automatic linking procedure [11]. This was the first automatic linking system to eliminate the linking efforts required by page authors. Declarative linking priorities and clauses are specified to enhance the linking precision. An approach, called *invalidation index*, was developed to invalidate entries belonging to those concepts where there are new entries. Reputation based collaborative filtering techniques could be used to provide personalized links.

Research on the Wikipedia has been undertaken in recent years. A set of experiments, based on Markov Chains [12], for finding related pages within Wikipedia collection was undertaken using two Green-based methods [13], *Green* and *SymGreen*, and three classical approaches, *PageRankOfLinks*, *Cosine with tf-idf weight* and *Co-citations*. The results show the Green method has better performance at finding similar nodes than only relying on the graph structure. Although page titles and category structure can be used to classify documents in Wikipedia, properties such as the internal text of the articles, the hierarchical category, and the linking structure should be used [14]. *Wikirelate* proposed by Strube and Ponzetto uses *Path*, *Information content* and *Text overlap* measures to compute the semantic relatedness of words [15]. These measures mainly rely on either the texts of the articles or the category hierarchy. Gabrilövich and Markövitch [16] introduce a new approach called Explicit Semantic Analysis (ESA), which computes relatedness by comparing two weighted vectors of Wikipedia concepts that represent words appearing within the content. Common to this research is the use of the *existing* linking structure and content (category, etc.).

Various link-based techniques based on the correlation between the link density and content have been developed for a diverse set of research problems including link discovery and relevance ranking [8]. Moreover, communities can be identified by analysing the link graph [17]. Beside co-citation used by Kumar et al. [18] to measure similarity, bibliographic coupling and SimRank based on citation patterns, and the similarity of structural context (respectively), have also been used to identify the similarity of web objects [19]. The companion algorithm derived from HITS has also been proposed for finding related pages (by exploiting links and their order on a page) [20, 21].

The assessment of IR results has been a challenge in IR experiments for many years because despite the existence of some standard procedures, relevance is hard to define and cross-assessor agreement levels are often low (so individual judgments come under dispute). Worse, it is difficult to compare IR methods which are able to retrieve highly relevant documents with those that retrieve less relevant documents because assessments are usually binary. The use of Precision-Recall (P-R) curves is typical in IR; however, Schamber [22] argues that traditional P-R based comparison using binary relevance cannot adequately capture the variability and complexity of relevance. Relevance is a multilevel circumstance where, for a user, the degree of relevance may vary from document to document.

Several studies done in [23] have examined components that influence judgments and the criteria of relevance (including graded relevance) in information seeking and retrieval. Kekalainen and Jarvelin [24] argued that evaluation methods should be flexible enough to handle different degrees of judgment scales. They proposed generalized precision and recall that can incorporate a continuous relevance scale into the traditional precision and recall measures. Their experiments demonstrate that the evaluation approach can distinguish between retrieval methods fetching highly relevant documents from those retrieving partially relevant documents.

3 The Link-the-Wiki Track

3.1 Tasks

Wikipedia is composed of millions of interlinked articles in numerous languages and offers many attractive features for retrieval tasks. The current INEX Wikipedia collection contains a snapshot of the Wikipedia English collection from 2006 and contains 660,000 documents and is about 4GB in size. In INEX 2007 the linking task used 90 topics, nominated from the existing collection by participants [25]. The task is two fold:

1. *Outgoing links*: Recommend anchor text and destinations within the Wikipedia collection.
2. *Incoming links*: Recommend incoming links from other Wikipedia documents.

Missing topics were regarded as having a score of zero for the purpose of calculating system rank when using all topics. Up to 250 outgoing links and 250 incoming links were allowed for each topic. Surplus links were discarded in computing the performance.

3.2 Submission

Each participant was encouraged to submit up to 5 runs. The submission example can be found in Appendix A-1. Finally, thirteen runs were submitted by 4 groups in 2007. The University of Amsterdam had 5 runs, Waterloo submitted 1 run, Otago 5 runs and QUT 2 runs.

4 Wikipedia Links

The 90 LTW topics were examined to discover the relationship between anchors and page titles. The result showed that in 81 of the 90 topics at least 50% of the links match an existing page name (see Table 1). This could be because the links were generated through careful construction by a user, or automatically by matching page names, either way such links are relatively easy to find. For example, the word, *Explorer*, in the document can be manually linked using Wiki markup [[destination page title| anchor text]] to the page titles, *Explorer (Learning for Life)*, *Explorer (album)* or *Explorer (novel)*, depending on the context. According to this investigation, we can expect a method that systematically matches potential anchor strings with page names to identify most links – and to achieve a better recall (e.g. near 1) when most page names and anchors are exact matches. Although this implies that we can expect high recall from simple page-name matching strategies, it does not necessarily mean that we can expect high precision – many matching links are not relevant (e.g. polyvalent terms). As the Wikipedia is a huge repository of definitions it is quite easy to find matching page names which are not relevant.

Table 1. Ratio of matching names between anchors and links

Ratio of Match	Number of Topics
90% ~ 100%	1
80% ~ 90%	8
70% ~ 80%	26
60% ~ 70%	35
50% ~ 60%	16
40% ~ 50%	2
30% ~ 40%	2

5 Assessment and Evaluation

The main focus of the Link-the-Wiki track in 2007 was to explore an automated evaluation procedure without human assessment effort. The incoming and outgoing links were retrieved directly from the existing collection links to form the result set for evaluation. This can completely eliminate the assessment effort. Accompanying the automatically generated result set (see Appendix A-2), the proposed evaluation tool was developed to examine the performance of the link discovery methods. We notice in the result set that incoming links outnumber outgoing links. This could be because the number of outgoing links may be restricted by the length of the document (only proper anchors will be specified) but there can be many different pages linked to the topic page.

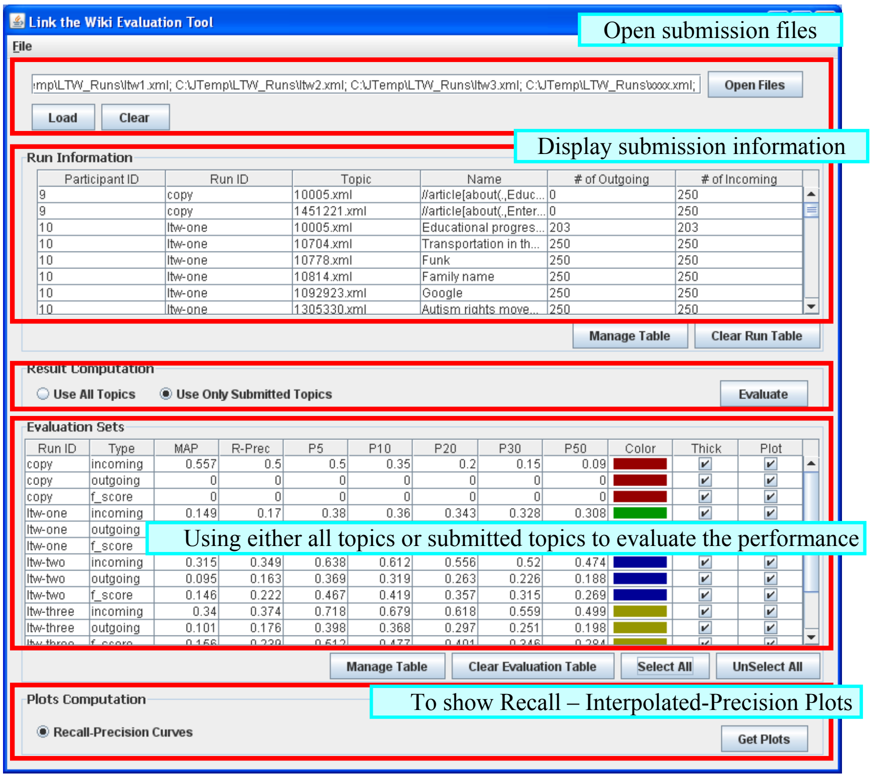

Figure 1. The Evaluation Tool

As we can see above, we treat the Wikipedia links as the ground-truth. However, they are obviously not perfect. Many links in Wikipedia are automatically generated but some of them do not have a clear topic relation. *Year* links, for example, are very often unrelated to the content of the document, but are easy to discover. Such links have probably no utility for IR. Problematically, when used in evaluation as the ground truth, they may also lead to optimistic evaluation results when easily identified by link-discovery systems. Many potentially good links that have not been identified by Wikipedia users are amenable to automatic discovery. Such useful returned links which are missing from the ground truth could result in poor evaluation scores for link discovery systems, hence leading to pessimistic evaluation results. So although it is not possible to quantify the absolute performance of link discovery by using automated result assessment, the procedure we used provides a trade-off between assessment effort and absolute accuracy of measurement.

It is reasonable to conjecture that *comparative evaluation* of methods and systems is still informative. Through the investigation of comparative analysis of automated linking system for the Wikipedia, it should remain possible to improve link discovery methods.

An evaluation tool, named *ltwEval*, was developed for LTW 2007 (see Figure 1). The performance measures include *Mean Average Precision (MAP)*, *precision at the point of the number of relevant documents (R-Prec)*, and *precision at varying numbers of documents retrieved (P@r)*. Plots of Interpolated Precision-Recall for incoming, outgoing and a combined score are also computed for comparison. By combined score we refer to the harmonic mean of the various values obtained for incoming and outgoing links. The *ltwEval* program was developed in Java for platform independence, but is GUI driven and provides more extensive functionality than traditional evaluation software. This assists participants by making result exploration and analysis easier.

6 Link Discovery Methods

6.1 Approaches and Evaluation Results

In this section we briefly describe the approaches that were taken by participants.

The University of Amsterdam system assumed that Wikipedia pages link to each other when articles are similar or related in content. For each of the 90 topics, the system queried the index of the entire collection, (excluding the topics). This was done by using the full topic as the query, but excluding stop words, and with important terms derived from a language model. The top 100 files (anchors) were selected for each topic. They experimented with line matching from the orphans to the anchor files. For the outgoing links, the system matched each line of a topic with the lines of the anchors until a matching line has been found. For the incoming links, the system iterated over all lines of each anchor for each line of the topic. The generated runs were based on the names of the pages, exact lines, and longest common substrings (LCSS) expanded with WordNet synonyms. The results show that the run based on restricting the line matching to the names of pages performed best. Therefore, submitted runs have a good performance on average, especially for incoming links (see Figure 2).

The University of Otago system identified terms within the document that were over represented by comparing term frequency in the document with the expected term frequency (computed as the collection frequency divided by document frequency). From the top few over-represented terms they generated queries of different lengths. A BM25 ranking search engine was used to identify potentially relevant documents. Links from the source document to the potentially relevant documents (and back) were constructed. They showed that using 4 terms per query was more effective than fewer or more.

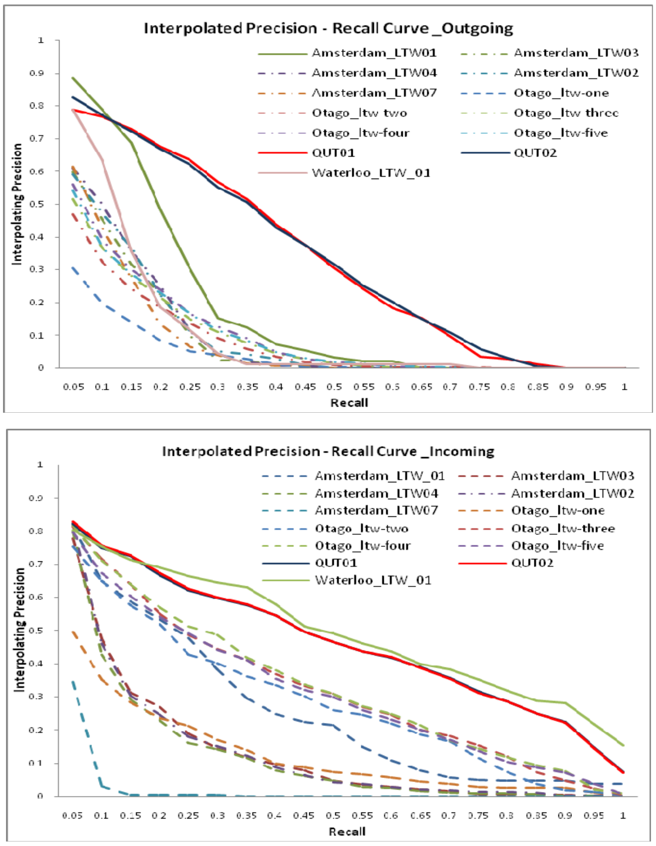

Figure 2. Interpolated Precision-Recall of Outgoing and Incoming links

The University of Waterloo system found the first 250 documents (in document collection order) that contain the topic titles and then generated article-to-article Incoming links. For outgoing links, they performed link analysis. The system computed the probabilities that each candidate anchor would be linked to a destination file. The probability that a candidate anchor would be linked was computed (essentially) as the ratio of the number of times that the anchor text was actually linked in the collection, to the number of times that the anchor text appeared in the collection.

The Queensland University of Technology (QUT) system identified incoming links using a ranked search for documents that were *about* the new document title. Outgoing links were identified by running a window over the new document text and looking for matching document titles in the collection. The window size varied from 12 words down to 1 word, and included stop words. Longer page names were ranked higher than shorter page names, motivated by the observation that the system was less likely to hit on a longer page name by accident. From the official results shown in Table 2, 3 and Figure 2, QUT runs have better performance than any other run proposed by participants, utilizing page title matching. However, the Waterloo system, using link analysis, was subsequently shown to outperform this system, for outgoing link discovery, when suitably configured.

Table 2. MAP of Outgoing and Incoming Links

MAP Outgoing Links			**MAP Incoming Links**		
1	QUT02	0.484	1	QUT02	0.318
2	QUT01	0.483	2	QUT01	0.314
3	Waterloo_LTW_01	0.465	3	Amsterdam_LTW_01	0.147
4	Otago_ltw-four	0.339	4	Otago_ltw-four	0.102
5	Otago_ltw-five	0.319	5	Otago_ltw-five	0.101
6	Otago_ltw-three	0.318	6	Waterloo_LTW_01	0.093
7	Otago_ltw-two	0.284	7	Otago_ltw-three	0.092
8	Amsterdam_LTW_01	0.226	8	Otago_ltw-two	0.081
9	Otago_ltw-one	0.123	9	Amsterdam_LTW04	0.080
10	Amsterdam_LTW03	0.110	10	Amsterdam_LTW02	0.080
11	Amsterdam_LTW02	0.108	11	Amsterdam_LTW03	0.073
12	Amsterdam_LTW04	0.093	12	Amsterdam_LTW07	0.067
13	Amsterdam_LTW07	0.004	13	Otago_ltw-one	0.048

Table 3. R-Precision of Outgoing and Incoming Links

R-Prec Outgoing Links			**R-Prec Incoming Links**		
1	QUT01	0.415	1	Waterloo_LTW_01	0.512
2	QUT02	0.411	2	QUT02	0.505
3	Otago_ltw-four	0.183	3	QUT01	0.503
4	Otago_ltw-five	0.183	4	Otago_ltw-four	0.379
5	Amsterdam_LTW_01	0.182	5	Otago_ltw-three	0.363
6	Otago_ltw-three	0.173	6	Otago_ltw-five	0.356
7	Otago_ltw-two	0.156	7	Otago_ltw-two	0.331
8	Amsterdam_LTW02	0.154	8	Amsterdam_LTW_01	0.258
9	Amsterdam_LTW04	0.149	9	Amsterdam_LTW02	0.165
10	Amsterdam_LTW03	0.141	10	Otago_ltw-one	0.153
11	Amsterdam_LTW07	0.127	11	Amsterdam_LTW03	0.144
12	Waterloo_LTW_01	0.103	12	Amsterdam_LTW04	0.142
13	Otago_ltw-one	0.098	13	Amsterdam_LTW07	0.020

6.2 Discussion of Best Approaches

In this section we concentrate on the performance of the two most successful approaches at INEX 2007 [25, 26], the Waterloo and QUT systems.

The best performing approaches were those that used either existing anchors to predict suitable anchors (Waterloo), or matching document titles to predict suitable anchors. The performance of these 2 approaches for discovering outgoing links (note: produced *after* the INEX 2007 workshop and some implementation corrections) are depicted in Figure 3. It can be seen that both approaches produce a very good result with high precision over a wide range of recall levels. This is precisely the kind of performance needed to satisfy a user.

There are considerable differences between the two approaches. The Waterloo approach relies on the availability of an extensive pre-existing web of anchor to document links in the collection. This pre-requisite may not always be satisfied, particularly when a new cluster of documents in a new domain is added to the collection in bulk, or when a new Wikipedia-like resource is created. However, the approach can discover links that are not solely based on a match between anchor text and a document title. If an anchor is frequently linked to a document with a different title, it will become a highly probable link. For instance, the Waterloo system was able to link *Educational Philosophy* to a document titled *The Philosophy of Education.* By contrast, the QUT approach only discovered matching document titles. In regard to the investigation described in Section 4, LTW approaches aiming at matching anchors with page titles can achieve a certain level of performance. Although the performance of QUT is somewhat lower than that of Waterloo, the approach is applicable to any collection, regardless of the pre-existing link structure. It could immediately be applied to any document collection, completely devoid of links or with pre-existing links.

Figure 3 presents the precision-recall curves for the two systems. *Anchors 90* is the Waterloo system and *Page Titles 90* is the QUT approach. Both are shown for the 90 INEX topics. The result shows that the anchor-based approach (Waterloo) is better at almost all recall points. In order to verify the scalability and reliability of the INEX evaluation itself, the QUT system was also tested with 6600 randomly selected topics (1% of the collection) – the plot entitled *Page Titles 6600* corresponds to this experiment. It demonstrates that the approach taken by INEX LTW in 2007 is robust and that 90 topics represent an adequate number of topics for the track.

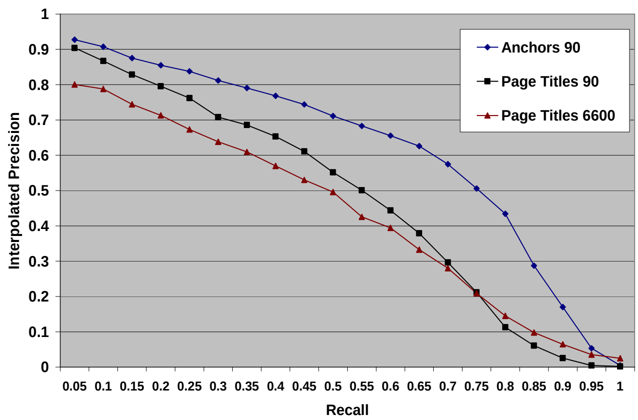

Figure 3. System performance of discovering outgoing links in scalability test

7 Conclusions and Future Direction

As far as we are aware, the Link the Wiki task at INEX is the first to offer extensive reusable independent evaluation resources for link discovery. Although in 2007 the LTW track still operated the evaluation at the document level, the LTW track has launched a forum to discuss the performance of results for extracting link discovery. The procedure of the LTW track has been defined and an evaluation tool has also been developed to speed the exploration of submission runs. Evaluation results were analyzed in this paper and the main findings described. Using a very large set of documents (1% of the collection), extensive linking experiment has been undertaken and the result has showed that linking is feasible and effective.

It is noticed that document-to-document link discovery systems are very good exhibiting high precision levels at most points of recall, systems are scalable and that several different techniques might be used. This result motivates us to examine (and outline future work) anchor to Best-Entry-Point identification. In future INEX evaluations the task will be defined as anchor to BEP link discovery, and allow multiple links per anchor (actually, the latter is essential for manual evaluation purposed where two systems might link the same anchor to different documents, both of which are correct). Traditional performance measures such as MAP will be adapted to address the performance differences of link-discovery methods in this new scenario. Manual assessment would allow us to study more deeply the nature of link discovery, to identify those links returned by automatic systems that have not yet been identified by Wikipedia authors, and those automatic links that already exist in the Wikipedia and which are not useful (e.g. *year* links are common, yet often of no use).

References

1. Denoyer, L., Gallinari, P.: The Wikipedia XML Corpus, SIGIR Forum, 40(1), June 2006, 64-69.
2. Wilkinson, R., Smeaton, A. F.: Automatic Link Generation, ACM Computing Surveys, 31(4), December 1999.
3. Ellis, D., Furner-Hines, J., Willett, P.: On the Measurement of Inter-Linker Consistency and Retrieval Effectiveness in Hypertext Database, In Proceedings of the 17th Annual International Conference on Research and Development in Information Retrieval, Dublin, Ireland, July 1994, 51-60.
4. Green, S. J.: Building Hypertext Links By Computing Semantic Similarity, IEEE Transactions on Knowledge and Data Engineering, 11(5), September/October 1999, 713-730.
5. Allan, J.: Building Hypertext using Information Retrieval, Information Processing and Management, 33(2), March 1997, 145-159.
6. Green, S. J.: Automated Link Generation: Can We Do Better than Term Repetition?, In Proceedings of the 7th International World Wide Web Conference, Brisbane, Australia, 14-18 April 1998, 75-84.
7. Zeng, J., Bloniarz, O. A.: From Keywords to Links: an Automatic Approach, In Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'04), 5-7 April 2004, 283-286.

8. Adafre, S. F., de Rijke, M.: Discovering missing links in Wikipedia, In Proceedings of the SIGIR 2005 Workshop on Link Discovery: Issues, Approaches and Applications, Chicago, IL, USA, 21-24 August 2005.
9. Jenkins, N.: Can We Link It, http://en.wikipedia.org /wiki/User:Nickj/Can_We_ Link_It, 2007.
10. Krowne, A: An Architecture for Collaborative Math and Science Digital Libraries, Thesis for Master of Science Virginia Polytechnic Institute and State University, 19 July 2003.
11. Gardner, J., Krowne, A., Xiong, L.: NNexus: Towards an Automatic Linker for a Massively-Distributed Collaborative Corpus, In Proceedings of the International Conference on Collaborative Computing: Networking, Applications and Worksharing, 17-20 November 2006, 1-3.
12. Norris, J. R.: Markov chains, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge: Cambridge University Press, 1999.
13. Ollivier Y., Senellart P.: Finding Related Pages Using Green Measures: An Illustration with Wikipedia, In Proceedings of the 22nd National Conference on Artificial Intelligence (AAAI'07), Vancouver, Canada, 22-26 July 2007.
14. Schönhofen P.: Identifying decument topics using the Wikipedia category network, In Proceedings of the 2006 IEEE/EIC/ACM International Conference on Web Intelligence (WI'06), Hong Kong, 18-22 December 2006.
15. Strube, M., Ponzetto, S. P.: WikiRelate! Computing Semantic Relatedness Using Wikipedia, In Proceedings of the 21th National Conference on Artificial Intelligence (AAAI'06), Boston, Massachusetts, USA, 16-17 July 2006, 16-20.
16. Gabrilovich, E., Markovitch, S.: Computing Semantic Relatedness using Wikipedia-based Explicit Semantic Analysis, In Proceedings of the 20^{th} International Joint Conference on Artificial Intelligence (IJCAI'07), Hyderabad, India, 6-12 January 2007.
17. Kumar, R., Raghavan, P., Rajagopalan, S., Tomkins, A.: Trawling the Web for Emerging Cyber-Communities, Computer Networks, 31(11-16), May 1999, 1481-1493.
18. Jeh, G., Widom, J.: SimRank: a measure of structural-context similarity, In Proceedings of the 8th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD'02), Edmonton, Canada, 23-26 July 2002, 538-543.
19. Kessler, M. M.: Bibliographic coupling between scientific papers, American Documentation, 14(10-25), 1963.
20. Dean, J., Henzinger, M. R.: Finding related pages in the World Wide Web, Computer Networks, 1999, 31(11-16), 1467-1479.
21. Kleinberg, J.: Authoritative sources in a hyperlinked environment, In Proceedings of the 9th Annual ACM–SIAM Symposium on Discrete Algorithms, San Francisco, CA, USA, 25-27 January 1998, 668-677.
22. Schamber, L.: Relevance and Information Behavior, Annual review of information science and technology 29, Medford, NJ: Information Today, 1994, 3-48.
23. Vakkari, P., Hakala, N.: Changes in relevance criteria and problem stages in task performance, Journal of Documentation, 56(5), September 2000, 540-562.
24. Kekäläinen, J., Järvelin, K.: Using graded relevance assessments in IR evaluation, Journal of the American Society for Information Science and Technology, 53(13), November 2002, 1120-1129.
25. Huang, W. C., Xu, Y., Geva, S.: Overview of INEX 2007 Link the Wiki track, In Pre-Proceedings of the INEX 2007 Conference, Dagstuhl, Germany, 2007.
26. Itakura, K. Y., Clarke, C. L. A.: University of Waterloo at INEX2007: Ad Hoc and Link-the-Wiki Tracks, In Pre-Proceedings of the INEX 2007 Conference, Dagstuhl, Germany, 2007, 380-387.

Appendix A

A.1 Example Submission

```
<inex-submission participant-id="12" run-id="LTW_01" task="LinkTheWiki">
  <details>
    <machine>
       <cpu>Intel(R) Pentium (R) D</cpu>
       <speed>3.00GHz</speed>
       <cores>2</cores>
       <hyperthreads>None</hyperthreads>
       <memory>2GB</memory>
     </machine>
     <time>166295 seconds</time>
  </details>
  <description>Using text chunking etc.</description>
  <collections>
    <collection>wikipedia</collection>
  <collections>
  <topic file="13876.xml" name="Albert Einstein">
    <outgoing>
      <link>
        <anchor>
          <file>13876.xml</file>
          <start>/article[1]/body[1]/p[3]/text()[2].10</start>
          <end>/article[1]/body[1]/p[3]/text()[2].35</end>
        </anchor>
        <linkto>
          <file>123456.xml</file>
          <bep>/article[1]/sec[3]/p[8]<bep>
        </linkto>
      </link>
      …
    </outgoing>
    <incoming>
      <link>
        <anchor>
          <file>654321.xml</file>
          <start>/article[1]/body[1]/p[3]/text()[2].10</start>
          <end>/article[1]/body[1]/p[3]/text()[2].35</end>
        </anchor>
        <linkto>
          <file>13876.xml</file>
          <bep>/article[1]/sec[3]/p[8]<bep>
        </linkto>
      </link>
      …
    </incoming>
  </topic>
</inex-submission>
```

These two sections of tags can be left as empty (***e.g.* <bep></bep>**) since the task is operated at the document level.

A.2 Official Result Set

Topics	# of Outgoing	# of Incoming	Topics	# of Outgoing	# of Incoming
Donald Bradman (87021.xml)	72	144	Dalai Lama (8133.xml)	71	237
Unified Modeling Language (32169.xml)	62	91	Within You Without You (1451526.xml)	13	11
Sukhoi Su-33 (552810.xml)	23	15	Software engineering (27010.xml)	107	404
Funk (10778.xml)	126	755	Philately (23681.xml)	41	108
Star Trek (26717.xml)	143	1649	Marie Curie (20408.xml)	75	127
Cartilage (166945.xml)	41	166	Stockholm syndrome (90910.xml)	49	36
Organic food (177593.xml)	73	50	Pink Floyd (24370.xml)	175	718
Pope Clement V (24102.xml)	69	56	Wavelet compression (50911.xml)	21	13
David (8551.xml)	124	513	Computer science (5323.xml)	241	1606
Aranyaka (321947.xml)	10	6	Pizza (24768.xml)	189	262
Greater Tokyo Area (354951.xml)	32	28	Joshua (16121.xml)	57	136
Xorn (322085.xml)	42	17	Skin cancer (64993.xml)	18	54
Kennewick Man (92818.xml)	47	10	Prince (artist) (57317.xml)	252	475
Frank Klepacki (752559.xml)	13	1	Family name (10814.xml)	165	474
University of London (60919.xml)	193	564	Search engine (27804.xml)	64	254
Latent semantic analysis (689427.xml)	16	10	Charleston, South Carolina (61024.xml)	200	947
Use case (300006.xml)	12	16	Elf (9896.xml)	235	378
Gout (55584.xml)	95	118	Akira Kurosawa (872.xml)	95	186
Thomas Edison (29778.xml)	132	358	Database (8377.xml)	99	186
Baylor University basketball scandal (493525.xml)	44	3	Radical feminism (25998.xml)	29	49
Search engine optimization (187946.xml)	49	45	Educational progressivism (10005.xml)	6	15
Civil Constitution of the Clergy (410450.xml)	40	34	Software development process (27565.xml)	49	33
Nokia (21242.xml)	48	196	Alastair Reynolds (69168.xml)	29	40
Achilles (305.xml)	124	219	Kazi Nazrul Islam (539155.xml)	31	20
Sunscreen (294419.xml)	38	46	Muammar al-Qaddafi (53029.xml)	159	149
Experiential education	16	17	Neo-Byzantine architecture	36	5

(447089.xml)			(1453013.xml)		
Yitzhak Rabin (43983.xml)	77	145	Waseda University (376791.xml)	67	85
Triple J's Impossible Music Festival (2542756.xml)	103	1	Text Retrieval Conference (1897206.xml)	9	2
World Wide Web Consortium (33149.xml)	23	181	Autism rights movement (1305330.xml)	86	27
Excel Saga (265496.xml)	74	73	Ballpoint pen (4519.xml)	53	55
Link popularity (210641.xml)	20	6	Digital library (8794.xml)	13	43
Coca-Cola (6690.xml)	171	506	Sloe gin (392900.xml)	13	7
Entertainment robot (1451221.xml)	17	3	Koala (17143.xml)	70	104
Indira Gandhi (15179.xml)	100	199	Billie Holiday (50420.xml)	53	196
Leukemia (18539.xml)	64	403	Softball (80763.xml)	50	368
Miss Universe (150340.xml)	159	182	Information retrieval (15271.xml)	40	45
Neuilly-sur-Seine (234647.xml)	18	80	Cheminformatics (575697.xml)	13	17
Jihad (16203.xml)	56	254	Requirement (544592.xml)	9	27
Google (1092923.xml)	192	541	Susan Haack (321979.xml)	27	10
Joseph Stalin (15641.xml)	373	1324	Math rock (221484.xml)	72	49
Seasonal energy efficiency ratio (2189642.xml)	8	0	Transportation in the Faroe Islands (10704.xml)	18	0
Sony (26989.xml)	136	965	Anthropology (569.xml)	129	808
Doctor of Philosophy (8775.xml)	64	2110	Red Bull (61123.xml)	75	74
Taiwanese aborigines (53787.xml)	68	86	Lithography (18426.xml)	32	281
Hyperlink (49547.xml)	60	118	Isaac Newton (14627.xml)	207	611