A Virtual Evaluation Track for Cross Language Link Discovery

Wei Che (Darren) Huang
Faculty of Science and Technology
Queensland University of Technology
Brisbane, Australia
w2.huang@student.qut.edu.au

Andrew Trotman
Department of Computer Science
University of Otago
Dunedin, New Zealand
andrew@cs.otago.ac.nz

Shlomo Geva
Faculty of Science and Technology
Queensland University of Technology
Brisbane, Australia
s.geva@qut.edu.au

ABSTRACT

The Wikipedia has become the most popular online source of encyclopedic information. The English Wikipedia collection, as well as some other languages collections, is extensively linked. However, as a multilingual collection the Wikipedia is only very weakly linked. There are few cross-language links or cross-dialect links (see, for example, Chinese dialects). In order to link the multilingual-Wikipedia as a single collection, automated cross language link discovery systems are needed – systems that identify anchor-texts in one language and targets in another. The evaluation of Link Discovery approaches within the English version of the Wikipedia has been examined in the INEX Link-the-Wiki track since 2007, whilst both CLEF and NTCIR emphasized the investigation and the evaluation of cross-language information retrieval. In this position paper we propose a new *virtual evaluation track*: *Cross Language Link Discovery* (CLLD). The track will initially examine cross language linking of Wikipedia articles. This virtual track will not be tied to any one forum; instead we hope it can be connected to each of (at least): CLEF, NTCIR, and INEX as it will cover ground currently studied by each. The aim is to establish a virtual evaluation environment supporting continuous assessment and evaluation, and a forum for the exchange of research ideas. It will be free from the difficulties of scheduling and synchronizing groups of collaborating researchers and alleviate the necessity to travel across the globe in order to share knowledge. We aim to electronically publish peer-reviewed publications arising from CLLD in a similar fashion: online, with open access, and without fixed submission deadlines.

Categories and Subject Descriptors: H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *Search Process*.

General Terms: Measurement, Performance, Experimentation

Keywords: Cross Language, Link Discovery, Information Retrieval, Evaluation

SIGIR'09, July 19–23, 2009, Boston, USA.

1. INTRODUCTION

1.1 Background

Collaborative hypertext knowledge management systems, such as the Wikipedia, offer an efficient means for creating, maintaining, and sharing information. Extensive linking between documents in these systems is essential for user navigation and assists readers to varying degrees. A reader with extensive background knowledge of a topic may be less likely to follow a link, while a less knowledgeable reader may choose to follow many links in order to expand their knowledge.

Links in the Wikipedia originate from two primary sources: the page authors' knowledge of the document collection; and automated link discovery systems such as *We Can Link It* [6] and *Wikify* [9]. The Link-the-Wiki track at INEX [22] was established as an independent evaluation forum for measuring the performance of these kinds of link discovery systems.

In 2007 the track explored document-to-document link discovery in the English Wikipedia, in 2008 the track also looked at anchor-text identification within the source document and the placement of the target point (the best-entry-point, or BEP) within the target document. This second kind of link discovery is known as anchor-to-BEP link discovery, or focused link discovery. The track also developed a standard methodology and metrics for the evaluation of link discovery systems.

The track's results show that excellent link discovery systems have been developed – that is, there are now published algorithms that can almost perfectly predict the links in a Wikipedia page. However, manual assessment revealed the highly unexpected result that many existing Wikipedia links are not relevant (at least not to the INEX assessors)! INEX now recommends manual assessment as the preferred procedure for the evaluation of link discovery systems. We note that an INEX assessor manually assessing links from the pool perfectly models a user who (after adding a new article to the Wikipedia) is navigating a list of links recommended by a link discovery system – accepting or rejecting as they go. This process lends itself to the interactive study of link discovery systems.

With the growth of the multilingual Wikipedia (and the multilingual web) there is a growing need for cross-language information retrieval including cross language interlinking of multilingual documents. Most Wikipedia pages are written in English and we, unsurprisingly but anecdotally, observe users whose first language is not English searching the English Wikipedia. Their need is two-fold: the Wikipedia documents to be translated into their first language; and links between documents

to reflect their language choices. Translation is already happening and some cross-language links already exist, however these problems are our research motivation. We are trying to:

- Identify Wikipedia documents that are all on the same topic, irrespective of language, and
- Identify hypertext links between documents, irrespective of language, so that a user can choose a target document based on a language preference.

1.2 Motivation

Many Internet users are multi-lingual. To satisfy their information needs the search engine should return documents in the different languages they read. Doing so is more thorough than returning results in just one language. As examples, the *Early history of the United States of America* can be found in the Chinese Wikipedia but the English Wikipedia has a much richer document on the topic; information about *Chinese Dynasties* may be found in several documents in the Chinese language Wikipedia, and indeed in several distinct Chinese language version of the Wikipedia. In both examples, a link between these different language versions will help the multi-lingual reader. In both examples focused cross-lingual anchor-to-BEP links would result in a more comprehensive interlinked knowledge base, especially if the links the multilingual reader sees are based on a personal language profile. Envisage a document being interlinked to any number of languages, but users only seeing links to languages that are defined in their personal profile.

Anchor-to-BEP linking is a feature of HTML that is rarely exploited in links – despite its existence since the beginnings of the web. Very few links in the Wikipedia actually take the user from the point of reference (the anchor) to the target location within another document (the BEP). Such interlinking is common within a single document and is used in navigation, but is rarely utilised when linking between documents. Such *focused* interlinking is particularly desirable when documents are large or when browsing on small mobile devices. For instance, in the article *South Eastern Main Line*, an orphaned anchor, *Folkestone Harbour*, is colored in red. It is a place-holder for a link to an article that does not yet exist. However, the article *Folkestone* does have a section titled *Folkestone Harbour*. This prospective anchor could be linked to this section until an article on Folkestone Harbour is created.

Extending the INEX Link-the-Wiki track to cross language linking will help turn the Wikipedia into a multi-lingual knowledge network. The section 地理 (English: *Geography*) in the article 英国 (English: *England*) has two anchor texts, 多佛港 (English: *Dover Harbour*) and 英吉利海峡隧道 (English: *Channel Tunnel*; French: *Le tunnel sous la Manche*). There is no link for 多佛港 (English: *Dover Harbour*) in the Chinese Wikipedia, but an article on the *Port of Dover* is linked from the redirect of the *Dover Harbour* page in the English Wikipedia. Information (images and geography) about Dover Harbour can also be found in the English article on *Dover*. The article, 英吉利海峡隧道, does not express much information about the channel tunnel, certainly not as much as the English *Channel Tunnel* page. These two examples show the need for cross-language links within the Wikipedia.

To the best of our knowledge, current link discovery systems such as *Wikify* [9] focus on monolingual Wikipedia and have not been extended to support multilingual link discovery. Cross-language tracks conducted in NTCIR and CLEF explore Information Retrieval and Question-Answering but not link discovery. Link Discovery is different from Information Retrieval although it does rely on similar technology: for link discovery a match of semantic context between the point of reference (the context of the anchor) and the target text (the BEP context) is essential. Unlike query based information retrieval, in link discovery the context of the anchor is always explicit since the anchor is always embedded in surrounding text, and similarly the context of the target location is highly focused and specific. In information retrieval the query is known but the context unknown, in link discovery it is necessary to identify both the query (the anchor-text) and the results list (the target document and BEP) and embedding contexts are available.

Link discovery provides a rich context in which NLP based approaches may well prove much more useful than they had been in the query based information retrieval. Furthermore, cross-language link discovery involves a set of technologies, including IR, NLP, semantic and similarity matching techniques, character encoding technologies, machine readable corpora and dictionaries, machine translation, focused and passage retrieval, and multiple links per anchor discovery. Cross Language Link Discovery (CLLD) demands the tight integration of techniques currently under examination at INEX, CLEF and NTCIR.

Herein we formally propose the CLLD track. This track will be run as a single collaborative web-based forum. Participants will be drawn from the existing forums, but be part of none (or all), it will be an online *virtual evaluation forum*. All collections, topics, submission and result analysis will be maintained via a remote repository. By using only public domain data (such as the Wikipedia and open source software) we can simplify participation and the sharing of resources. The community of participants will provide software tools and assessments; as well as a peer-reviewed online publication for approaches and results. The forum will not be tied to any particular timeline or venue but will be run as a continuous evaluation track – without a dedicated annual event (although there is no reason not to hold such meetings, perhaps as surrogate to larger events). This proposal represents a dramatic philosophical change from the traditional TREC paradigm.

2. RELATED WORK

As suggested by Wilkinson & Smeaton [1], 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 in related information. However, since digital resources are distributed it has become difficult 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 [2]. 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 [3, 4, 5]. These approaches were

focused on the document-to-document linking scenario, rather than identifying which parts of which documents were related.

Jenkins [6] developed a link suggestion tool, *Can We Link It*. This tool extracts a number of anchors which have not been discovered in the current article and that might be linked to other Wikipedia documents**.** 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 a Wikipedia article.

A collaborative knowledge management system, called *PlanetMath*, based on the *Noosphere* system has been developed for mathematics [7]. It is encyclopedic, (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. Based on the Noosphere system, *NNexus* (Noosphere Networked Entry eXtension and Unification System) was developed to automate the process of the automatic linking procedure [8]. This was the first automatic linking system which eliminates the linking efforts required by page authors.

The *Wikify* [9] system which integrates technologies of automatic keyword extraction and word sense disambiguation can identify the important concepts in a document and link these concepts to corresponding documents in the Wikipedia. Mihalcea and Csomai stated that many of applications such as the annotation of semantic webs and academic systems can benefit from this kind of system.

Since the inception of TREC in 1992 interest in IR evaluation has increased rapidly and today there are numerous active and popular evaluation forums. It is now possible to evaluate a diverse range of information retrieval methods including: ad-hoc retrieval, passage retrieval, XML retrieval, multimedia retrieval, question answering, cross language retrieval, link discovery, learning to rank, and so on.

The CLIR (Cross-Language Information Retrieval) track was first introduced to TREC in 2000. It offered document collections in English, French and German and queries in English, French, German, Spanish and Dutch. Three fundamental resources, machine translation, machine readable dictionaries and corpus-based resources, have been used. There are three common approaches to match queries with the resource documents [10]: machine translation technology using dictionaries and statistical information or example-based translation; machine readable bilingual dictionaries; and relying on corpus resources to train retrieval mechanism by using Latent Semantic Indexing (LSI), Generalized Vector Space Model or Similarity Thesauri for translating queries. As a performance baseline corresponding language queries were also submitted against the same language collections.

The cross-language track investigated the retrieval of relevant documents that pertain to a particular topic or query regardless of the language in which both the topic and documents were written. The last TREC CLIR track was run in 2002, whoever ongoing effort can be found in both NTCIR and CLEF [11].

NTCIR started in late 1997 and is focused on Japanese and other East Asian languages [12]. The workshop runs on an 18-month cycle. The aim is to build an infrastructure for large-scale experimental evaluation of Information Access (IA) research. IA in the workshop has been indicated as the process of searching, browsing and looking for relevant information, and utilizing the information. Technologies, like Information Retrieval (IR), Cross-Language Information Retrieval (CLIR), Question-Answering (QA), text summarization and text mining, are considered part of the IA family.

The goal is to develop a module-based infrastructure for evaluation integrating IR and QA technologies to propose answers in a suitable format for given questions in any language. It intends to model significant work from every participant and build a set of APIs (or modules) to facilitate the development of cross-language (QA) systems. A platform, called EPAN (Evaluation Platform for ACLIA and NTCIR), was adopted by NTCIR to perform the collaborative evaluation [13]. Through module-by-module evaluation it is possible to identify problems in parts of participants' otherwise complicated CLIR-QA systems – something not possible in end-to-end evaluation. For example, many CLIR-QA systems failed to retrieve relevant documents when named entities in queries did not appear in ordinary translation dictionaries. The module-based approach also makes it possible for participants to collaborate by working on different modules.

The annual CLEF forum aims to create a research forum in the field of multilingual system development [14]. The experiments range from monolingual text retrieval to multilingual multimedia retrieval. The collection and available languages vary depending on different tasks. They include 3 million News articles in 13 languages, a social science database in English and German, the Cambridge Sociological Abstracts, and the Russian ISISS collection, 3.5 million web pages in multi-languages, and a photograph database with captions in different languages [15]. Various sets of topics in different languages are available for respective tasks.

The DIRECT system used by CLEF manages data (collection, topics, and metrics), builds statistics (analysis, plots and results) and provides different entries for various roles [16, 17]. Of particular note is the dynamic user interface through which participants can interact with their-own and others' experimental data and results.

3. MULTILINGUAL WIKIPEDIAS

The Wikipedia is a multilingual online encyclopedia that offers a free and flexible platform for developing a collaborative knowledge repository [18]. Currently, it has entries written in more than 200 different languages. Overell [19] shows that the geographic coverage of the Wikipedia very much depends on the language version – places in the UK are best covered by the English language version of the Wikipedia while places in Spain are best covered by the Spanish language version. There are more than 2,874,919 articles (May 2009) in the English Wikipedia which is the largest language version in the Collection. By the end of 2008, no fewer than 13 languages have more than 200,000 articles. As can be seen in Table 1, both European and Asian language versions have reached a substantially size – the collection is already useful for multilingual research [20].

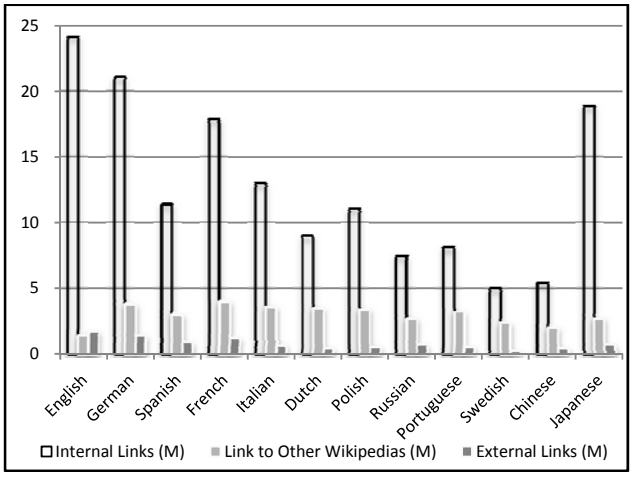
Table 1: Language subsets of the Wikipedia at end 2008

	Articles (K)	Database (GB)	Bytes per Article	Over 2KB
English*	1100	3.2	3092	363 (33%)
German	858	3.5	3489	420 (49%)
French	746	2.7	2995	269 (36%)
Polish	567	1.5	1988	130 (23%)
Japanese	555	2.6	1890	128 (23%)
Italian	533	1.9	2914	192 (36%)
Dutch	507	1.3	2021	147 (29%)
Portuguese	448	1.1	1866	90 (20%)
Spanish	430	1.8	3492	189 (44%)
Russian	347	1.9	2822	125 (36%)
Swedish	301	0.609	1535	54 (18%)
Chinese	209	0.717	1451	36 (17%)
Norwegian	203	0.475	1781	43 (21%)

* WikiMedia state the English version as at Oct. 2006!

Most documents have a rich set of same-language links but a much lower number cross-language links. Figure 1 shows that by October 2006 (WikipMedia have not provided stats for English since that date) only 1.4 million English anchor texts have been linked to entries in other languages while the amount of same-language links in the English Wikipedia had reached 24.2 million.

Around a third of anchor texts in the Chinese Wikipedia were linked to other language collections in the Wikipedia. There are several different Chinese dialects available as different collections and many of English terms are linked to the English Wikipedia. Other versions shown in Figure 1 have less than 20 percent of their links pointing to other languages. For the Article-to-Article, Sorg and Cimano state that around 50% of Wikipedia articles in German version are linked to the English Wikipedia articles whilst only 14% of English articles have links to the German version [21]. Linking the Wikipedia entries across different languages is still very limited. Efficient and accurate cross language link discovery is yet to be demonstrated and evaluated.



* The statistics of the English version was dated in Oct. 2006.

Figure 1: Different linking types in different language versions of the Wikipedia at the end of 2008

In order to achieve a comprehensive cross-language knowledge network, cross language link discovery is essential. The document collection is large (there are many millions of documents) and broad (there are many languages covered). Although only one link per anchor is typically displayed by existing HTML based web browsers, there is no inherent restriction for this limit in the HTML standard. Anchor text can be linked not only to multiple targets, but also in different languages, and the extension of browsers to support this functionality is long overdue.

4. INEX *LINK-the-WIKI* TRACK

The INEX Link-the-Wiki track offers a standard forum for the evaluation of link discovery in both document-to-document and anchor-to-BEP linking. The task is to discover a set of anchor texts, based on the context of the topic, which can semantically be linked to best entry points in other documents. Besides outgoing links, candidate incoming links from other articles into the topic document are also required.

The INEX 2008 English Wikipedia collection consisting of 659,388 documents was used as the corpus for the experiments since Wikipedia is composed of millions of interlinked articles in numerous languages and have been proved as a valuable collection for IR research. 50 documents nominated by participants were used in the anchor-to-BEP task. For the document-level link discovery, 6,600 documents were randomly selected but pre-filtered (for suitability) by size and the number of links. For 2009 this collection has been updated and now consists of over 2.6 million articles, spanning over 60GB of text (without images).

The documents were converted into topics by removing all links in the documents and removing all links from the remainder of the collection to those documents, they were said to have been orphaned from the collection. The original documents (with the links) were said to be pre-orphans.

Links within the collection to and from the pre-orphans were extracted and used as the ground truth to which runs were compared. This is the automatic evaluation method. Using standard methodology, the topics were sent to participating groups. Participant's link discovery systems were required to return a *ranked* list of at most 50 outgoing text anchors (each of which targeted up-to 5 documents). The results were exhaustively pooled and the ground truth of existing links in the Wikipedia was added to the pools. Pools were manually assessed to completion, by the groups that nominated the topics. This formed the manual evaluation set. Assessment against the ground-truth result set generated a score for the performance of a submission relative to the Wikipedia. Additional evaluation was performed against the manually assessed links (including the manually assessed existing Wikipedia links). The outcome showed that the Wikipedia ground-truth does not agree with user expectation. The manual assessment process is necessary in order to produce a reliable test base for evaluation [22].

5. PROPOSED CLLD METHODOLOGY

The cross language link discovery track will work in a similar manner to the INEX track. Participating groups will be asked to submit a list of languages they can read (from those covered by the Wikipedia). This also indicates the languages these participants can submit runs for and can assess in. A subset of topics from the different language Wikipedia collections will be chosen and distributed to participants. To support the creation of topics, a selection tool will be developed to help choose and to orphan documents in indicated languages. Participants will be able to choose any combination of languages, for example

German documents linking to English or Dutch documents, or Chinese documents linking to Japanese documents.

Programs will be provided that allow participants to view their runs. These programs will show proposed linked documents with the anchors and their respective target document best-entry-points, as would be seen by an assessors (and ultimately end-users of their system).

The challenge for the organizers is to obtain a critical mass of participants and assessors to facilitate robust and reliable manual evaluation in multiple languages. The track must, therefore, be a close and extensive collaboration between NTCIR, CLEF, INEX, and other evaluation forums.

5.1 Tasks Specification

Initially two linking tasks will be formalized:

- MULTILINGUAL topical linking:
  This is a form of document clustering – the aim is to identify (regardless of language) all the documents in all languages that are *on the same topic*. The Wikipedia currently shows these links in a box on the left hand side of the page.
- BILINGUAL anchor linking:
  It is exemplified by the Chinese article 诺森伯兰郡, having a link from the anchor 国会选区 to the English article *List of United Kingdom Parliament constituencies*. The link discovery system must identify the anchor text in one language version of the Wikipedia and the destination article within any other language version of the Wikipedia.

In the case of MULTILINGUAL topical linking, the participants are encouraged to discover as many documents as they can.

In the case of BILINGUAL anchor linking, at most 50 anchors may be identified in a orphaned document and up to 5 BEPs may be linked to one target language (e.g. English to German). Initially only outgoing links will be examined since incoming links from a single language may not make sense.

5.2 Test Collections and Submission

The set of multilingual Wikipedia collections will be used as the corpus for cross language link discovery. The size and the number of documents are listed in Table 1. Nominated topics will be collected and the ground-truth extracted from the collection.

Participants will be encouraged to share in the development of appropriate procedures for topic selection, multilingual topic discovery, ground truth link extraction, and the assessment method.

The submission format may be derived from the format currently used by INEX. The existing INEX tools will be ported to support CLLD.

5.3 Evaluation

It is essential to define a standard methodology and procedure to assess link quality and to quantitatively evaluate different approaches.

5.3.1 Static Evaluation

When Trotman & Geva [24] introduced the Link-the-Wiki track at INEX they also noted that the evaluation required no human assessment. The same is true with cross-language link discovery.

Topics in the INEX Link-the-Wiki track are chosen directly from the document collection. All links in those documents are removed (the document is orphaned). The task is to identify links for the orphan (both from and to the collection). Performance is measured relative to the pre-orphan (the document before the links were removed).

For MULTILINGUAL linking the links on the left hand side of the Wikipedia page could be used as the ground truth. The performance could be measured relative to the alternate language versions of the page already known to exist.

BILINGUAL anchor linking from one document to another could also be automatically evaluated. Links from the pre-orphan to a destination page in an alternate language would be used as the ground truth – but there are unlikely to be many such links.

A same-language link from a pre-orphan to a target provides circumstantial evidence that should the target exist in multiple languages then the alternate language versions are relevant. This is essentially a triangulation: $A \xrightarrow{t} B \xrightarrow{l} C \Rightarrow A \xrightarrow{tl} C$ where *A*, *B*, and *C* are documents; and *t* designates a topical link, *l* a cross language link, and *tl* a topical cross language link. By extension, if $A \xrightarrow{t} B \xrightarrow{T} C \Rightarrow A \xrightarrow{tT} C$, where $B \xrightarrow{T} C$ designated two documents that are not linked but have the same title.

Static assessment requires no human interaction. A web site with orphan sets chosen using some criteria (perhaps randomly), with the assessment sets (extracted from the pre-orphans), and that will evaluate a run will be built and provided. Such an evaluation methodology raises the possibility of running the track continually and without any deadlines.

Huang *et al*. [22] question automatic evaluation. Their investigation suggests that many of the links in the Wikipedia are not topical, but are trivial (such as dates), and that users do not find them useful. Manual assessment is, consequently, necessary. Although the automatic evaluation method in the INEX Link-the-Wiki is less accurate, it is still very practical and provides a reliable way to evaluate different link discovery methods against the ground truth. The submission runs that have better performance in automatic evaluation tend to have better results in manual evaluation.

5.3.2 Continual Evaluation

Manual assessment raises new challenges for cross language link discovery because finding assessors fluent in multiple languages is difficult – especially for a track with a relatively small number of participants but in a large number of languages (the Wikipedia has over 200 languages).

We propose a novel form of evaluation called *continual evaluation* in which participants can download topics and submit runs at any time; and in which contribution to manual assessment is an on-going concern. The document collection will, initially, be static. Topics will either be chosen at random from the collection, or nominated by participants. For any given run a participant will download a selection of topics and submit a run. The evaluation will be based on metrics that consider the un-assessed document problem (such as a variant on rank-biased precision [23]), and comparative analysis will be relative to an incomplete, but growing, assessment set.

To collect assessments two methods are proposed: first, in order to submit a run the participant will be required to assess some anchor-target pairs in languages familiar to them; second, we will run an assessment Game With A Purpose (GWAP). Kazai et al. used a GWAP for the INEX Book track; Von Ahn & Dabbish [25] discuss GWAPS in other contexts (including the Google Image Labeler). Regardless of the method of assessment collection, we are trying to validate the minimum number of links necessary to disambiguate the relative rank order of the runs (within some known error).

6. PUBLICATION

Both automatic and manual assessment of cross language link discovery can be done on a continual rolling basis; there is no need for topic submission deadlines, run deadlines, assessment deadlines; or paper publication deadlines.

At INEX the time difference between run-submission and the workshop paper submission date is long (6 July – 23 Nov). With automatic assessment it is possible to achieve a result, write, and then publish a paper with a short turn around. As part of the virtual track we propose an open-access virtual workshop workbook to which registered participants can immediately submit their papers for peer-review and publication.

7. CONCLUSIONS AND OUTLOOK

As far as we are aware, the cross-language link discovery track is the first to offer extensive reusable independent evaluation resources. In this paper we introduce this new evaluation task.

A fully automated procedure for anchor-to-document link analysis, using the existing Wikipedia linking network is described. The procedure was used at INEX 2007 and allowed us to create a fast evaluation procedure with a turnaround time of days and not months because it had no manual assessment. The procedure allows for a very large number of documents to be used in experiments. This overcomes the assessment bottleneck which is encountered in most other tasks in collaborative evaluation forums such as INEX and TREC. We further proposed to extend the task to Cross Language Link Discovery, and propose the concept of automatic evaluation. We describe the requirement for evaluating such a task.

These activities may not be held in a fixed place but can be done by gathering participants from INEX, CLEF and NTCIR through a virtual web-based system. The CLLD track will be dedicate to supporting efficient methods and tools for CLLD evaluation. The collections, submission and result data will be well managed for further analysis and experiments. Participants from different nations are expected to work collaboratively to achieve the development of multilingual link discovery systems.

Baseline automatic evaluation methods seen at INEX do not require human intervention as the assessments are extracted directly from the collection and performance is measured relative to these. The new track can, therefore, bootstrap and run online with continuous evaluation, free from the problems of scheduling groups of collaborating researchers. Overtime manual assessments will be collected and improve the available resources. We also propose to publish the results of the track in a similar fashion to the CLLD track itself – online, with open access, and with quality control.

8. REFERENCES

[1] Wilkinson, R. and Smeaton, A. F., *Automatic Link Generation*, ACM Computing Surveys, 31(4), December 1999.

[2] Green, S. J., *Building Hypertext Links By Computing Semantic Similarity*, IEEE Transactions on Knowledge and Data Engineering, September/October 1999, 11(5), pp. 713-730.

[3] Allan, J., *Building Hypertext using Information Retrieval*, Information Processing and Management, 33(2) pp. 145-159.

[4] Green, S. J. (1998) *Automated Link Generation: Can We Do Better than Term Repetition?*, In Proceedings of the 7th International World Wide Web Conference, Brisbane, Australia, pp. 75-84.

[5] Zeng, J. and Bloniarz, O. A. (2004) *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, pp. 283-286.

[6] Jenkins, N., *Can We Link It*, http://en.wikipedia.org /wiki/User:Nickj/Can_We_ Link_It.

[7] 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.

[8] Gardner, J., Krowne, A. and Xiong, L., *NNexus: Towards an Automatic Linker for a Massively-Distributed Collaborative Corpus*, In Proceedings of the International Conference on Collaborative Computing, 17-20 November 2006, pp. 1-3.

[9] Mihalcea, R. and A. Csomai, *Wikify!: linking documents to encyclopedic knowledge*, CIKM 2007, pp. 233-242.

[10] Schäuble, P. and Sheridan, P., *Cross-Language Information Retrieval (CLIR) Track Overview*, In Proceedings of the Sixth Text Retrieval Conference, pp. 31-44.

[11] Agosti, M., Di Nunzio, G. M., Ferro, N., Harman, D. and Peters, C., *The Future of Large-Scale Evaluation Campaigns for Information Retrieval in Europe*, LNCS 4675, 2007, pp. 509-512.

[12] Kando, N., *Overview of the Seventh NTCIR Workshop*, In Proceedings of NTCIR-7 Workshop Meeting, 16-19 December 2008, Tokyo Japan, pp. 1-9.

[13] Mitanura, T., Nyberg, E. Shima, H., Kato, T., Mori, T., Lin, C.Y., Song, R., Lin, C. J., Sakai, T., Ji, D. and Kando, N., *Overview of the NTCIR-7 ACLIA Taska: Advanced Cross-Language Information Access*, In Proceedings of NTCIR-7 Workshop Meeting, 16-19 December 2008, Tokyo Japan, pp. 11-25.

[14] Ferro, N. and Peters, C., From *CLEF to TrebleCLEF: the Evolution of the Cross-Language Evaluation Forum*, In Proceedings of NTCIR-7 Workshop Meeting, 16-19 December 2008, Tokyo Japan, pp. 577-593.

[15] Di Nunzio, G. M., Ferro, N., Mandl, T. and Peters, C., *CLEF 2007: Ad Hoc Track Overview*, CLEF 2007, LNCS 5152, pp. 13-32.

[16] Di Nunzio, G. M. and Ferro, N., *DIRECT: A System for Evaluating Information Access Components of Digital Libraries*, ECDL 2005, LNCS 3652, pp. 483-484.

[17] Dussin, M. and Ferro, N., *Design of the User Interface of a Scientific Digital Library System for Larger-Scale Evaluation Campaigns*, Post-proceedings of the Fourth Italian Research Conference on Digital Library Systems (IRCDL 2008), pp. 105-113.

[18] Wikipedia, the free encyclopedia, 2009. http://wikipedia.org/.

[19] Overell, S.E., *Geographic Information Retrieval: Classification, Disambiguation and Modelling*, in *Department of Computing*. 2009, Imperial College London: London. pp. 175.

[20] Wikipedia Statistics, Wikipedia, the free encyclopedia, 2009, http://stats.wikimedia.org/EN/Sitemap.htm.

[21] Sorg, P. and Cimiano, P., *Enriching the Cross-lingual Link Structure of Wikipedia – A Classification-Based Approach* -, In Proceedings of the AAAI 2008 Workshop on Wikipedia and Artificial Intelligence, June 2008.

[22] Huang, W.C., A. Trotman, and S. Geva, *The Importance of Manual Assessment in Link Discovery*, in *SIGIR 2009*. 2009, ACM Press: Boston, USA.

[23] Moffat, A. and J. Zobel, *Rank-biased precision for measurement of retrieval effectiveness*. ACM Trans. Inf. Syst., 2008. 27(1):1-27.

[24] Trotman, A. and S. Geva. *Passage Retrieval and other XML-Retrieval Tasks*. In SIGIR 2006 Workshop on XML Element Retrieval Methodology. 2006. Seattle, Washington, USA.pp. 43-50.

[25] von Ahn, L. and L. Dabbish, *Designing games with a purpose*. Commun. ACM, 2008. **51**(8):58-67.