Cross-Lingual Information Access and Its Evaluation

Noriko Kando
Research and Development Department
National Center for Science Information Systems (NACSIS), Japan
URL: http://www.rd.nacsis.ac.jp/~{ntcadm,kando}/

Abstract

This paper introduces the outline of the first NTCIR Workshop[1] [1], August 30 - September 1, 1999, which is the first evaluation workshop designed to enhance research in Japanese text retrieval and cross-lingual information retrieval, then suggests some thoughts on the future directions of cross-lingual information access in the research and development of digital libraries.

Keywords: Cross-Lingual Information Retrieval, NTCIR, Large-Scale Lexicon, Evaluation

1. Introduction

1.1 Cross-Lingual Information Retrieval

Cross-lingual information retrieval (CLIR) is a search that a user submit a query in a language and the system retrieves documents which may represented in other languages as well as the documents in the query language. As shown in Fig.1, the term "multilingual information retrieval" may represent CLIR in this sense, but it may represent a paired single language search. To avoid such ambiguity, the term "Cross-Lingual" (or, Cross-Language, or Cross-Linguistic) is used in the IR community.

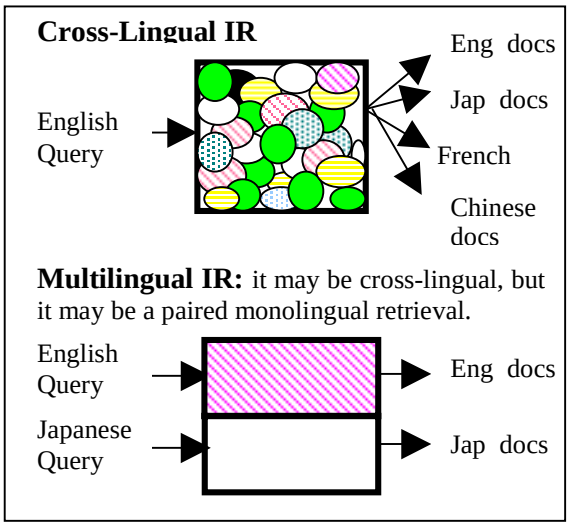

Fig.1 Cross-Lingual *vs* Multilingual

Importance of CLIR in digital libraries is extremely increasing in the Internet environment. It is more practical than translating every documents into every other languages. A user can issue a query in the most fluent language and avoid multiple queries in each language. It would be useful for people who are not fluent enough to create a query in a foreign language, but who can read the language to understand document contents and judge their relevance. CLIR also helps to reduce the cost of manual translation by discarding irrelevant documents before translation. Translating retrieved documents may be unnecessary since (1) users may be able to read other languages; (2) relevant information objects may be non-textual ones, such as photos, paintings, etc.; (3) to know the existence of the documents may be sufficient for the users, such as bibliometrics, visualization, etc.; and (4) translations may be provided, and so on.

Scholarly or scientific documents are one of the focuses of research and implementation of the digital libraries in these days. CLIR technique is critical in information access and delivery of Japanese scholarly information since scholarly information produced in Japan is substantially multilingual in both document and lexical levels.

For example, a journal may contain both English articles and Japanese articles. Japanese papers often contain English abstracts and author-given-keywords. In Japanese documents, a concept may represented in four different forms; Japanese term, English term in original spelling using roman alphabet, English acronyms, and transliterated form of English term using Japanese phonetic characters of Katakana. Especially technical terms or new concepts in Japanese papers are quite often expressed in English or other foreign languages with original spelling, and they are important as search keys. Word mismatch caused by the variation may decrease the search effectiveness and to overcome the problems, CLIR technique is needed[2].

1.2 Evaluation Workshop

An "evaluation workshop" usually provides a set of data usable for experiments and unified procedure of evaluation of experiment results. Each participating research group conducts research and

[1] This project is supported by “Research for the Future” Program JSPS-RFTF96P00602 of the Japan Society for the Promotion of Science

experiments using the data provided in various approaches. The first and one of the most successful examples of the evaluation workshops in information retrieval is the workshop series called Text Retrieval Conference (TREC) organized by National Institute of Standards and Technology (NIST) in the United States [4] since 1992. Importance of large-scale standard test collections in IR research has been widely recognized. Providing a large scale test collection, which is equivalent to the size of documents data searched in the operational setting, and unified evaluation procedures, various new techniques have been developed through TREC, and both exchanging research ideas among research groups and technology transfer from research laboratories to commercial products have been facilitated so much.

Regarding evaluation of CLIR, TREC has hosted Cross Language Retrieval Subtrack since 1996. CLIR between English and European languages such as Spanish, French, German, Italian has been investigated and English-Mandarin Chinese is planed for the next year, 2000. It has leaded the research and development of techniques of CLIR. Topic Detection and Tracking (TDT) [5] uses newswire documents on both English and Chinese. Fundamental IR procedures like stopping, stemming, and query analysis contains language dependent procedures. Therefore to test IR systems which can handle Japanese documents, a large-scale test collection containing Japanese documents is needed.

In the followings, the next section envisages the outline of the first NTCIR Workshop, Section 3 introduces the research on CLIR at NACSIS, and Section 4 discusses the research trends towards cross-lingual information access.

2. NTCIR Workshop

2.1 Overview of the NTCIR Workshop 1

The First NTCIR Workshop was held on August 30-31, 1999, in Tokyo. The participation to the Workshop was limited to the active participants, *i.e.* the members of the research groups that submitted the results of the tasks, advisors and members of the organizing group. It had fifty-two participants and ended in great enthusiasm. The NTCIR/IREX Joint Workshop and IREX Workshop, the another evaluation workshop of IR and information extraction (named entity) using Japanese newspaper articles were held consecutively.

The purposes of the NTCIR Workshop are as follows;

(1) to encourage research in information retrieval, CLIR and related areas by providing a large-scale Japanese test collection and a common evaluation setting that allows cross-system comparisons

(2) to provide a forum for research groups interested in comparing results and exchanging research ideas or opinions in an informal atmosphere

(3) to investigate methods for constructing large-scale test collections and IR laboratory-type testing, and create large scale test collections based on the investigation.

The test collection used in this Workshop is called “NACSIS Test Collection 1” or “NTCIR-1” and consists of more than 330,000 documents, with more than half presented as English-Japanese pairs. Although there is a Japanese test collection called BMIR-J2 consisting of 5,080 newspaper articles [6] and its contribution to IR research in Japan is tremendous, enhancement of the Japanese test collection in both variety of text types and scale was needed. We place emphasis on CLIR since it is critical in the Internet environment and for Japanese scientific information retrieval as discussed in Section 1.

Thirty-one groups, including participants from six countries, have enrolled in the first NTCIR Workshop. Among these, 28 groups have enrolled in IR tasks (23 in the Ad Hoc Task and 16 in the Cross-Lingual Task) and nine in the Term Recognition Task.

Regarding IR tasks, twenty-three groups submitted the search results. The search results of 48 runs were submitted for the Ad Hoc Retrieval Task from 17 groups and 69 runs for the Cross-Lingual Retrieval Task from 10 groups. Nine groups are from Japanese companies, 11 are from Japanese universities or national research institutes, and four are non-Japanese groups. Two groups are from the United States, one group is from Australia, one group is from Taiwan and 19 groups are from Japan; some of this latter group have non-Japanese members or have collaborated with research groups outside Japan. Two groups worked without any Japanese language expertise.

2.2 The Tasks

Each participant has conducted one or more of the following tasks.

The Ad Hoc Information Retrieval Task: to investigate the retrieval performance of systems that search a static set of documents using new search topics

The Cross-Lingual Information Retrieval Task: an *ad hoc* task in which the documents are in English and the topics are in Japanese

The Automatic Term Recognition and Role Analysis Task: (1) to extract terms from titles and

abstracts of documents, and (2) to identify the terms representing the "object", "method" and "main operation" of the main topic of each document

The Ad Hoc Task is also substantially cross-lingual since in the Task, as shown in Fig. 2, JE collection, mixture of Japanese documents and English documents, is used since in Japanese operational IR environments, especially for retrieval of scholarly, scientific or technical documents and World Wide Web documents, retrieving both Japanese and English documents at the same time is quite natural. The E Collection, English document collection, is used in the Cross-Lingual Task. The J Collection, Japanese document collection, is used in monolingual retrieval, which serves as a baseline for comparing the search effectiveness with the results in the Cross-Lingual Task.

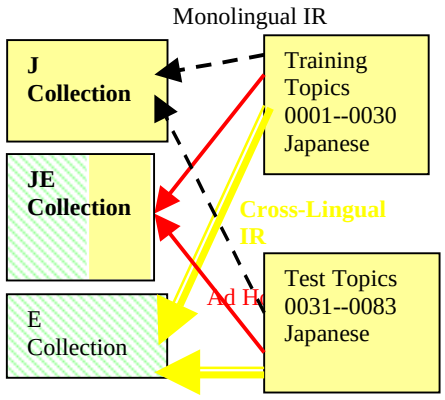

Fig 2. Relation of the Document Collections and Tasks

The Procedures

From November 1, 1998, delivery to each IR tasks-participant of the document data, 30 search topics (0001–0030) and their relevance judgments began, so the participants could train their systems. Among them, 21 topics were used as cross-lingual topics. The topics were written in Japanese. Fifty-three new test topics (0031–0083) were distributed to participants on February 8, 1999, and the search results for these new topics were submitted from each participant by March 4 as official test runs. The test topics are common for both IR tasks.

A participant could submit the results of more than one run. Both automatic and manual query constructions were allowed. In the case of automatic construction, the participants were required to submit at least one set of results for searches using the <DESCRIPTION> fields of the topics only as the mandatory runs as baseline of cross-system comparison. For optional automatic runs and manual runs, any fields of the topics could be used. In addition, each participant was required to complete and submit a form describing the detailed features of their system.

Top ranked documents of each topic were collected from each run and formed document pools of candidates of relevant documents. Human analysts assessed the relevance of documents in the pools for each topic. Based on the assessments, interpolated recall and precision at 11 points, average precision (non-interpolated) over all relevant documents, and precision at 5, 10, 15, 20, 30, 100 documents were calculated using TREC's evaluation program, which is available from the ftp site of Cornell University.

2.3. The Test Collection

The test collection used in the Workshop, "Test Collection 1" or "NTCIR-1", consists of documents, topics, and relevance assessments for each search topic.

Documents

The documents are author abstracts of conference papers that were presented at academic meetings hosted by 65 Japanese academic societies [7]. Subject domain, length and format of the documents are diversified according to each society. A wide range of subject domains from pure sciences, technology and engineering, social sciences and humanities is included, although about half of the documents are from electronic engineering and computer sciences.

The Collection contains three document collections: the JE Collection; the J Collection; and the E Collection. The JE Collection contains 339,483 documents, of which more than half are English-Japanese paired, the most of the rest are Japanese, and rather small number of documents are English documents. The author(s) of the document prepared both Japanese and English parts. The J and E Collections are constructed by extracting the Japanese or English parts, respectively, of the documents in the JE Collection.

Documents are plain text with SGML-like tags. A record may contain document ID, title, a list of author(s), name and date of the conference, abstract, keyword(s) that were assigned by the author(s) of the document, and the name of the host society.

Topics

A topic is a formatted description of a user's information need. We defined the topics as statements of "user need" rather than "queries", which are the strings actually submitted to the system, since we would like to allow both manual and automatic query construction from the topics.

<REC>
<ACCN>gakkai-0000011144</ACCN>
<TITL TYPE="kanji">電子原稿・電子出版・電子図書館-「SGML 実験誌」の作成実験を通して</TITL>
<TITE TYPE="alpha">Electronic manuscripts, electronic publishing and electronic library </TITE>
<AUPK TYPE="kanji">根岸 正光</AUPK>
<AUPE TYPE="alpha">Negishi,Masamitsu</AUPE>
<CONF TYPE="kanji">研究発表会(情報学基礎)</CONF>
<CNFE TYPE="alpha">The Special Interest Group Notes of IPSJ</CNFE>
<CNFD>1991. 11. 19</CNFD>
<ABST TYPE="kanji"><ABST.P>電子出版というキーワードを中心に、文献の執筆、編集、印刷、流通の過程の電子化について、その現状を整理して今後の動向を検討する。とくに、電子出版に関する国際規格である SGML(Standard Generalized Markup Language)に対するわが国での動きに注目し、学術情報センターにおける「SGML 実験誌」およびその全文 CD-ROM 版の作成実験を通じて得られた知見を報告する。また電子図書館について、その諸形態を展望する。出版文化に依拠するこの種の社会システムの場合、技術的な問題というのは、その技術の社会的な受容・浸透の問題であり、この観点から標準化の重要性を論じる。</ABST.P></ABST>
<ABSE TYPE="alpha"><ABSE.P>Current situation on electronic processing in preparation, editing, printing and distribution of documents is summarized and its future trend is discussed, with focus on the concept: "Electronic publishing. "Movements in the country concerning an international standard on electronic publishing, SGML (Standard Generalized Markup Language), are assumed to be important, and the results from an experiment at NACSIS to publish "SGML Experimental Journal" and to make its full-text CD-ROM version are reported. Various forms of "Electronic library" are also investigated. The author puts emphasis on standardization, as technological problems for those social systems based on cultural settings of publication of the country, are the problems of acceptance and penetration of the technology in the society.</ABSE.P></ABSE>
<KYWD TYPE="kanji">電子出版 // 電子図書館 // 電子原稿 // SGML // 学術情報センター // 全文データベース</KYWD>
<KYWE TYPE="alpha">Electronic publishing // Electronic library // Electronic manuscripts // SGML // NACSIS // Full text databases</KYWE>
<SOCN TYPE="kanji">情報処理学会</SOCN>
<SOCE TYPE="alpha">Information Processing Society of Japan</SOCE>
</REC>

Fig. 3. A Sample of the Document Record

The query format is similar to that used in TREC-1 and 2 and contains SGML-like tags. A query consists of a title of the topic, a description, a detailed narrative and a list of concepts and field(s). The title is a very short description of the topic and can be used as a very short query that resembles those often submitted by end-users of Internet search engines. Each narrative may contain a detailed explanation of the topic, term definitions, background knowledge, the purpose of the search, criteria for judgment of relevance, and so on.

Topics were collected from users. Their subject domains are diversified such as, computer sciences, pharmacology, biochemistry, social sciences such as education, linguistics, and so on.

The Collection contains 30 training topics and 53 test topics. Among them, 21 training topics and 39 test topics are usable for cross-lingual retrieval. All the topics are written in Japanese. English and Korean versions will be available.

<検索課題q=0005>
<タイトル >
特徴次元リダクション
</タイトル >
<検索要求
クラスタリングにおける特徴次元リダクション
</検索要求
<検索要求説明
オブジェクトのクラスタリングを行なうとき オブジェクトを特徴ベクトルで表現することが望まれる。アプリケーションによっては オブジェクトの次元は数千、数万となることがある。このような場合、事前に次元を落とすことが必要になる。正解文書は 特徴次元リダクションの方法について、理論 面から、または実験によって提案、比較などを行なっているもの。画像処理などの実験の操作の一部として特徴次元リダクションを用いているだけでは要求を満たさない
</検索要求説明
<概念>
特徴選択、主成分分析、情報の粒度、幾何クラスタリング
</概念>
<分野>
1.電子・情報・制御
</分野>
</検索課題>

Fig. 4 A sample Topic

The criteria of topic preparation are as follows.

(1) Statements of “user need” rather than “queries”

(2) <Description> containing every concept needed to describe the topic

(3) Not too easy:

(3-1) Simple word matching of query terms cannot retrieve every relevant document. (Since a concept may be represented by various terms.)

(3-2) A document containing query terms can be non-relevant. (Since a word may have different meanings.)

(4) Five or more relevant documents in the top 100 documents retrieved by the NACSIS retrieval system

Sentences were modified when they were too restricted or ambiguous. The function category of each topic was analyzed and assigned based on the function category proposed by BMIR [6]. The category indicates the required level of techniques and knowledge to conduct a search of the topic. We tried to balance the topic length, number of relevant documents, and “difficulty”; however, we found that estimating “difficulty” is difficult.

Relevance Judgments (Right Answers)

The relevance judgments were undertaken by pooling methods. A certain number of top-ranked documents were collected from each submitted run and created a pool of possibly relevant documents. Human analysts assessed the relevance of each

document in the pool against the topic. The relevance assessment was undertaken using three grades: relevant, partially relevant, non-relevant. Two analysts assessed the relevance to a topic separately, and then the primary analyst of the topic, who created the topic, made the final judgment based on the negociation between two analysts.

Relevance judgment files contain not only the relevance of each document in the pool but also contain extracted phrases or passages showing the reason why the analyst assessed the document as "relevant". Since a narrative of topics may contain some description related to the user's situation or the purpose of the search, situational-oriented relevance judgments were conducted as well as topic-oriented relevance judgments, which are more common in ordinary IR systems laboratory testing. However, only topic-oriented judgments are used in the formal evaluation of this Workshop. The relevance judgments are often called "right answers" in the Japanese IR community.

Linguistic Analysis

Part of the J Collection contains detailed hand-tagged part-of-speech (POS) tags [8]. Because of the absence of explicit boundaries between words in Japanese sentences, we set three levels of lexical boundaries (i.e., word boundaries, strong and weak morpheme boundaries), and assigned detailed POS tags based on the boundaries and types of origin, so that the Collection could be used to examine suitable term segmentations of Japanese texts for retrieval purpose. The Tagged Corpus was not used for official evaluation of the IR tasks but was distributed to the IR tasks participants on request.

Robustness of the System Evaluation using the Test Collection 1

The Test Collection 1 itself has been tested from the following aspects so that it is usable as a reliable tool for IR system testing:

(A) exhaustivity of the document pool
(B) inter-analysts consistency and its effect for system evaluation
(C) topic-by-topic evaluation.

The results of these studies have been reported and published on various occasions [9-13]. A brief summary of the studies is shown here.

1. The initial pooling, which was held in NACSIS to create relevance judgments for the training topics, worked well and covered 97% of the total relevant documents.

2. Interactive searches were effective for some particular topics and found 17.5% of unique relevant documents.

3. In terms of exhaustiveness, pooling the top 100 documents from each run worked well for topics with less than 50 relevant documents. For the topics with more than 100 relevant documents, although the top 100 pooling covered only 51.9% of the total relevant documents, the coverage reached 90% if combined with interactive searches.

Regarding test topics, based on the analyses above we decided to use the top 100 pooling and interactive search for the topic with more than 50 relevant documents.

4. The top 100 pooling method has the effect of reducing the size of the document pool to 21.9% of its possible size for training topics and 20.9% for test topics.

5. Very high similarity was found between the system rankings produced using different sets of relevance assessments, regardless of the different coverage and pooling methods and regardless of inconsistencies between relevance assessments of human analysts for both training and test topics.

Regardless of the inconsistency of the relevance assessments, we found strong correlation between the system rankings produced using the relevant judgments by the primary analysts, the secondary analysts, and the final judges [9-10,13]. Regarding pooling methods, we compared the relevance judgments based on the document pool created by collecting top X documents from one run per system and top X documents from every run, regardless of the number of runs submitted by a system.

As a result, regardless of the different coverage of pooled methods, we found strong correlation between the system rankings produced based on the different pooling methods for both training topics and test topics [10,13]. We concluded that the test collection is reliable as a tool for system evaluation based on these analyses.

2.4. Retrieval Results

This section reports an overview of the retrieval results from the aspects of system effectiveness and analyzes some of the similarities and differences of the approaches taken by each participating group. Since one of the main purposes of the NTCIR Workshop is to enhance research in Japanese text retrieval and cross-lingual retrieval, examination of various approaches using the NTCIR Collection has been encouraged. For the official test runs, 53 new topics (topics 0031–0083) were delivered. Forty-eight *ad hoc* runs and 69 cross-lingual runs were submitted. Relevance judgments were undertaken by human analysts and delivered to the active participating groups by the beginning of June 1999. For further details of each approach, please consult

each system paper in the Workshop Proceedings. The Proceeding will be available online; http://www.rd.nacsis.ac.jp/~ntcadm/

Cross-Lingual Runs

The Cross-Lingual Retrieval Task is an *ad hoc* retrieval task in which the documents are in English and the topics are in Japanese. Sixty-nine retrieval result sets were submitted from 10 participating groups and the NTCIR IR organizer. The organizer did this to assess the effectiveness of the system used for the initial pooling for this collection without any extra tuning.

Cross-Lingual All Runs

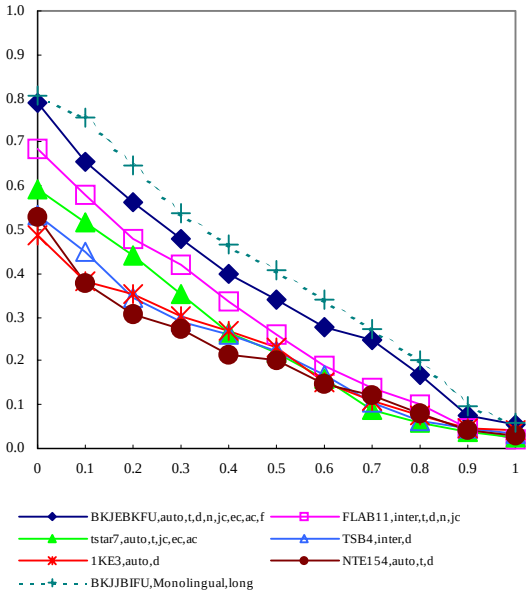

Fig. 5 Cross-Lingual All Runs (Relevant) Top Runs （Best run for each group）

In the Workshop, various approaches towards tasks were reported. Results are as follows;

1. A search using a longer topic tended to obtain better results, however, some runs gave opposite results and could not utilize the <NARRATIVE> field of the search topics to improve search effectiveness.
2. Interactive systems obtained better results than automatic system
3. A search using <CONCEPT> in the topic obtained better results than a search without <CONCEPT>, but the search using <CONCEPT> only worked poor
4. Indexing methods used were various; *bi-gram*, morphological analysis, and extended n-gram such as, adaptive segmentation, etc. were used.
5. Retrieval model: Vector space, probabilistic model, extension of probabilistic model, etc.

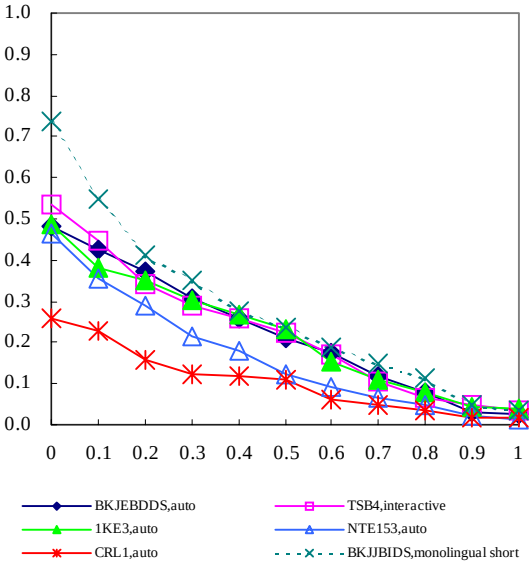

Fig.6 Cross-Lingual ShortQueries w/o Concepts (Relevant) Top Runs (Best run for each group)

(1) Approaches

Every runs took the approach of “Query Translation” as overall approach for CLIR. Most of runs were dictionary-based or using machine translation. One group used corpus-based approach.

For the dictionaries, most groups prepared bilingual dictionaries internally from various resources. The size of them were from 20,636 entries to 579,115 entries. EDR and EDICT are also used by a few groups.

(2) Technical Terms

One of the difficult components of the NTCIR-1 is that of translating technical terms. Terms are sometimes too specific or technical, or too new to look up in ordinary dictionaries.

On the other hand, quasi-paired documents using a native language and English, such as ntc1-je0 documents, are rather common in non-English speaking countries. These documents can be found on the Web and in scholarly documents, commercial documents describing a company’s products, and government documents. Using these documents to prepare bilingual or multilingual lexical resources that are usable for cross-lingual information access will be one of the practical approaches to the problem.

(3) Query Expansion and Disambiguation

Query expansion and word sense disambiguation were conducted by several groups. For example, one group looked at post-translation expansion and disambiguation using various kinds of English

monolingual corpora. On the other hand, the other group looked at pre-translation expansion and disambiguation using J Collection or another parallel corpus. Another group incorporated automatic local feedback. Several groups employed more natural and naïve query expansion, *i.e.*, translating terms into more than one target language, translation and expansion at the same time.

(4) Transliteration

A group used transliteration of Katakana words and worked well. Especially it seemed to work well on technical terms and is expected to effective to reduce the problems caused by the word mismatch because of the various ways of expression of a concept in Japanese document, which were discussed in Section 1.1. More investigation is expected on this matter.

2.5 Next NTCIR Workshop

NACSIS will change its name into National institute of Informatics in the next spring and plans to run a second evaluation after the change. It will include at least Japanese and English with training data available in May 2000, test data available in approximately in September, and the workshop itself scheduled for February or March, 2001. Some part of the schedule may be changed through the effort to cordinating other evaluation project like TREC, TDT, or TIDES.

Meanwhile, we are planing details of the tasks, subtasks, evaluation scheme, collection, and resources. The needs of training courses and tutorials on evaluation of information retrieval systems including interactive systems for Japanese new comers in Japanese language and ones on Japanese text processing and available resources for non Japanese researchers in English are suggested from the advisory group. Any comments, advises, and leads are welcome.

3. Large-Scale Multilingual Lexicon

One of the problems of the CLIR is drawbacks associated with the availability of resources [16-17]. For dictionary-based methods, the coverage of dictionaries or thesauri are often not sufficiently broad and deep, thus domain specific terms or new concepts, which are critical for retrieval of technical and scientific documents used here, tend not to be listed. Lack of resources is also problematic in corpus-based methods; parallel corpora are not always readily available. Regarding machine translation techniques, we can not ignore the cost of linguistic analysis. Preparing large-scale multilingual lexicon through international collaboration and encourage the research on the methods to generate such lexicon automatically are one of the promising directions to facilitate CLIR in various types of information.

In the NTCIR Workshop, many groups constructed bilingual lexical resource internally. Some of them were from the document collection in NTCIR-1. We are very grateful that a group kindly offered to share the resources with participants of the next workshop. We ourselves also conducted research on CLIR using automatically generated bilingual keyword clusters based on graph-theory [18]. The clusters were generated using Japanese and English keywords, which are assigned to scientific papers by the authors.

In the contrast to the problems of resource shown above, our approach for the CLIR using keyword clusters has several advantages: regarding the resources, the keyword data we used has advantages, *i.e.*, *subject-spe*cific bilingual keyword corpora are *readily available* in machine-readable form for a great many subject domains. They are also readily segmented into terms, and well aliened (albeit with some noise). Therefore, they are rather easy to handle for our purpose. In addition to this, the graph-theoretic approach as advantages: (1) by utilizing topological features of the graph, low-frequency keywords can be treated properly and are usable in IR; (2) the clusters contain not only J-E pairs, but also J-J and E-E pairs so that they are usable for query expansion in monolingual retrieval as well, and (3) this was achieved with reasonable computational cost.

We used keyword clusters based on only less than 10% of database as a corpus and achieved the effectiveness of 52.4% of monolingual at average precision for all relevant documents at “relevant” level and 65.7% for "relevant & partial relevant” level with automatic phrase identification and synonym processing, without any manual interaction, and, in monolingual IR, 13.4% over the baseline at the average precision for relevant level, and 14.2% for "relevant & partial relevant”. This is comparable to existing researches and fairy well as a start point. However this still leaves a lot of room for improvement both in the aspect of IR and clustering as further discussion is found in [18].

The majority of scientific papers published in Japan have such bilingual keyword lists and such lists are widely found in non-English speaking countries. These are one of the good candidates to enhance the lexical resources usable for CLIR through international collaboration.

4. Towards Cross-Lingual Information Access.

The occasion of acting as one of the panel speakers at ACM-DL'99 and SIGIR'99 Joint Workshop on Multilingual Information Discovery and Access (MIDAS) brought me to think about "information discovery and access". From my point of view, one of the focuses of research and development in digital libraries is to enhance the universal information discovery and access. It includes (1) to know the appropriate place to search necessary information, (2) to search and acquire it, and (3) to access it.

To know the right place to search, or "where to go", is the first step of the information discovery. Any of Internet search engines can not be perfect in the aspect of coverage of the documents indexed. Traditional libraries or information centers have provided such service called "referral service". This kind of function should be incorporated to digital libraries. To search and acquire is covered by information retrieval.

To access the information has two meanings, i.e., to access the information physically and intellectually. The latter includes that users do information work such as decision making, problem solving, paper writing using the information retrieved. To support and facilitate the users to do information work using retrieved information is one of the new areas of investigation in digital libraries. For example, translation of the retrieved documents, skimming, automatic summarization, extracting answer passages, clustering the retrieved docs, visualization, text mining, comparison of the contents in the documents etc

Cross-lingual information discovery and access is one of the key aspects to achieve the universal information discovery and access. CLIR is a part of the whole process of the cross-lingual information discovery and access. Expansion of the scope of research from CLIR to cross-lingual information discovery and access is one of the interesting directions of future research.

ACKNOWLEDGMENTS

We thank all the participants for their contributions and the assessors. Special thanks are due to Donna Harman, Ellen Voorhees, Ross Wilkinson, Sung H. Myaeng, and Mun Kew Leong for their substantial advice and continuous support. Hard work of the NTCIR project members should be acknowledged; Kazuko Kuriyama, Toshihiko Nozue, Koji Eguchi, Souichiro Hidaka, Hideyuki Kato, and Yuko Miyamoto.

REFERENCES

[1] NTCIR Project. http://www.rd.nacsis.ac.jp/~ntcadm/

[2] Kando, N. et al. NACSIS Test Collection Workshop (NTCIR-1) [poster]. Proceeding of ACM-SIGIR '99. Aug.,1999, Berkeley, CA

[3] Kando, N. Cross-Linguistic Scholarly Information Transfer and Database Services in Japan. Annual Meeting of the ASIS. (Nov. 1997) Washington DC.

[4] TREC URL: http://www.nist.gov/trec

[5] TDT URL: http://www.nist.gov/speech and selecting "benchmark tests" and then "TDT-3"

[6] Sakai, T. et al. BMIR-J2: Test Collection for Evaluation of Japanese Information Retrieval Systems. SIGIR Forum (to appear).

[7] The list of 65 academic societies is available at URL, http://www.rd.nacsis.ac.jp/~ntcadm/acknowledge/thanks1-en.html.

[8] Kageura, K. Overview of Term Recognition Task, In Proceedings of the First NTCIR Workshop, August 30-September 1, 1999, Tokyo (to appear).

[9] Kando, N. et al. NTCIR-1: Its Policy and Practice. IPSJ SIG Notes, Vol.99, No.20, pp. 33-40, 1999 [in Japanese].

[10] Kuriyama, K. et al. Pooling for a Large Scale Test Collection: Analysis of the Search Results for the Pre-test of the NTCIR-1 Workshop. IPSJ SIG Notes, Vol. 99, (May, 1999) [in Japanese].

[11] Kuriyama, K. et al. Construction of a Large Scale Test Collection: Analysis of the Training Topics of the NTCIR-1. IPSJ SIG Notes, Vol. 99 (July, 1999) [in Japanese].

[12] Kando, N. et al. Construction of a Large Scale Test Collection: Analysis of the Test Topics of the NTCIR-1. In Proceedings of IPSJ Annual Meeting (to appear) [in Japanese].

[13] Kuriyama, K. et al. Pooling for a Large Scale Test Collection: Analysis of the Search Results for the first NTCIR Workshop. In Proceedings of IPSJ Annual Meeting (to appear) [in Japanese].

[14] Rorvig, M. et al. The N-Gram Hypothesis Applied to Matched Sets of Visualized Japanese-English Technical Documents. In Proceeding of the ASIS Annual Meeting (to appear).

[15] Voorhees, E.M. Variations in Relevance Judgments and the Measurement of Retrieval Effectiveness. In Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Melbourne, Australia, August, 1998, pp. 315-323.

[16] Ballesteros, L. et al. Resolving ambiguity for cross-language retrieval. Proceedings of the 21st Annual International ACM SIGIR Conference of Research and Development in Information Retrieval, Aug., 1998, Melbourne, Australia, p.64-71

[17] Shauble, P. et al."Cross-language Information Retrieval (CLIR) Track overview" In Proceedings of TREC6, p.25-30, 1998

[18] Kando, N. et al., Cross-Lingual Information Retrieval using Automatically Generated Multilingual Keyword Clusters. In Proceedings of the 3rd International Workshop of IR with Asian Language, Oct., 1998, Singapore