Overview of CLIR Task at the Sixth NTCIR Workshop

Kazuaki Kishida*, Kuang-hua Chen**, Sukhoon Lee***,
Kazuko Kuriyama****, Noriko Kando[+], Hsin-Hsi Chen[++]

*School of Library and Information Science, Keio University
Tokyo, 108-8345, Japan
kishida@slis.keio.ac.jp

**Department of Library and Information Science, National Taiwan University
Taipei 10617, Taiwan
khchen@ntu.edu.tw

***Department of Statistics, Chungnam National University
Taejon 305-764, Korea
shlee@stat.cnu.ac.kr

****Shirayuri College
Chofu 182-8525, Japan
kuriyama@shirayuri.ac.jp

[+]National Institute of Informatics (NII)
Tokyo 101-8430, Japan
kando@nii.ac.jp

[++]Department of Computer Science and Information Engineering, National Taiwan University,
Taipei 10617, Taiwan
hh_chen@csie.ntu.edu.tw

Abstract

The purpose of this paper is to overview research efforts at the NTCIR-6 CLIR task, which is a project of large-scale retrieval experiments on cross-lingual information retrieval (CLIR) of Chinese, Japanese, Korean, and English. The project has three sub-tasks, multi-lingual IR (MLIR), bilingual IR (BLIR), and single language IR (SLIR), in which many research groups from ten countries or regions are participating. This paper describes the system of the NTCIR-6 CLIR task and its test collection (document sets, topic sets, and method for relevance judgments), and reviews CLIR techniques used by participants and search performance of runs submitted for evaluation.
Keywords: *Cross-lingual information retrieval; Evaluation; Retrieval experiment*

1 Introduction

The purpose of the NTCIR-6 CLIR Task is to contribute to developments of cross-linguistic information retrieval (CLIR) that enables us to search documents in East Asian languages, i.e., Chinese (C), Japanese (J), and Korean (K). In this time, unlike the previous workshops, two separate stages have been arranged in order to understand much more characteristics of CLIR, i.e.,

- STAGE1: ordinary ad hoc search tasks on multilingual IR (MLIR), bilingual IR (BLIR), and single language IR (SLIR),
- STAGE2: cross-collection analysis using old test collections from NTCIR-3 to -5.

In STAGE1, CJK document sets consisting of news paper articles published in 2000 to 2001 were used as in the last workshop. Meanwhile, for STAGE2, the participants in this workshop were asked to search for old topic sets using their current systems. Observations from the STAGE2 are expected to provide us with useful insight on reliability of evaluation based on experiments using the Cranfield-type test collection, and we would obtain more substantial knowledge on performance of techniques for CLIR including monolingual IR.

This paper aims at reporting on the CLIR task in the NTCIR-6 workshop. In the section 2, the design of the task is explained. The section 3 discusses the document collection and search topics. The outline of

submission of results is described in the section 4, and the section 5 dedicates to an explanation of relevance judgments. The section 6 reviews retrieval techniques used by participating research groups. Finally, search performance of each subtask is discussed in the section 7 and 8.

2 Design of the CLIR Task

2.1 Schedule

Registration for the NTCIR-6 CLIR task started on April in 2006. The time schedule for the NTCIR-6 CLIR task is as follows.

2006-06-01: Document sets (CJK) Release
2006-06-23: Registration Due
2006-07-01: Topic Release for STAGE1
2006-08-01: Submission of Search Results for STAGE1
2006-08-01: Topic Release for STAGE2
2006-09-20: Submission of Search Results for STAGE2
2006-11-29: Delivery of Evaluation Results for STAGE1
2007-03-01: Paper Due (for Proceedings)
2007-05-15 to 18: NTCIR Workshop 6 Meeting

2.2 Subtasks in STAGE1 and STAGE2

In this workshop, English document sets are removed from our test collection for both STAGE1 and 2. Also, in STAGE2, only BLIR and SLIR are allowed, i.e., MLIR is not involved.

2.2.1 Multilingual IR (MLIR). In general, the document set of MLIR task consists of two or more languages. For the NTCIR-6 CLIR task, the multilingual search is limited to use the CJK collection, which consists of Chinese (C), Japanese (J), and Korean (K) documents. Regarding the topic set, participants can select one language from CJKE for each run. Therefore, there are four combinations of topic sets and the document set, i.e.,

Topic set: C or J or K or E >> Doc set: CJK

2.2.2 Bilingual IR (BLIR). BLIR means that the document set in a single language is searched for a topic in a different language, e.g., searching Japanese documents for Korean topics (K-J run). The combinations of topics and documents for the BLIR subtask are as follows:

Topic set: C >> Doc set: J or K
Topic set: J >> Doc set: C or K
Topic set: K >> Doc set: C or J
Topic set: E >> Doc set: C or J or K

2.2.3 Single language IR (SLIR). The topic set and document sets of SLIR are written in a same language. The combinations of topics and documents for the SLIR subtask are as follows:

Topic set: C >> Doc set: C
Topic set: J >> Doc set: J
Topic set: K >> Doc set: K

2.3 Outline of STAGE2

The objective of STAGE2, which is a new challenge in this workshop, is to take the more reliable measurement of search performance by repeatedly examining each technique or method on three old test collections, NTCIR-3, -4, and -5. To do this, the task organizers asked participants to submit search results for these test collections, respectively, using their current IR systems without any special adjustment of parameters. As a result, the same IR system is to be evaluated three times independently. We suppose that a participant submits search results of System A and B, and their scores of mean average precision are as follows:

	NTCIR-3	NTCIR-4	NTCIR-5
System A	0.5	0.4	0.6
System B	0.4	0.3	0.4

This result shows more clearly dominance of System A than that obtained from an experiment using just a single test collection.

The document sets and the numbers of topics included in the old test collections will be shown in the section 3.1 and 3.2. Also, we can know about them much more by referring to past task overviews in the proceedings of NTCIR-3, -4, and -5 Workshops.

2.4 Topic fields and run types

2.4.1 Types of runs. Basically, each topic consists of four fields, i.e., “T” (TITLE), “D” (DESC), “N” (NARR) and “C” (CONC) (see below for details). We can categorize search runs based on the fields used for the execution. In the NTCIR-6 CLIR task, the following types of runs are adopted:

- *Mandatory runs*: T-run and D-run
  Each participant must submit two types of run for each combination of topic language and document language(s);
  T-run, for which only TITLE field is used,
  D-run, for which only DESC field is used.
  The purpose of asking participants to submit these mandatory runs is to make research findings clear by comparing systems or methods under a unified condition.
- *Recommended runs*: DN-run
  Participants are also recommended to execute DN run that employs both <DESC> and <NARR> fields.

- *Optional runs*

  Other any combinations of fields are allowed to submit as optional runs according to each participant's research interests, e.g., TDN-run, DC-run, TDNC-run and so on.

2.4.2 Number of runs. Each participant can submit up to 5 runs in total for each language pair regardless of the type of run, and participants are allowed to include two T-runs in maximum and also two D-runs in maximum into the 5 runs. The language pair means the combination of topic language and document language(s). For example,

Language combination -> Topic: C and Docs: CJK (C-CJK)

Submission -> two T-runs, a D-run, a DN-run and a TDNC run (5 runs in total).

Table 1 Document sets for STAGE1 of the NTCIR-6 CLIR Task

Sources	No. of Docs
Chinese 2000-01	
CIRB040r (United Daily News (udn), United Express (ude), Ming Hseng News (mhn), Economic Daily News (edn))	901,446
Total	**901,446**
Japanese 2000-01	
Mainichi	199,681
Yomiuri	658,719
Total	**858,400**
Korean 2000-01	
Hankookilbo	85,250
Chosunilbo	135,124
Total	**220,374**

2.4.3 Identification and priority of runs. Each run has to be associated with a RunID, which is an identity for each run. The rule of format for RunID is as follows.

Group's ID - Topic Language - Document Language - Run Type - pp

The 'pp' is a string of two digits used to represent the priority of the run. It is used as a parameter for pooling. The participants have to decide the priority for each submitted run among them on each language pair. The "01" means the highest priority. For example, we suppose that a participating group, LIPS, submits 3 runs for C-CJK, where the first is a T run, the second is a D run and the third is a DN run. Therefore, the Run ID for each run is LIPS-C-CJK-T-01, LIPS-C-CJK-D-02, and LIPS-C-CJK-DN-03, respectively.

Furthermore, in STAGE2, participants were asked to add a notation for indicating which test collection was used for executing the search run, e.g., LIPS-C-C-D-01-N3 in which 'N3' means the NTCIR-3 ('N4' and 'N5' are used for NTCIR-4 and NTCIR-5, respectively).

Table 2 Document sets used for STAGE2 of the NTCIR-6 CLIR Task

(a) For NTCIR-5 Topic sets

See Table 1.

(b) For NTCIR-4 Topic sets

Sources	No. of Docs
Chinese 1998-99	
CIRB020 (United Daily News)	249,508
CIRB011 (China Times, China Times Express, Commercial Times, China Daily News, Central and Daily News)	132,173
Total	**381,681**
Japanese 1998-99	
Mainichi	220,078
Yomiuri	375,980
Total	**596,058**
Korean 1998-99	
Hankookilbo	149,921
Chosunilbo	104,517
Total	**254,438**

(c) For NTCIR-3 Topic sets

Sources	No. of Docs
Chinese 1998-99	
CIRB020 (United Daily News)	249,508
CIRB011 (China Times, China Times Express, Commercial Times, China Daily News, Central and Daily News)	132,173
Total	**381,681**
Japanese 1998-99	
Mainichi	220,078
Total	**220,078**
Korean 1994	
Korea Economic Daily	66,146
Total	**66,146**

3 Test Collection

3.1 Document Sets

Table 1 shows the sources and the numbers of records in the document collections for STAGE1, which are the same for the NTCIR-5 CLIR task except English document sets. As mentioned above, STAGE2 employs also old document sets shown in Table 2. Therefore, the NTCIR-6 CLIR task again uses sets of news articles from various news agencies as docu-

ment collections for evaluation like previous workshops.

The tags used for separating each field in a record of these sets are indicated in Table 3.

Table 3 Tags used for identifying each filed

Mandatory tags	
<DOC>	The tag for each document
<DOCNO>	Document identifier
<LANG>	Language code: CH, JA, KR
<HEADLINE>	Title of this news article
<DATE>	Issue date
<TEXT>	Text of news article
Optional tags	
<P>	Paragraph marker
<SECTION>	Section identifier in original newspapers
<AE>	Contain figures or not
<WORDS>	Number of words in 2 bytes (for Mainichi Newspaper)

Table 4 Numbers of topics used for STAGE1 and STAGE2 in the NTCIR-6 CLIR Task

	Target document collections			
	C	J	K	CJK
STAGE1	50	50	50	50
STAGE2				
NTCIR-5	50	47[1]	50	-
NTCIR-4	59[2]	55[3]	57[4]	-
NTCIR-3	42[5]	42[6]	30	-

Note:
1) 021, 023 and 039 are excluded.
2) 025 are excluded.
3) 001, 002, 022, 025 and 038 are excluded.
4) 001, 010 and 011 are excluded.
5) 016, 026, 028, 029, 030, 031, 041 and 044 are excluded.
6) 001, 003, 006, 009, 011, 013, 048 and 049 are excluded.

3.2 Topic

3.2.1 Numbers of topics. The numbers of topics used for STAGE1 and STAGE2 are shown in Table 4, which also includes lists of topics excluded from evaluation due to few relevant documents. That is, the NTCIR CLIR task has been adopting so-called the "3-in-S+A" criterion (see section 5.2), and according to the criterion, some topics may have been excluded for final evaluations from the original sets of topics delivered initially to participants.

It should be noted that old topics used for past workshops are also reused in STAGE1, i.e., for the first time, the participants were provided a set of 140 topics, which consists of 60 topics of NTCIR-4, 50 topics for Chinese and Japanese documents in NTCIR-3 and 30 topics for Korean documents in NTCIR-3. These topics can be reused for 2000-01 document sets because the document sets in NTCIR-3 and -4 are those published in 1998-99, which are different from sets used for STAGE1 (see Table 1 and 2).

After the participants submitted their search results for the 140 topics, the organizers selected final 50 topics for evaluation, based on relevance assessments of "shallow" document pools, which consist of only top-ten documents of search runs from every participant, i.e., topics having enough relevant documents in the shallow pools were chosen for evaluation. The IDs of 50 topics selected finally for STAGE1 are listed in Appendix 1.

3.2.2 Format of topics. Each topic has four fields; 'T' (TITLE), 'D' (DESC), 'N' (NARR), 'C' (CONC). The tags used in topics are shown in Table 5. The following shows a sample topic.

<TOPIC>
<NUM>009</NUM>
<SLANG>CH</SLANG>
<TLANG>EN</TLANG>
<TITLE>Japan, South Korea, Fishery Agreement</TITLE>
<DESC>Find articles on the content of the final fishery agreement between Japan and South Korea</DESC>
<NARR>
<BACK>There are frequent disputes between Japan and South Korea because of the 35 years of colonized reign. Things worsened in January of 1998 when Japan announced the abolishment of the fishery agreement of 1965. Finally, in September of 1998, a new fishery agreement between Japan and South Korea was reached despite disputes over the sovereignty of the isles. It marked an end to eight months of serious disputes between the two countries. Please query the content of this new agreement for things such as allocation of fishing areas and results of negotiation.</BACK>
<REL>Documents of reports on the final fishery agreement are relevant. Reports on historical disputes and events between Japan and South Korea are not relevant.</REL>
</NARR>
<CONC>Japan, South Korea, Fishery Agreement, Isles, Fishing Area</CONC>
</TOPIC>

It should be noted that topics reused from the sets of NTCIR-3 do not include <BACK>, <REL> and <TERM> subfields in the <NARR> filed.

Table 5 Topic tags used in the NTCIR-6 CLIR task

<TOPIC>	The tag for each topic
<NUM>	Topic identifier
<SLANG>	Source language code: CH, EN, JA, KR
<TLANG>	Target language code: CH, EN, JA, KR
<TITLE>	A concise representation of information request, which is composed of noun or noun phrase.
<DESC>	A short description of the topic. A brief description of information need, which is composed of one sentence.
<NARR>[1]	A much longer description of topic. The <NARR> may have three parts; (1)<BACK>...</BACK>: background information about the topic. (2)<REL>...</REL>: further interpretation of the request and proper nouns, a list of relevant or irrelevant items, the specific requirements or limitations of relevant documents, and so on. (3)<TERM>...</TERM>: a definition or explanation of proper nouns, scientific terms and so on.
<CONC>	keywords relevant to the whole topic.

Note: 1) topics reused from the sets of NTCIR-3 do not include <BACK>, <REL> and <TERM> subfields.

Table 6 Regional Distribution of Participants

Country / Region	STAGE		Total
	1	2	
Australia	1	1	1
Canada	1	2	2
China	1	1	2
Japan	6	6	6
Korea	1	1	1
Singapore	1		1
Switzerland	1	1	1
Taiwan	5	5	5
UK	1	1	1
USA	2	1	2
Total	20	19	22

4 Submission of Results

In total, search results were submitted by 22 groups from 10 countries or regions (see Table 6). Regarding the numbers of participants, Japan is dominant (6 groups), followed by Taiwan (5 groups). Appendix 2 shows the names of groups submitting the search results. Unfortunately, other 7 groups that applied to participate in the NTCIR-6 CLIR task could not submit final results for some reasons.

Table 7 shows the numbers of submitted runs and groups. In STAGE1, 151 runs were submitted, of which 94 (62.3%) are for SLIR, 55 (36.4%) BLIR, and 2 (1.3%) MLIR. Also, the numbers of submitted runs and groups for old NTCIR-5 test collection in STAGE2 are indicated in Table 7 (b). In total, 175 runs were submitted, of which 87 (49.7%) are for SLIR, 88 (50.3%) BLIR. The same numbers of runs and groups were submitted for NTCIR-4, but for NTCIR-3, a C-K run and an E-K run are missing.

5 Results of Relevance Judgments

Like the past workshop, evaluation in STAGE1 of the NTCIR-6 CLIR task is based on the TREC-like procedures using results of relevance assessment of each pool of retrieved documents for topics (Appendix 3 shows the size of each pool for identifying relevant documents). The trec_eval program was used to score search results submitted by participants.

Traditionally, the NTCIR CLIR task has been adopting multi-grade relevance judgments, i.e., the organizers have been assigning four degrees to each document in the process of relevance judgments; “S: highly relevant”, “A: relevant”, “B: partially relevant”, and “C: irrelevant”. Like the previous workshops, two kinds of relevance degree,

- *Rigid relevant*: S+A,
- *Relaxed relevant*: S+A+B,

are used for computing traditional metrics, average precision and R-precision, since trec_eval scoring program adopts only binary relevance. Hence, two files of relevance judgments (rigid and relaxed) for each collection (C, J, K and CJK) are prepared by the task organizers.

In addition, an original module for computing evaluation metrics based on multi-grade relevance judgments (e.g., nDCG and so on) was introduced in the NTCIR-6 (see section 7.4).

Appendix 4 indicates the numbers of relevant documents included in the document sets at STAGE1.

6 Overview of CLIR Techniques

In CLIR research field, various techniques have been proposed for enhancing the search performance. This section dedicates to review techniques or methods used in the NTCIR-6 CLIR task.

6.1 Indexing methods

6.1.1 Indexing of CJK text. For Chinese text, unigrams or overlapped bi-grams are often used by par-

ticipants. UniNE group [2] employs a Mandarin tool for segmenting Chinese text. Meanwhile, ISCAS group [19] seems to use an HMM-based technique for indexing Chinese text.

Table 7 Statistics on submissions for the NTCIR-6 CLIR task

(a) STAGE1

	Run types	# of runs	# of groups
SLIR	C-C	30	9
	J-J	44	10
	K-K	20	4
	Total	94	16
BLIR	C-J	17	5
	C-K	0	0
	J-C	4	1
	J-K	5	1
	K-C	2	1
	K-J	5	1
	E-C	8	4
	E-J	9	2
	E-K	5	1
	Total	55	10
MLIR	C-CJK	2	1
	J-CJK	0	0
	K-CJK	0	0
	E-CJK	0	0
	Total	2	1
Total		151	20

(b) STAGE2 (Only submissions for NTCIR-5 test collection[1])

	Run types	# of runs	# of groups
SLIR	C-C	36	10
	J-J	35	9
	K-K	16	4
	Total	87	15
BLIR	C-J	19	6
	C-K	6	2
	J-C	2	1
	J-K	8	2
	K-C	4	2
	K-J	7	2
	E-C	28	7
	E-J	11	4
	E-K	5	2
	Total	88	13
Total		175	19

Note: 1) The numbers of runs and groups for NTCIR-4 test collections are the same as those for NTCIR-5 test. For NTCIR-3, the numbers are almost the same, but a C-K run and an E-K run are missing.

For Japanese text, in general, morphological analyzers such as ChaSen, Mecab and JUMAN are employed. JSCCL group [13] explores phrasal indexing for Japanese text, and investigates some weighting systems for the phrases.

For Korean text, KLE group [11] applies an unsupervised segmentation method, which is based on collection statistics. In the KLE experiments, a data fusion technique is used for merging multiple results from different indexing methods. UniNE group [2] uses a Hangle analyzer.

6.2 Translation

6.2.1 Translation methods. In general, query translation by MT or bilingual dictionaries is adopted. RALI group [16] applies language modeling based on translation probabilities computed from a parallel corpus and a bilingual dictionary by the IBM Model 1. OASIS group [21] tries to merge results from some online translation systems.

6.2.2 Out-of-vocabulary (OOV) problem. NICT group [4] tries to solve the OOV problem using Web resources based on the joint validation model [23]. ISQUT group [9] uses a term extraction strategy for Web-based OOV solution, in which a kind of term disambiguation technique is employed for low-frequency terms. WTG group [6] examines a term extraction method based on a pattern matching technique for identifying OOV terms in Web resources.

Wikipedia is used for solving the OOV problem by CYTU group [10] and IASL group [12].

6.2.3 Conversion of *Kanji* codes. For C-J runs, BRKLY group [20] applies "no translation" technique, in which only conversion of character coding from Chinese to Japanese is executed.

6.2.4 Pivot language approach. BRKLY group [20] uses English as a pivot for J-C runs. Also, OASIS group [21] executed C-J runs as a pivot search via English.

6.3 Retrieval model

Various IR models or algorithms are used; vector space model, Okapi formula, language modeling, logistic regression model, PIRCS and so on. UniNE group [2] investigates extensively search performance of various combinations of retrieval models, data fusion methods and indexing techniques.

YLMS group [3] tries to estimate optimal parameters of the BM25 model by using a genetic algorithm (GA). NICT group [4] applies a variation of Okapi formula, in which term location information and other information are incorporated. OKSAT group [15] combines the Okapi weighting with a term variety factor (TVF) for computing document scores.

6.4 Query expansion and re-ranking

6.4.1 Pseudo-relevance feedback (PRF). HUM group [7] attempts to analyze effectiveness of PRF by using new metrics; "Success@10" and "Generalized Success@10". KLE group [11] applies a model-based feedback by Zhai and Lafferty [24] for long queries.

TSB group [1] introduces a complicated PRF technique for searching news paper article, in which the system merges three search results from "selective sampling query", "headline query" and "lead sentence query". Also, in TSB experiment, a "graded relevance feedback" based on multi-grade relevance assessment is examined.

6.4.2 Document re-ranking. I2R group [8] explores a document re-ranking method based on propagation-based semi-supervised learning algorithm. Also, NTNU group [14] compares empirically performance between label propagation (LP), k-nearest neighboring (kNN) and relevance feedback for document re-ranking.

Meanwhile, CCNU group [17] applies the group-average agglomerative clustering for document expansion in their re-ranking algorithm.

6.5 Others

HUM group [7] examines precision score in the first 9000 retrieved items.

7 Search Results and Performance at STAGE1

In this section, we review briefly search performance at STAGE 1. Note that search runs submitted after the deadline are marked with an asterisk "*".

7.1 SLIR runs

7.1.1 C-C runs. In total, 30 Chinese monolingual runs (C-C runs) were submitted by 9 groups (see Table 7). Table 8 shows average, median, maximum and minimum values of mean average precision (MAP) by types of run. We use the following notations;

C-C: all C-C monolingual runs
C-C-T: all C-C <TITLE>-only runs (T-runs)
C-C-D: all C-C <DESC>-only runs (D-runs)
C-C-O: all runs other than T- and D-runs.

Note that these notations will be used for other languages (J, K, and E).

Table 9 indicates best runs of top eight groups ranked according to MAP scores of D-runs based on rigid relevance.

Table 8 MAP of overall C-C runs

(a) Average and median

	Average		Median	
	Rigid	Relax	Rigid	Relax
C-C	0.2285	0.3187	0.2314	0.3348
C-C-T	0.2321	0.3214	0.2430	0.3399
C-C-D	0.2379	0.3339	0.2420	0.3447
C-C-O	0.1968	0.2751	0.2266	0.3296

(b)Min and max

	Min		Max	
	Rigid	Relax	Rigid	Relax
C-C	0.0183	0.0165	0.3136	0.4118
C-C-T	0.1146	0.1468	0.3097	0.4090
C-C-D	0.1561	0.2389	0.3136	0.4118
C-C-O	0.0183	0.0165	0.2794	0.3832

7.1.2 J-J runs. In total, 44 J-J monolingual runs were submitted by 10 groups (see Table 7). Table 10 shows average, median, maximum and minimum values of MAP by types of runs. Table 11 indicates top eight groups ranked according to MAP scores of D-runs based on rigid relevance.

Table 9 Top-ranked 8 groups (C-C, Rigid, D-runs)

Run-ID	MAP
I2R-C-C-D-01	0.313
UniNE-C-C-D-05	0.289
pircs-C-C-D-02*	0.259
CCNU-C-C-D-02	0.256
HUM-C-C-D-05	0.227
BRKLY-C-C-D-03*	0.199
WTG-C-C-D-04	0.186
HUM-C-C-D-03	0.186

Table 10 MAP of overall J-J runs

(a) Average and median

	Average		Median	
	Rigid	Relax	Rigid	Relax
J-J	0.2553	0.3267	0.2654	0.3420
J-J-T	0.2707	0.3427	0.2738	0.3485
J-J-D	0.2480	0.3214	0.2519	0.3206
J-J-O	0.2378	0.3023	0.2767	0.3420

(b)Min and max

	Min		Max	
	Rigid	Relax	Rigid	Relax
J-J	0.0172	0.0158	0.3600	0.4393
J-J-T	0.1560	0.1955	0.3600	0.4393
J-J-D	0.1768	0.2249	0.3255	0.4138
J-J-O	0.0172	0.0158	0.2969	0.3898

7.1.3 K-K runs. In total, 20 K-K monolingual runs were submitted by 4 groups (see Table 7). Table 12 shows average, median, maximum and minimum values of MAP by types of run. Table 13 indicates top eight groups ranked according to MAP scores of D-runs based on rigid relevance.

Table 11 Top-ranked 8 groups (J-J, Rigid, D-runs)

Run-ID	MAP
TSB-J-J-D-02	0.325
UniNE-J-J-D-03	0.289
YLMS-J-J-D-03	0.274
NICT-J-J-D-02	0.268
BRKLY-J-J-D-03*	0.252
JSCCL-J-J-D-04	0.245
KLE-J-J-D-02	0.243
HUM-J-J-D-05	0.226

Table 12 MAP of overall K-K runs

(a) Average and median

	Average		Median	
	Rigid	Relax	Rigid	Relax
K-K	0.3850	0.4626	0.4113	0.4876
K-K-T	0.3834	0.4644	0.3920	0.4775
K-K-D	0.3892	0.4678	0.4139	0.4870
K-K-O	0.3821	0.4544	0.4531	0.5368

(b)Min and max

	Min		Max	
	Rigid	Relax	Rigid	Relax
K-K	0.0442	0.0222	0.4789	0.5883
K-K-T	0.3252	0.3772	0.4236	0.5179
K-K-D	0.2898	0.3423	0.4535	0.5375
K-K-O	0.0442	0.0222	0.4789	0.5883

7.2 BLIR

7.2.1 BLIR runs on Chinese document sets. In total, 4 J-C runs were submitted by 1 group, 2 K-C runs by 1 group, and 8 E-C runs by 4 groups (see Table 7). Table 14, 15 and 16 show MAP scores of these runs.

7.2.2 BLIR runs on Japanese document sets. In total, 17 C-J runs were submitted by 5 groups, 5 K-J runs by 1 groups, and 9 E-J runs by 2 groups (see Table 7). Table 17, 18 and 19 show MAP scores of these runs.

Table 13 Top-ranked 4 groups (K-K, Rigid, D-runs)

Run-ID	MAP
UniNE-K-K-D-02	0.454
KLE-K-K-D-02	0.429
NICT-K-K-D-03	0.414
HUM-K-K-D-05	0.332

Table 14 Best run of each group (J-C, Rigid, D-runs)

Run-ID	MAP
BRKLY-J-C-D-03*	0.078

Table 15 Best run of each group (K-C, Rigid, D-runs)

Run-ID	MAP
IASL-K-C-D-01	0.102

Table 16 Best run of each group (E-C, Rigid, D-runs)

Run-ID	MAP
I2R-E-C-D-01	0.191
pircs-E-C-D-04*	0.167
WTG-E-C-D-02	0.120
ISQUT-E-C-D-03	0.090

7.2.3 BLIR runs on Korean document sets. In total, 5 J-K runs by 1 group, and 5 E-K runs by 1 group (see Table 7). Unfortunately, there is no C-K run. Table 20 and 21 show MAP scores of these runs.

Table 17 Best run of each group (C-J, Rigid, D-runs)

Run-ID	MAP
TSB-C-J-D-02	0.312
BRKLY-C-J-D-03*	0.252
NICT-C-J-D-05	0.207
AINLP-C-J-D-02	0.062
OASIS-C-J-D-04*	0.054

Table 18 Best run of each group (K-J, Rigid, D-runs)

Run-ID	MAP
NICT-K-J-D-02	0.267

Table 19 Best runs of each group (E-J, Rigid, D-runs)

Run-ID	MAP
TSB-E-J-D-02	0.307
NICT-E-J-D-05	0.251

Table 20 Best run of each group (J-K, Rigid, D-runs)

Run-ID	MAP
NICT-J-K-D-05	0.287

Table 21 Best runs of each group (E-K, Rigid, D-runs)

Run-ID	MAP
NICT-E-K-D-05	0.292

7.2.5 Summary on BLIR. Table 22 shows best runs of BLIR with best SLIR (monolingual) runs. C-J, K-J and E-J searches show very high performance in comparison with SLIR runs (95.8%, 82.1% and 94.4%, respectively). Also, E-C, J-K, and E-K keep

moderately high performance (over 60%). In contrast, there is a room for further research efforts on other combinations of languages.

Table 22 Summary on BLIR: Best runs of each language combination (Rigid, D-runs)

	Documents		
	C	J	K
Mono. (base)	0.313 (100%)	0.325 (100%)	0.454 (100%)
C > X	-	0.312 (95.8%)	N/A
J > X	0.078 (24.7%)	-	0.287 (63.2%)
K > X	0.102 (32.6%)	0.267 (82.1%)	-
E > X	0.191 (61.0%)	0.307 (94.4%)	0.292 (64.3%)

7.3 MLIR

In the case of MLIR, only 1 group submitted search results (C-CJK) whose performance is shown in Table 23.

Table 23 Best runs of each group by run type (MLIR, Rigid, D-runs)

Run-ID	MAP
CYUT-C-CJK-D-02	0.0584

Table 24 Numerical values for each relevance degree

	Highly relevant	Relevant	Partially relevant	Irrelevant
Value	3	2	1	0

7.4 Evaluation Metrics Based on Multi-grade Relevance

Appendix 5 shows scores of some metrics based on multi-grade relevance for top runs in Table 9, 11, 13, and 14 to 21. As described above, our relevance assessors have determined four degree relevance (highly relevant, relevant, partially relevant and irrelevant) for each document. In this paper, numerical values are assigned to each degree as shown in Table 24.

The following metrics are used in the Appendix 5.

- nDCG: normalized DCG (Kekalainen and Jarvelin[25])
- QM: Q-Measure (Sakai et al.[26])
- gAP: generalized average precision (Kishida[27])

Mathematical definitions of these metrics are described in Kishida [27]. In addition, scores of standard average precision (or MAP) based on rigid relevance and relaxed relevance are also shown in this appendix, which are labeled as AP_g, and AP_x, respetively (it should be noted that the runs are sorted according to scores of AP_g like Table 9, 11, 13, and 14 to 21).

This calculation should be considered as a trial. We have developed a program module for computing these scores. Note that traditional MAP scores from our module may be slightly different from those by trec_eval due to a different way for data processing.

8 Search Results and Performance at STAGE2

Appendix 6 shows a result of STAGE2 at NTCIR-6 CLIR task. In this table, MAP scores (based on rigid relevance) of each runs are listed by types of run.

In order to check "stability" of the MAP score over different test collections, we compute correlation coefficients by types of runs (see Table 25). Except J-J runs and E-C runs, correlation is very high, which means that the MAP score is stable.

The detailed analysis of the stability is our future work.

Table 25 Correlation matrix of MAP scores (rigid) of D-runs

(a) C-C runs (n=9)

	NTCIR-5	NTCIR-4	NTCIR-3
NTCIR-5	1.000		
NTCIR-4	0.956	1.000	
NTCIR-3	0.952	0.957	1.000

(b) J-J runs (n=14)

	NTCIR-5	NTCIR-4	NTCIR-3
NTCIR-5	1.000		
NTCIR-4	0.861	1.000	
NTCIR-3	0.760	0.751	1.000

(c) K-K runs (n=6)

	NTCIR-5	NTCIR-4	NTCIR-3
NTCIR-5	1.000		
NTCIR-4	0.992	1.000	
NTCIR-3	0.962	0.976	1.000

(d) E-C runs (n=6)

	NTCIR-5	NTCIR-4	NTCIR-3
NTCIR-5	1.000		
NTCIR-4	0.562	1.000	
NTCIR-3	0.645	0.946	1.000

(e) C-J runs (n=7)

	NTCIR-5	NTCIR-4	NTCIR-3
NTCIR-5	1.000		
NTCIR-4	0.992	1.000	
NTCIR-3	0.975	0.973	1.000

(f) E-J runs (n=4)

	NTCIR-5	NTCIR-4	NTCIR-3
NTCIR-5	1.000		
NTCIR-4	0.988	1.000	
NTCIR-3	0.946	0.983	1.000

Acknowledgment

Part of the research project carried out in Taiwan is partially supported by Excellent Research Projects of National Taiwan University under grant no. 95R0062-AE00-02.

References

[1] Tetsuya Sakai, Makoto Koyama, Tatsuya Izuha, Akira Kumano, Toshihiko Manabe and Tomoharu Kokubu. Toshiba BRIDJE at NTCIR-6 CLIR: The head/lead method and graded relevance feedback. In *Proceedings of NTCIR-6 Workshop*, 2007.

[2] Abdou Samir and Jacques Savoy. Monolingual experiments with far-east languages in NTCIR-6. In *Proceedings of NTCIR-6 Workshop*, 2007..

[3] Sumio Fujita. NTCIR-6 CLIR-J-J experiments at Yahoo! Japan. In *Proceedings of NTCIR-6 Workshop*, 2007.

[4] Masaki Murata, Jong-Hoon Oh, Qing Ma and Hitoshi Isahara. Applying multiple characteristics and techniques in the NICT Information Retrieval System in NTCIR-6. In *Proceedings of NTCIR-6 Workshop*, 2007.

[5] Yu-Chieh Wu, Kun-Chang Tsai and Jie-Chi Yang. NCU in bilingual information retrieval experiments at NTCIR-6. In *Proceedings of NTCIR-6 Workshop*, 2007.

[6] Dong Zhou, Mark Truran, Tim Brailsford and Helen Ashman. NTCIR-6 experiments using pattern matched translation extraction. In *Proceedings of NTCIR-6 Workshop*, 2007.

[7] Stephen Tomlinson. Sampling precision to depth 9000: evaluation experiments at NTCIR-6. In *Proceedings of NTCIR-6 Workshop*, 2007.

[8] Yang Lingpeng, Ji Donghong and Nie Yu. I2R at NTCIR6. In *Proceedings of NTCIR-6 Workshop*, 2007.

[9] Chengye Lu, Yue Xu and Shlomo Geva. Improving translation accuracy in web-based translation extraction. In *Proceedings of NTCIR-6 Workshop*, 2007.

[10] Chen-Yu Su, Shih-Hung Wu and Tien-Chien Lin. Using Wikipedia to translate OOV term on MLIR. In *Proceedings of NTCIR-6 Workshop*, 2007.

[11] Seung-Hoon Na and Jong-Hyeok Lee. POSTECH at NTCIR-6: Combining evidences of multiple term extractions for mono-lingual and cross-lingual retrieval in Korean and Japanese. In *Proceedings of NTCIR-6 Workshop*, 2007.

[12] Yu-Chun Wang, Cheng-Wei Lee, Richar Tzong-Han Tsai and Wen-Lian Hsu. IASL System for NTCIR-6 Korean-Chinese cross-language information retrieval. In *Proceedings of NTCIR-6 Workshop*, 2007.

[13] Tetsuya Tashiro. JustSystems in Japanese monolingual information retrieval at NTCIR-6. In *Proceedings of NTCIR-6 Workshop*, 2007.

[14] Yuen-Hsien Tseng, Chen-Yang Tsai and Ze-Jing Chuang. On the robustness of document re-ranking techniques: a comparison of label propagation, KNN, and relevance feedback. In *Proceedings of NTCIR-6 Workshop*, 2007.

[15] Takashi Sato. NTCIR-6 CLIR experiments at Osaka Kyoiku University: term expansion using online dictionaries and Weighting Score by Term Variety. In *Proceedings of NTCIR-6 Workshop*, 2007.

[16] Lixin Shi and Jian-Yun Nie. Using unigram and bigram language models for monolingual and cross-language IR. In *Proceedings of NTCIR-6 Workshop*, 2007.

[17] Tingting He, Li Li, Guozhong Qu and Yong Zhang. Chinese information retrieval based on document expansion. In *Proceedings of NTCIR-6 Workshop*, 2007.

[18] Chen-Hsin Cheng, Reuy-Jye Shue, Hung-Lin Lee, Shu-Yu Hsieh, Guann-Cyun Yeh and Guo-Wei Bian. AINLP at NTCIR-6: evaluations for multilingual and cross-lingual information retrieval. In *Proceedings of NTCIR-6 Workshop*, 2007.

[19] Ruihong Huang, Le Sun, Jing Li, Longxi Pan and Junlin Zhang. ISCAS in CLIR at NTCIR-6: experiments with MT and PRF. In *Proceedings of NTCIR-6 Workshop*, 2007.

[20] Fredric Gey. Search between Chinese and Japanese text collections. In *Proceedings of NTCIR-6 Workshop*, 2007.

[21] Vitaly Klyuev. OASIS at NTCIR-6: on-line query translation for Chinese-Japanese cross-lingual information retrieval. In *Proceedings of NTCIR-6 Workshop*, 2007.

[22] Kui-Lam Kwok and Norbert Dinstl. NTCIR-6 Monolingual Chinese and English-Chinese cross language retrieval experiments using PIRCS. In *Proceedings of NTCIR-6 Workshop*, 2007.

[23] J. –H. Oh and H. Isahara. Mining the web for transliteration lexicons: joint-validation approach.

In *2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI' 06)*, pages 254-261, 2006.

[24] C. Zhai and J. Lafferty. Model-based feedback in the language modeling approach to information retrieval. In *CIKM '01*, pages 403-410, 2001.

[25] Kalervo Jarvelin and Jaana Kekalainen. IR evaluation methods for retrieving highly relevant documents. In *Proceedings of the 23rd Annual Conference on Research and Development in Information Retrieval*, pages 41-48, 2000.

[26] T. Sakai, M. Koyama, and A. Kumano. Toshiba BRIDJE at NTCIR-4 CLIR: monolingual / bilingual IR and flexible feedback. In *Proceedings of NTCIR-4 Workshop*, 2004.

[27] Kazuaki Kishida. Property of average precision and its generalization: an examination of evaluation indicator for information retrieval. NII Technical Reports, NII-2005-014E, 2005. http://research.nii.ac.jp/TechReports/05-014E.html

Appendix 1. Topic IDs in final test collection of STAGE1

Topic ID in NTCIR-6	Original ID in source topic sets			Topic ID in NTCIR-6	Original ID in source topic sets		
	NTCIR-4	NTCIR-3 for CJ	NTCIR-3 for K[1]		NTCIR-4	NTCIR-3 for CJ	NTCIR-3 for K[1]
003	003			048	048		
014	014			050	050		
015	015			053	053		
016	016			058	058		
017	017			059	059		
018	018			060	060		
019	019			064		004	
020	020			065		005	
021	021			070		010	
023	023			074		014	
024	024			075		015	
026	026			077		017	
027	027			079		019	
030	030			080		020	
033	033			083		023	
036	036			095		035	
037	037			096		036	
039	039			097		037	
041	041			099		039	
042	042			100		040	
043	043			102		042	
044	044			103		043	
045	045			105		045	
046	046			106		046	
047	047			110		050	

Note: 1) As a result, it was not needed to use NTCIR-3 topics for Korean document sets.

Appendix 2. List of participating groups

	ID	Name of Group	Country	STAGE	
				1	2
1	AINLP	Artificial Intelligence Laboratory; Huafan University	Taiwan	*	*
2	BRKLY	UC DATA; University of California	USA	*	
3	CCNU	Department of Computer Science; Huazhong Normal University	China	*	
4	CYUT	Department of Computer Science and Information Engineering; Chaoyang University of Technology	Taiwan	*	*
5	HUM	Open Text Corporation (Hummingbird)	Canada	*	*
6	I2R	Institute for Infocomm Research	Singapore	*	
7	IASL	Intelligent Agent Systems Lab, Institute of Information Science, Academia Sinica	Taiwan	*	*
8	ISCAS	Institute of Software; Chinese Academy of Sciences	China		*
9	ISQUT	School of SEDC, Queensland University of Technology	Australia	*	*
10	JSCCL	Justsystem Corporation	Japan	*	*
11	KLE	Dept. of Computer Science & Engineering; POSTECH	Korea	*	*
12	NCUTW	Department of Computer Science; National Central University	Taiwan	*	*
13	NICT	Computational Linguistics Group; National Institute of Information and Communications Technology	Japan	*	*
14	NTNU	Digital Media Center; National Taiwan Normal University	Taiwan	*	*
15	OASIS	Software Engineering Laboratory, University of Aizu	Japan	*	*
16	OKSAT	Information Processing Center; Osaka Kyoiku University	Japan	*	*
17	pircs	Queens College, City University of New York	USA	*	*
18	RALI	RALI; Dept. IRO; University of Montreal	Canada		*
19	TSB	Knowledge Media Laboratory; Toshiba Corporate R&D Center	Japan	*	*
20	UniNE	Computer Science Department; University of Neuchatel	Switzerland	*	*
21	WTG	Web Technology Group, University of Nottingham	UK	*	*
22	YLMS	Yahoo Japan Corporation; Japanese Language Processing Department	Japan	*	*

Appendix 3. Pool size and the numbers of documents judged by each language (STAGE1)

topic	C		J		K		CJKE	
	size	#doc	size	#doc	size	#doc	size	#doc
003	100	508	100	547	100	508	100	1431
014	100	863	100	510	100	863	100	1763
015	100	717	100	749	100	717	100	1955
016	100	893	100	696	100	893	100	2122
017	100	861	100	1147	100	861	100	2462
018	100	2960	100	2576	100	2960	100	6937
019	100	1939	100	2905	100	1939	100	6402
020	100	1147	100	2020	100	1147	100	4299
021	100	1006	100	945	100	1006	100	2402
023	100	1259	100	1186	100	1259	100	2885
024	100	1250	100	1277	100	1250	100	3295
026	100	1187	100	970	100	1187	100	2860
027	100	1266	100	821	100	1266	100	2823
030	100	856	100	753	100	856	100	2104
033	100	824	100	1402	100	824	100	2680
036	100	715	100	801	100	715	100	1830
037	100	1298	100	1951	100	1298	100	4559
039	100	1703	100	1834	100	1703	100	4168
041	100	1580	100	854	100	1580	100	3362
042	100	1646	100	1180	100	1646	100	3289
043	100	2380	100	1155	100	2380	100	4395
044	100	448	100	740	100	448	100	1646
045	100	1752	100	900	100	1752	100	3324
046	100	1026	100	686	100	1026	100	2224
047	100	795	100	1110	100	795	100	2306
048	100	1219	100	634	100	1219	100	2166
050	100	1034	100	777	100	1034	100	2437
053	100	967	100	655	100	967	100	1993
058	100	933	100	1240	100	933	100	3006
059	100	1103	100	1141	100	1103	100	2808
060	100	1019	100	1685	100	1019	100	3350
064	100	1290	100	1322	100	1290	100	3457
065	100	1205	100	1712	100	1205	100	4072
070	100	1010	100	1484	100	1010	100	3436
074	100	800	100	1394	100	800	100	2679
075	100	754	100	683	100	754	100	1916
077	100	816	100	1289	100	816	100	2951
079	100	986	100	874	100	986	100	2680
080	100	513	100	602	100	513	100	1766
083	100	690	100	906	100	690	100	2482
095	100	1330	100	1627	100	1330	100	3498
096	100	2154	100	1624	100	2154	100	4465
097	100	1187	100	1020	100	1187	100	2577
099	100	1184	100	1781	100	1184	100	3985
100	100	1477	100	1959	100	1477	100	4438

102	100	1081	100	1564	100	1081	100	3369
103	100	1659	100	1645	100	1659	100	4031
105	100	1467	100	2050	100	1467	100	4388
106	100	2500	100	1435	100	2500	100	4503
110	100	1816	100	1843	100	1816	100	4750

Appendix 4. Numbers of relevant documents

	C		J		K		CJK	
topic	S+A	S+A+B	S+A	S+A+B	S+A	S+A+B	S+A	S+A+B
003	59	77	24	60	59	77	89	150
014	9	17	6	20	9	17	33	62
015	11	18	16	30	11	18	43	90
016	18	62	8	47	18	62	42	218
017	42	77	6	46	42	77	66	163
018	267	361	43	100	267	361	355	617
019	21	32	105	173	21	32	155	274
020	15	25	27	51	15	25	68	125
021	32	54	16	26	32	54	59	104
023	50	74	60	64	50	74	130	201
024	23	29	29	36	23	29	120	153
026	62	85	31	36	62	85	110	169
027	64	97	22	27	64	97	103	167
030	62	96	92	239	62	96	185	450
033	27	42	69	99	27	42	132	215
036	59	71	127	127	59	71	203	236
037	15	31	142	146	15	31	251	302
039	19	35	58	59	19	35	96	163
041	142	253	210	216	142	253	380	658
042	138	179	23	24	138	179	171	221
043	138	194	38	48	138	194	341	453
044	25	34	24	33	25	34	66	114
045	196	322	177	233	196	322	418	624
046	37	51	68	94	37	51	119	186
047	77	163	75	154	77	163	199	454
048	20	44	71	122	20	44	114	215
050	8	16	119	174	8	16	140	268
053	91	164	43	104	91	164	240	406
058	40	59	100	129	40	59	176	244
059	71	130	168	311	71	130	262	546
060	49	78	154	266	49	78	320	508
064	226	400	29	105	226	400	385	666
065	53	80	27	29	53	80	109	143
070	10	15	4	4	10	15	21	30
074	108	185	171	198	108	185	465	580
075	18	24	40	57	18	24	81	130
077	25	32	11	14	25	32	43	61
079	83	94	80	112	83	94	177	238
080	18	83	83	122	18	83	107	288
083	61	73	38	45	61	73	144	399

095	16	32	56	63	16	32	128	225
096	329	498	158	180	329	498	507	760
097	23	38	66	71	23	38	128	164
099	35	51	101	128	35	51	250	364
100	41	62	13	19	41	62	74	106
102	25	77	8	29	25	77	41	154
103	50	87	41	106	50	87	275	460
105	33	50	58	104	33	50	211	345
106	292	507	39	65	292	507	413	689
110	47	83	6	19	47	83	95	165

Appendix 5. Scores of evaluation metrics based on multi-grade relevance

(1) C-C (D-runs)

Run-IDs	AP_g	AP_x	nDCG	gAP	QM
I2R-C-C-D-01	0.313	0.407	0.596	0.332	0.417
UniNE-C-C-D-05	0.289	0.412	0.586	0.322	0.422
pircs-C-C-D-02*	0.259	0.366	0.556	0.288	0.383
CCNU-C-C-D-02	0.256	0.351	0.543	0.279	0.376
HUM-C-C-D-05	0.227	0.292	0.496	0.238	0.314
BRKLY-C-C-D-03*	0.199	0.302	0.467	0.229	0.319
WTG-C-C-D-04	0.186	0.259	0.452	0.204	0.275
HUM-C-C-D-03	0.186	0.249	0.459	0.199	0.275
NCUTW-C-C-D-04	0.137	0.203	0.314	0.157	0.215

(2) J-J (D-runs)

Run-IDs	AP_g	AP_x	nDCG	gAP	QM
TSB-J-J-D-02	0.325	0.414	0.626	0.368	0.440
UniNE-J-J-D-03	0.289	0.369	0.579	0.329	0.403
YLMS-J-J-D-03	0.274	0.362	0.557	0.318	0.389
NICT-J-J-D-02	0.268	0.353	0.535	0.310	0.380
BRKLY-J-J-D-03*	0.252	0.316	0.514	0.282	0.347
JSCCL-J-J-D-04	0.245	0.321	0.522	0.285	0.346
KLE-J-J-D-02	0.243	0.315	0.501	0.278	0.341
HUM-J-J-D-05	0.226	0.284	0.487	0.255	0.315

(3) K-K (D-runs)

Run-IDs	AP_g	AP_x	nDCG	gAP	QM
UniNE-K-K-D-02	0.454	0.538	0.705	0.467	0.548
KLE-K-K-D-02	0.429	0.523	0.683	0.456	0.533
NICT-K-K-D-03	0.414	0.487	0.664	0.431	0.510
HUM-K-K-D-05	0.332	0.413	0.612	0.357	0.428

(4)J-C (D-runs)

Run-ID	AP_g	AP_x	nDCG	gAP	QM
BRKLY-J-C-D-03*	0.078	0.107	0.204	0.084	0.120

(5) K-C (D-runs)

Run-ID	AP_g	AP_x	nDCG	gAP	QM
IASL-K-C-D-01	0.102	0.127	0.283	0.104	0.148

(6) E-C (D-runs)

Run-IDs	AP_g	AP_x	nDCG	gAP	QM
I2R-E-C-D-01	0.191	0.281	0.480	0.216	0.302
pircs-E-C-D-04*	0.167	0.237	0.394	0.188	0.257
WTG-E-C-D-02	0.120	0.156	0.324	0.125	0.175
ISQUT-E-C-D-03	0.090	0.114	0.253	0.092	0.119

(7) C-J (D-runs)

Run-IDs	AP_g	AP_x	nDCG	gAP	QM
TSB-C-J-D-02	0.312	0.371	0.577	0.336	0.403
BRKLY-C-J-D-03*	0.252	0.316	0.514	0.282	0.347
NICT-C-J-D-05	0.207	0.255	0.405	0.227	0.277
AINLP-C-J-D-02	0.062	0.085	0.201	0.072	0.096
OASIS-C-J-D-04*	0.054	0.060	0.199	0.055	0.076

(8) K-J (D-runs)

Run-ID	AP_g	AP_x	nDCG	gAP	QM
NICT-K-J-D-02	0.267	0.346	0.519	0.304	0.368

(9) E-J (D-runs)

Run-ID	AP_g	AP_x	nDCG	gAP	QM
TSB-E-J-D-02	0.307	0.369	0.578	0.332	0.402
NICT-E-J-D-05	0.251	0.321	0.505	0.283	0.349

(10) J-K (D-runs)

runID	AP_g	AP_x	nDCG	gAP	QM
NICT-J-K-D-05	0.287	0.356	0.522	0.310	0.377

(11) E-K (D-runs)

runID	AP_g	AP_x	nDCG	gAP	QM
NICT-E-K-D-05	0.292	0.374	0.532	0.319	0.389

Appendix 6. Cross-collection analysis: MAP of D-runs (rigid relevance)

(1) For Chinese collection

(a) C-C runs

NTCIR-5		NTCIR-4		NTCIR-3	
run-IDs	AP_g	run-IDs	AP_g	run-IDs	AP_g
UniNE-C-C-D-03-N5	0.424	UniNE-C-C-D-03-N4	0.254	UniNE-C-C-D-03-N3	0.325
UniNE-C-C-D-05-N5	0.422	UniNE-C-C-D-05-N4	0.251	UniNE-C-C-D-05-N3	0.324
pircs-C-C-D-02-N5	0.374	pircs-C-C-D-02-N4	0.218	pircs-C-C-D-02-N3	0.294
HUM-C-C-D-05-N5	0.327	HUM-C-C-D-05-N4	0.172	AINLP-C-C-D-02-N3	0.220
AINLP-C-C-D-02-N5	0.297	RALI-C-C-D-01-N4	0.171	HUM-C-C-D-05-N3	0.199
RALI-C-C-D-01-N5	0.281	HUM-C-C-D-02-N4	0.154	RALI-C-C-D-01-N3	0.198
NCUTW-C-C-D-02-N5	0.275	AINLP-C-C-D-02-N4	0.142	NCUTW-C-C-D-02-N3	0.191
HUM-C-C-D-02-N5	0.270	NCUTW-C-C-D-02-N4	0.120	HUM-C-C-D-02-N3	0.190
WTG-C-C-D-04-N5	0.237	WTG-C-C-D-04-N4	0.082	WTG-C-C-D-04-N3	0.094
Ave.	0.323	Ave.	0.174	Ave.	0.226

(b) J-C runs

NTCIR-5		NTCIR-4		NTCIR-3	
run-IDs	AP_g	run-IDs	AP_g	run-IDs	AP_g
NCUTW-J-C-D-02-N5	0.182	NCUTW-J-C-D-02-N4	0.082	NCUTW-J-C-D-02-N3	0.085

(c) K-C runs

NTCIR-5		NTCIR-4		NTCIR-3	
run-IDs	AP_g	run-IDs	AP_g	run-IDs	AP_g
IASL-K-C-D-01-N5	0.108	IASL-K-C-D-01-N4	0.105	IASL-K-C-D-01-N3	0.096
NCUTW-K-C-D-02-N5	0.087	NCUTW-K-C-D-02-N4	0.061	NCUTW-K-C-D-02-N3	0.026
Ave.	0.098	Ave.	0.083	Ave.	0.061

(d) E-C runs

NTCIR-5		NTCIR-4		NTCIR-3	
run-IDs	AP_g	run-IDs	AP_g	run-IDs	AP_g
pircs-E-C-D-04-N5	0.242	pircs-E-C-D-04-N4	0.142	pircs-E-C-D-04-N3	0.175
NCUTW-E-C-D-02-N5	0.179	ISQUT-E-C-D-01-N4	0.114	RALI-E-C-D-01-N3	0.120
AINLP-E-C-D-02-N5	0.155	RALI-E-C-D-01-N4	0.102	ISQUT-E-C-D-01-N3	0.116
ISQUT-E-C-D-01-N5	0.151	NCUTW-E-C-D-02-N4	0.063	AINLP-E-C-D-02-N3	0.094
WTG-E-C-D-02-N5	0.141	AINLP-E-C-D-02-N4	0.062	NCUTW-E-C-D-02-N3	0.072
RALI-E-C-D-01-N5	0.128	WTG-E-C-D-02-N4	0.054	WTG-E-C-D-02-N3	0.051
Ave.	0.166	Ave.	0.089	Ave.	0.105

(2) For Japanese collection

(a) J-J runs

NTCIR-5		NTCIR-4		NTCIR-3	
run-IDs	AP_g	run-IDs	AP_g	run-IDs	AP_g
UniNE-J-J-D-05-N5	0.402	TSB-J-J-D-02-N4	0.408	TSB-J-J-D-02-N3	0.407
UniNE-J-J-D-03-N5	0.400	TSB-J-J-D-01-N4	0.405	TSB-J-J-D-01-N3	0.394
YLMS-J-J-D-02-N5	0.398	UniNE-J-J-D-03-N4	0.389	UniNE-J-J-D-03-N3	0.392
TSB-J-J-D-02-N5	0.389	UniNE-J-J-D-05-N4	0.375	JSCCL-J-J-D-02-N3	0.382
YLMS-J-J-D-03-N5	0.386	YLMS-J-J-D-02-N4	0.375	YLMS-J-J-D-02-N3	0.373
TSB-J-J-D-01-N5	0.380	JSCCL-J-J-D-02-N4	0.373	UniNE-J-J-D-05-N3	0.360
JSCCL-J-J-D-02-N5	0.346	YLMS-J-J-D-03-N4	0.365	JSCCL-J-J-D-04-N3	0.347
JSCCL-J-J-D-04-N5	0.332	NICT-J-J-D-02-N4	0.360	NICT-J-J-D-02-N3	0.332
KLE-J-J-D-02-N5	0.328	JSCCL-J-J-D-04-N4	0.354	YLMS-J-J-D-03-N3	0.328
NICT-J-J-D-02-N5	0.316	HUM-J-J-D-05-N4	0.321	HUM-J-J-D-05-N3	0.320
HUM-J-J-D-05-N5	0.290	HUM-J-J-D-02-N4	0.300	OKSAT-J-J-D-04-N3	0.319
HUM-J-J-D-02-N5	0.282	KLE-J-J-D-02-N4	0.299	HUM-J-J-D-02-N3	0.307
OKSAT-J-J-D-04-N5	0.257	OKSAT-J-J-D-04-N4	0.207	OKSAT-J-J-D-01-N3	0.282
OKSAT-J-J-D-01-N5	0.226	OKSAT-J-J-D-01-N4	0.071	KLE-J-J-D-02-N3	0.280
Ave.	0.338	Ave.	0.329	Ave.	0.344

(b) C-J runs

NTCIR-5		NTCIR-4		NTCIR-3	
run-IDs	AP_g	run-IDs	AP_g	run-IDs	AP_g
TSB-C-J-D-02-N5	0.332	TSB-C-J-D-02-N4	0.330	TSB-C-J-D-02-N3	0.362
TSB-C-J-D-01-N5	0.315	TSB-C-J-D-01-N4	0.323	TSB-C-J-D-01-N3	0.358
NICT-C-J-D-05-N5	0.225	NICT-C-J-D-02-N4	0.241	NICT-C-J-D-02-N3	0.302
NICT-C-J-D-02-N5	0.225	NICT-C-J-D-05-N4	0.235	NICT-C-J-D-05-N3	0.290
AINLP-C-J-D-02-N5	0.088	AINLP-C-J-D-02-N4	0.107	AINLP-C-J-D-02-N3	0.131
OASIS-C-J-D-01-N5	0.048	CYUT-C-J-D-02-N4	0.016	OASIS-C-J-D-01-N3	0.116
CYUT-C-J-D-02-N5	0.036	OASIS-C-J-D-01-N4	0.014	CYUT-C-J-D-02-N3	0.023
Ave.	0.181	Ave.	0.181	Ave.	0.226

(c) K-J runs

NTCIR-5		NTCIR-4		NTCIR-3	
run-IDs	AP_g	run-IDs	AP_g	run-IDs	AP_g
NICT-K-J-D-05-N5	0.249	NICT-K-J-D-05-N4	0.258	NICT-K-J-D-05-N3	0.265
NICT-K-J-D-02-N5	0.247	NICT-K-J-D-02-N4	0.244	NICT-K-J-D-02-N3	0.253
Ave.	0.248	Ave.	0.251	Ave.	0.259

(d) E-J runs

NTCIR-5		NTCIR-4		NTCIR-3	
run-IDs	AP_g	run-IDs	AP_g	run-IDs	AP_g
TSB-E-J-D-02-N5	0.321	TSB-E-J-T-03-N4	0.360	TSB-E-J-D-02-N3	0.370
TSB-E-J-D-01-N5	0.310	TSB-E-J-T-04-N4	0.359	TSB-E-J-D-01-N3	0.366
NICT-E-J-D-02-N5	0.266	NICT-E-J-T-01-N4	0.268	NICT-E-J-D-02-N3	0.267
AINLP-E-J-D-02-N5	0.081	AINLP-E-J-T-01-N4	0.099	AINLP-E-J-D-02-N3	0.175
Ave.	0.244	Ave.	0.271	Ave.	0.295

(3) For Korean collection

(a) K-K runs

NTCIR-5		NTCIR-4		NTCIR-3	
run-IDs	AP_g	run-IDs	AP_g	run-IDs	AP_g
KLE-K-K-D-02-N5	0.525	KLE-K-K-D-02-N4	0.519	UniNE-K-K-D-04-N3	0.330
UniNE-K-K-D-03-N5	0.525	UniNE-K-K-D-03-N4	0.516	UniNE-K-K-D-03-N3	0.324
UniNE-K-K-D-04-N5	0.522	UniNE-K-K-D-04-N4	0.509	KLE-K-K-D-02-N3	0.315
NICT-K-K-D-03-N5	0.494	NICT-K-K-D-03-N4	0.467	NICT-K-K-D-03-N3	0.300
HUM-K-K-D-05-N5	0.416	HUM-K-K-D-05-N4	0.390	HUM-K-K-D-05-N3	0.250
HUM-K-K-D-02-N5	0.354	HUM-K-K-D-02-N4	0.347	HUM-K-K-D-02-N3	0.244
Ave.	0.473	Ave.	0.458	Ave.	0.294

(b) C-K runs

NTCIR-5		NTCIR-4		NTCIR-3	
run-IDs	AP_g	run-IDs	AP_g	run-IDs	AP_g
AINLP-C-K-D-02-N5	0.073	CYUT-C-K-D-04-N4	0.038	CYUT-C-K-D-04-N3	0.009
CYUT-C-K-D-04-N5	0.016	AINLP-C-K-D-02-N4	0.030		
Ave.	0.045	Ave.	0.034		

(c) J-K runs

NTCIR-5		NTCIR-4		NTCIR-3	
run-IDs	AP_g	run-IDs	AP_g	run-IDs	AP_g
NICT-J-K-D-05-N5	0.494	NICT-J-K-D-05-N4	0.467	NICT-J-K-D-05-N3	0.300
NICT-J-K-D-02-N5	0.490	NICT-J-K-D-02-N4	0.456	NICT-J-K-D-02-N3	0.274
KLE-J-K-D-02-N5	0.463	KLE-J-K-D-02-N4	0.403	KLE-J-K-D-02-N3	0.273
Ave.	0.482	Ave.	0.442	Ave.	0.283

(d) E-K runs

NTCIR-5		NTCIR-4		NTCIR-3	
run-IDs	AP_g	run-IDs	AP_g	run-IDs	AP_g
NICT-E-K-D-03-N5	0.494	NICT-E-K-D-03-N4	0.467	NICT-E-K-D-03-N3	0.300
AINLP-E-K-D-02-N5	0.278	AINLP-E-K-D-02-N4	0.058		
Ave.	0.386	Ave.	0.263		