



## NTCIRテストコレクション

Collection	task	Documents			topic./Q		Relevance/ Answer
		Genre	Size	Language	Language	Language	
NTCIR-1	IR	Academic	577MB	JE	J	J	3
CIRB010	IR	News	132MB	Ct	CtE	CtE	4
NTCIR-2	IR	Academic	800MB	JE	JE	JE	4
NTCIR-2 Summ	Summ	News	180 docs	J	J	J	
NTCIR-3 CLIR	IR	News	884MB	CtKJE	CtKJE	CtKJE	4
NTCIR-3 PATENT	IR	Patent	18GB(+5GB)	J(JE)	CsCtKJE	CsCtKJE	3
NTCIR-3 QA	QA	News	282MB	J	J(E)	J(E)	exact
NTCIR-3 Summ	Summ	News	60 docs + 50 sets	J	-	-	
NTCIR-3 WEB	IR	WEB	100GB	Multiple	J(E)	J(E)	4+relative
NTCIR-3 CLIR	IR	News	ca 3GB	CtKJE	CtKJE	CtKJE	4
NTCIR-3 PATENT	IR	Patent	45GB	J(JE)	CsCtKJE	CsCtKJE	3
NTCIR-3 QA	QA	News	776MB	J	J(E)	J(E)	4
NTCIR-3 Summ	Summ	News	30 sets	J	-	-	
NTCIR-3 WEB	IR	WEB	100GB	Multiple	J(E)	J(E)	

Ct: Traditional Chinese, Cs: Simplified Chinese, K: Korean, J: Japanese, E: English  
ntcir4-ov 2004-06-02

7

## Tasks of NTCIR-4

- CLIR: Chinese, Korean, Japanese, English**
- Single Language IR; Bilingual CLIR; Multilingual CLIR; Pivot CLIR
- Patent Retrieval**
- Main : Invalidity Search (Search patent by patent)
  - Feasibility: Automatic Patent Map Generation
- Question Answering**
- 5 possible answers; 1 set of all the answers; series of Qs
- Text Summarization**
- Multiple document summarization --Automatic evaluation!
- Web**
- Informal; Navigational; Geographic; Clustering

ntcir4-ov 2004-06-02

8

## NTCIR Workshop 4 (2003-2004) Organizers



General chair: Jun Adachi, NII  
Program chair: Noriko Kando, NII

ntcir4-ov 2004-06-02

**+CLIR**  
Kuang-hua Chen, NTU  
Sukhoon Lee, NCU  
Kazuaki Kishida, Surugadai U  
Hsin-Hsi Chen, NTU  
Sung Hyon Myaeng, I IU  
Kazuko Kuriyama, Shirayuri U  
Noriko Kando, NII

**+PATENT**  
Atsushi Fujii, Tsukuba U  
Makoto Iwayama, Hitachi  
Noriko Kando, NII

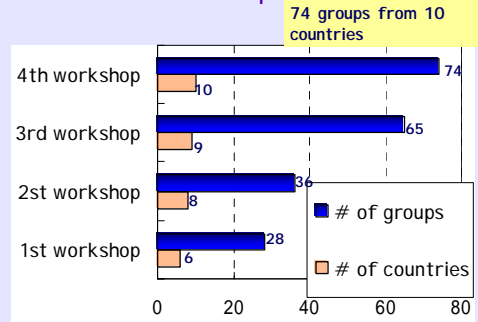
**+Question Answering**  
Junichi Fukumoto, Ritsumeikan U  
Tsuneaki Kato, U Tokyo  
Fumito Masui, Mie U

**+Text Summarization**  
Tsutomu Hirao, NTT-CS  
Takahiro Fukusima, Otemongakuin U  
Hidetsugu Nanba, Hiroshima C U  
Manabu Okumura, TITEC

**+WEB**  
Koji Eguchi, NII  
Keizo Oyama, NII

9

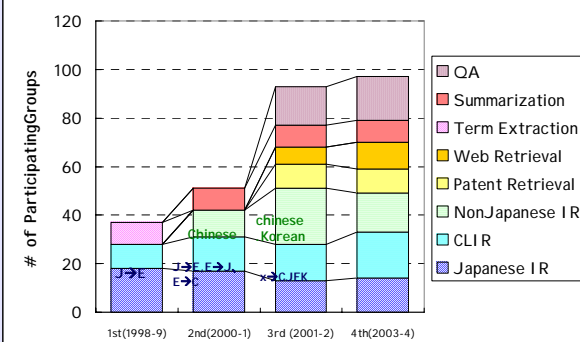
## NTCIR workshop: Number of Participating Groups



ntcir4-ov 2004-06-02

10

## Number of Participants by Tasks



ntcir4-ov 2004-06-02

11



12

## Schedule for NTCIR-4

April 2003: Document Release  
June-Sept, 2003: Dry Run  
Oct-Dec, 2003: Formal Run  
Feb 20, 2004: Evaluation Results Return  
Late March 2004: Paper Submission  
Open Submission Session  
ACM-TALIP Special Issue Recommendation  
2-5 June 2004: Conference, at NII, Tokyo Japan

15 July 2004: ACM-TALIP Due  
31 July 2004: Revised papers for Proceedings

ntcir4-ov 2004-06-02

14

## What's New to NTCIR

- Open Submission Session
- ACM-TALIP Special Issue Recommendation
- Open Attendance
- Research Purpose Use of the Submission Raw Data  
Started with NTCIR-3 CLIR, and then will enlarge
- Online Working Notes and Slides (from today)  
... pls send a PDF of Slides if you agree

ntcir4-ov 2004-06-02

15

## Meeting Local Arrangement

*Local Arrangement: Keiko Watanabe	*Publication/Web: Haruko I shikawa
*Digital Poster: Tsuneaki Kato	*Oral Presentation: Takashi Koga
*Boaster: Fumito Masui	*Secretariat: Yuko Hayashi Ai Kuba Shigeko Tokuda Tomoko Sonobe
*Pre-Meeting Lecture: Koji Eguchi	
*Open Submission: Keita Tsuji	

ntcir4-ov 2004-06-02

16

## Cross-Language Information Retrieval (CLIR) Task

### Task Organizers

Kazuaki Kishida\*, Kuang-hua Chen, Sukhoon Lee,  
Hsin-Hsi Chen, Koji Eguchi, Noriko Kando  
Kazuko Kuriyama, Sung Hyon Myaeng

In Cooperation with: National Taiwan Univ, Korea  
Institute of Science and Technology Information,

## Design of CLIR Task

- **Purpose**
  - To promote researches of cross-lingual information retrieval (CLIR) on East-Asian languages
- **Languages**
  - Chinese (C), Japanese (J), Korean (K), English (E)
- **Subtasks**
  - *Multilingual CLIR* (MLIR) : e.g., C - CJKE
  - *Bilingual CLIR* (BLIR): e.g., C - J
  - *Single Language IR* (SLIR): e.g., C - C
  - *Pivot Bilingual CLIR* (PLIR): e.g., C - E - J

ntcir4-ov 2004-06-02

18

## Test Collection

- **Document sets** - News articles (1998-99)
  - Chinese: 381,681 docs
  - Japanese: 596,058 docs
  - Korean: 254,438 docs
  - English: 347,550 docs
- **Queries** - 60 topics
  - TITLE-only run, DESC-only run, others
- **Relevance Judgments** - 4 grades
  - Highly Relevant (S), Relevant (A), Partial Relevant (B), Non-Relevant (C)

ntcir4-ov 2004-06-02

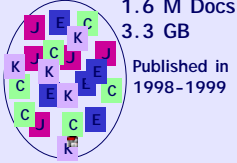
19

## NTCIR-4 CLIR (2003-2004)

60 topics



Documents



- Short Q: D-only and T-only are mandatory
- Background info of search requests
- Balance btw topic-types:
  - specific (ex. Particular event) vs generic
  - proper nouns vs without PN
  - domestic/regional/international

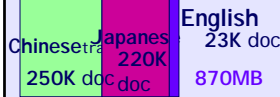
ntcir4-ov 2004-06-02

20

## Documents for CLIR at NTCIR

NTCIR-3 (2002)

Published in 1998-1999



Published in 1994



NTCIR-4 (2003-2004)

Published in 1998-1999



3.3GB

Good balance btw 4 langs

Every language is multi-source

ntcir4-ov 2004-06-02

## Submission of results

- 26 groups submitted results
  - From Australia, Canada, China, Hong Kong, Japan, Korea, Singapore, Switzerland, Taiwan, USA (10 countries and areas)
- No. of runs
  - SLIR: 182 runs from 19 groups
  - BLIR (or PLIR): 149 runs from 17 groups
  - MLIR: 37 runs from 5 groups
  - TOTAL: 368 runs

ntcir4-ov 2004-06-02

22

## Techniques for CLIR

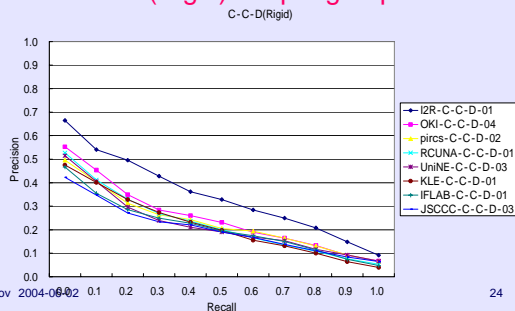
- Indexing, Stop Words, Decomposing
- Query and Document translation
  - MT, MRD, Parallel corpora
- Translation disambiguation
- Out-of-vocabulary problem
  - Use of Web resources
  - Transliteration - Cognate
- Query expansion techniques
  - Pseudo-relevance feedback
  - Use of Knowledge ontology
- Merging strategies

ntcir4-ov 2004-06-02

23

## Evaluation (2)

- SLIR: C-C-D (Rigid) – top 8 groups

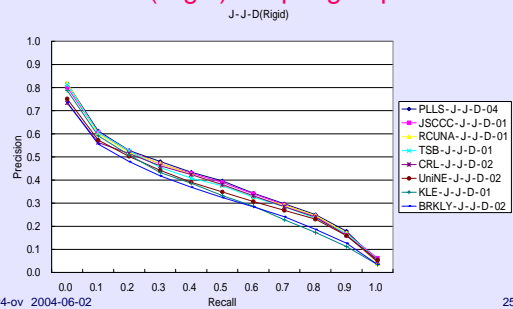


ntcir4-ov 2004-06-02

24

## Evaluation (3)

- SLIR: J-J-D (Rigid) – top 8 groups

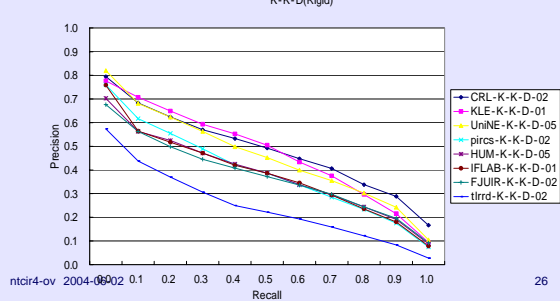


ntcir4-ov 2004-06-02

25

## Evaluation (4)

- SLIR: K-K-D (Rigid) – top 8 groups

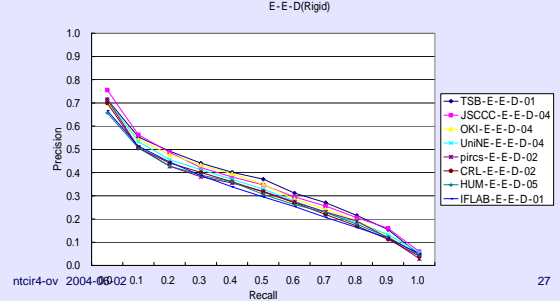


ntcir4-ov 2004-06-02

26

## Evaluation (5)

- SLIR: E-E-D (Rigid) – top 8 groups

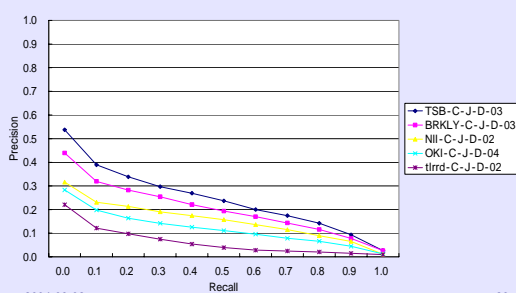


ntcir4-ov 2004-06-02

27

## Evaluation (6)

- SLIR: C-J-D (Rigid) – top groups



ntcir4-ov 2004-06-02

28

## Evaluation (6)

- BLIR – Comparison of MAP values between best SLIR and best BLIR runs (D-run, Rigid)

	C-C .3255		J-J .3804
J-C	.0548 16.8%	C-J	.2309 60.7%
K-C	.1447 44.5%	K-J	.2935 77.2%
E-C	.0663 20.4%	E-J	.2674 70.3%
	K-K .4685		E-E .3469
C-K	.3973 84.8%	C-E	.2238 64.5%
J-K	.3984 85.0%	J-E	.3340 96.2%
E-K	.3249 69.3%	K-E	.2250 64.9%

ntcir4-ov 2004-06-02

29

## Conclusion

- Various techniques for improving search performance were used.
- BLIR on Chinese document sets shows relatively low performance. Meanwhile, BLIR on Korean documents seems to reach at a very high level.
- Performance of pivot language approach is lower than non-pivot runs.
- Performance of MLIR was low.

ntcir4-ov 2004-06-02

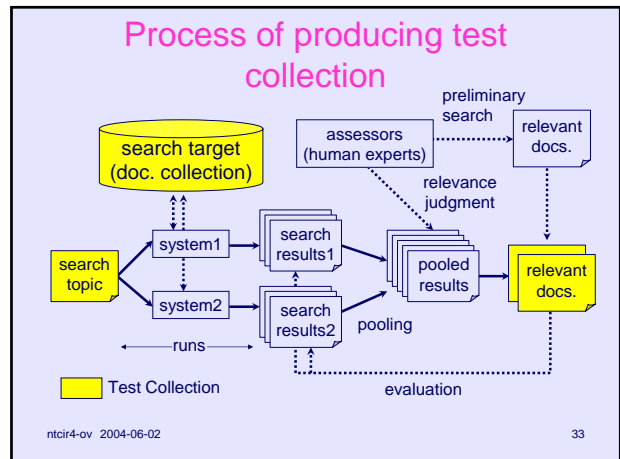
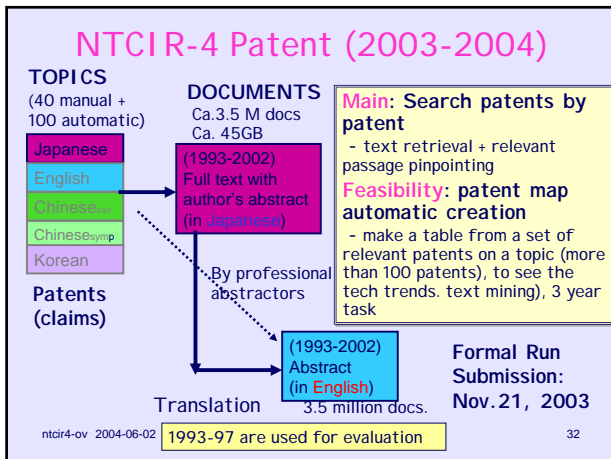
30

## Patent Retrieval Task

### Task Organizers

Atsushi Fujii (Univ of Tsukuba)  
Makoto Iwayama (TIT/Hitachi)  
Noriko Kando (NII)

In Cooperation with: Japan Intellectual  
Property Association (JIPA)



- ### Search topics
- Japanese patent application rejected by Japanese Patent Office (JPO)
  - 34 topics were selected and judged by members of "Japan Intellectual Property Association" (JIPA)
  - 69 additional topics: applications rejected by JPO/ used the citations only
  - Quite few relevant documents
  - English, Korean, and simplified/traditional Chinese translations were also produced for cross-language patent IR
- ntcir4-ov 2004-06-02 34

### Example search topic

Date of filing

```

<TOPIC>
<NUM>008</NUM>
<LANG>EN</LANG>
<FDATE>19960527</FDATE>
<CLAIM>(Claim 1) A sensor device, characterized in that an open recessed part is formed on a box-shaped forming base, a conductive film of a designated pattern is formed on the surface of the forming base including the inner surface of the recessed part, an element for a sensor is bonded to the recessed part, and the forming base is closed with a cover.</CLAIM>
...
</TOPIC>

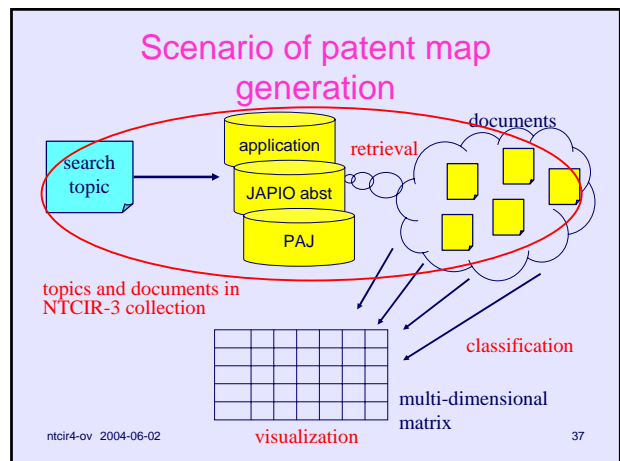
```

Relevant documents must be prior art, which had been open to the public before the topic patent was filed

Target for invalidation

ntcir4-ov 2004-06-02 35

- ### Relevance judgment
- **Document-based** relevant judgment
    - A: patent that can invalidate the topic claim
    - B: patent that can invalidate the topic claim, when used with other patents
  - **passage-based** relevant judgment:
    - **combinational relevance**
  - Submitted runs were evaluated by mean average precision (MAP)
- ntcir4-ov 2004-06-02 36



### Example map (blue light-emitting diode)

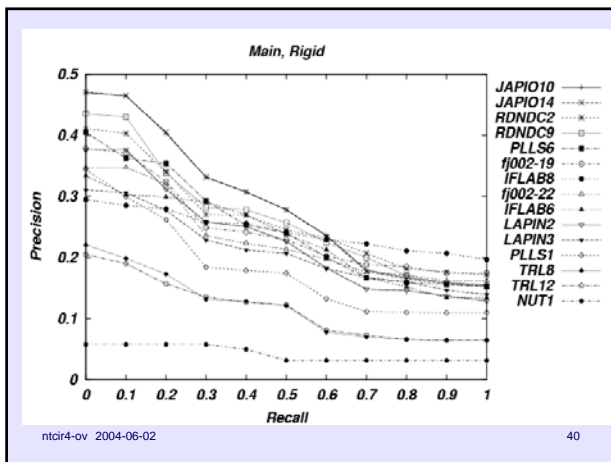
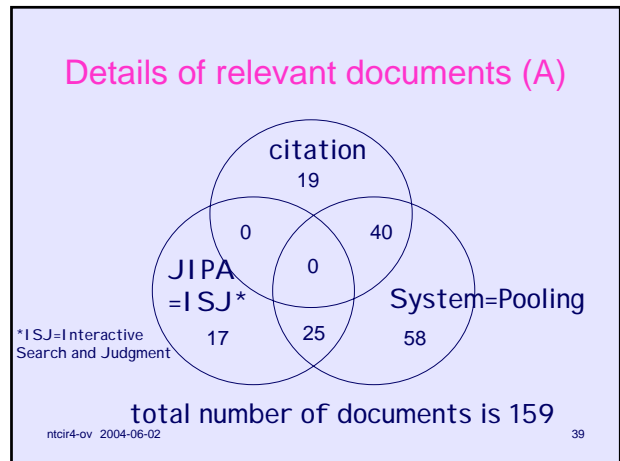
given → problems to be solved

	crystalline	reliability	long operating life	emission stability	emission intensity
structure of active layer			1998-145000 1998-233554		
electrode composition		1998-107318		1998-190063 1998-209498	1998-209495
electrode arrangement		1998-215034 1998-223930	1998-242518	1998-173230 1998-209499 1998-256602	1998-242515 1998-270757
structure of light emitting element	1998-135516 1998-242586 1998-247761		1998-135514 1998-256668		1998-012923 1998-247745 1998-256597

solutions

participants identify lines and columns

ntcir4-ov 2004-06-02 38



### Question Answering Challenge (QAC-1)

Task Organizers  
 Jun'ichi FUKUMOTO  
 Tsuneaki KATO  
 Fumito MASUI

### Question Answering Challenge at NTCIR

**Subtask 1: 5 ordered answers: Eval by MRR 195Q**

**Subtask 2: 1 set of all the answers: 199Q**

Return 1 set of only and all the correct answers. Q may have multiple answers or no answers. Penalty given for wrong answers. Eval by F-measure

**Subtask 3: A series of questions. 251 Q (36 series)**

Report writing task: topic centered vs browsing, Eval by F-measure

- Exact Answers - Return in 48 hours
- Doc IDs are required as support information

NTCIR-3 (2002): Using one news source  
 NTCIR-4 (2003-2004): Two different news sources

ntcir4-ov 2004-06-02 42

### Submission

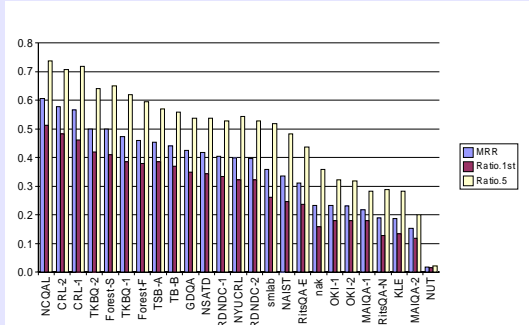
Subtask-1: 25 runs from 17 groups  
 Subtask-2: 14 runs from 9 groups  
 Subtask-3: 14 runs from 7 groups

Attributes of participating groups

- Univ 11 (9 Japan, 2 International)
- Company 5
- National Labs 2

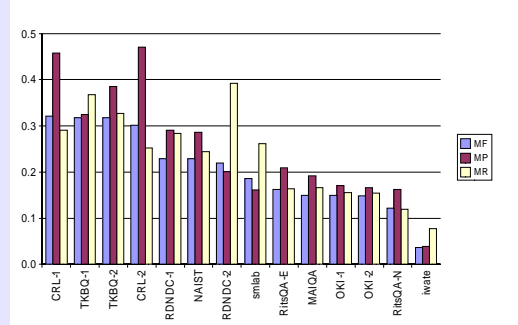
ntcir4-ov 2004-06-02 43

## Results: Subtask 1



MRR of correct ratio of 1<sup>st</sup> ranked answer and among 5<sup>th</sup> ranked ones  
ntcir4-ov 2004-06-02 44

## Results: Subtask 2



MRR of correct ratio of 1<sup>st</sup> ranked answer and among 5<sup>th</sup> ranked ones  
ntcir4-ov 2004-06-02 45

## Subtask 3 : series of Qs Background

- To make open-domain QA systems to answer series of related questions rather than isolated questions
  - To be used for gathering/browsing information interactively
  - A useful aid to report writing and summarization
- To evaluate QA systems' abilities for participating in information access dialogues
  - Context processing such as anaphora resolution and ellipses handling
  - Objective and quantitative

ntcir4-ov 2004-06-02

46

## Subtask 3 : series of Qs Situation Settings (User's Task)

- (Topic-oriented) Information gathering for writing a report on a specific topic
  - One (hidden) global topic and series of Qs on subtopics of the global topic
- Browsing along transitive interests
  - Topic or focus of the Qs are shifting through the interaction of the user and system.
  - Local coherence with the previous Q only

Answering a series of Qs has close relation with **Multi-Doc Summarization**:

- Series of Qs covers subtopics shall be contained in a summary; can be used as "quality questions",
- Summarization as pre-processing of QA?
- QA for pre-processing of Abstract-type summary generation?

## Example of Series of Questions

- When was Seiji Ozawa born?
- Where was he born?
- Which university did he graduate from?
- Who did he study under?
- Who recognized him?
- Which orchestra was he conducting in 1998?
- Which orchestra will he begin to conduct in 2002?

Series 14: Strictly Gathering Type<sub>48</sub>

ntcir4-ov 2004-06-02

## Example of Series of Questions

- Which stadium is home to the New York Yankees?
- When was it built?
- How many persons' monuments have been displayed there?
- Whose monument was displayed in 1999?
- When did he come to Japan on honeymoon?
- Who was the bride at that time?
- Who often draws pop art using her as a motif?
- What company's can did he often draw also?

Series 22: Browsing Type<sub>49</sub>

ntcir4-ov 2004-06-02



## Problem on Evaluation

- Questions: Which stadium is home to the NY Yankees?  
When was it built?
- Fact: The Yankee stadium was built in 1923  
The Shea stadium was built in 1964
- System A's Answer
  - Which stadium is home to the NY Yankees? The Shea stadium x
  - When was it built? 1964 x
- System B's Answer
  - Which stadium is home to the NY Yankees? The Shea stadium x
  - When was it built? 1923

ntcir4-ov 2004-06-02

50

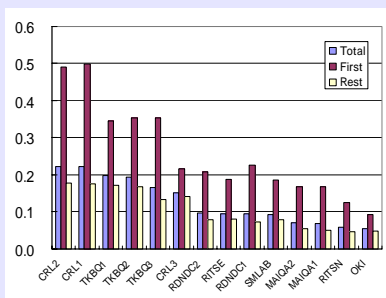
## Pragmatic Phenomena Observed

Phenomena	Occurrence
No Reference Expression	36
Pronouns	76
Zero Pronouns	134
Definite Noun Phrases	11
Ellipses	7
Total # of Questions	251

ntcir4-ov 2004-06-02

51

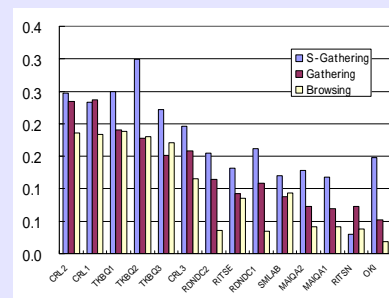
## Evaluation by MMF



ntcir4-ov 2004-06-02

52

## Differences on Series Type



ntcir4-ov 2004-06-02

53

## Text Summarization Challenge

Takahiro FUKUSIMA  
\*Tutomu HIRAO  
Hidetsugu NANBA  
Manabu OKUMURA

## Tasks

- Two kinds of summarization -

- **Extraction**
    - Extracting important sentences from document sets
    - limitation: No of sentences
  - **Abstraction**
    - Producing summaries from document sets
    - limitation: No of characters
- The limitations were given by the organizers
- The length: short, long

ntcir4-ov 2004-06-02

55

## Data

- 30 document clusters
  - 30 event document clusters selected by organizers from both Mainichi and Yomiuri Newspaper articles
    - **redundant source**
  - almost 10 documents / topic
- Data given for participants
  - 30 document clusters
  - Titles of the clusters
  - (Questions about important parts concerning document clusters generated by human summarizers)\*
  - \* The participants judges whether use this information or not

ntcir4-ov 2004-06-02

56

## Participants

- 9 participants from university and governmental research org.
  - 10 systems for extraction
  - 9 systems for abstraction
  - University: 6, Gov.+Univ.:3

ntcir4-ov 2004-06-02

57

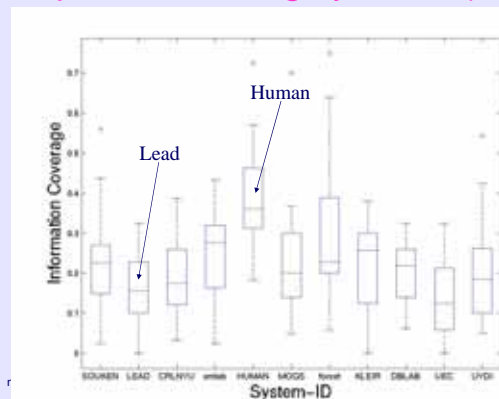
## Evaluation Methods

- Extraction
  - Intrinsic evaluation
    - Precision, Coverage
- Abstraction
  - Intrinsic evaluation
    - Content : Information Coverage
    - Readability : Quality Questions (We modified DUC's QQ for Japanese text)
  - Extrinsic evaluation
    - Pseudo-Question-Answering

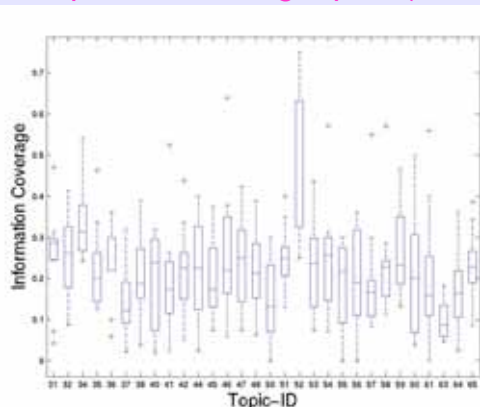
ntcir4-ov 2004-06-02

58

## Comparison among systems (short)



## Comparison among topics (short)



## Evaluation Results (Readability)

- Q04: How many expressions which have same meanings but different term are there?
  - System average: 2.05 (short), 4.16 (long)
  - Human : 0.433, 1.133
- Q08: Does the summary have wrong chronological ordering (yes→ -1, no→+1, other→0)
  - System average: -0.21, -0.58
  - Human : 0.933 0.800

ntcir4-ov 2004-06-02

61

## Web Retrieval Task

### Task Organizers

Koji Eguchi (NII), Co-Chair  
 Keizo Oyama (NII), Co-Chair  
 Akiko Aizawa (NII), Haruko Ishikawa (NII), Masatoshi Arikawa (Tokyo U), Tsuyoshi Sagara (Tokyo U), Hayato Yamana (WasedaU)

ntcir4-ov 2004-06-02

62

## NTCIR-4 WEB

### WEB Task in NTCIR-4 at a Glance

(Subtask A) Informational Retrieval Task 2

(Subtask B) Navigational Retrieval Task 1

[Pilot Task](Subtask C) Geographical Task 1

[Pilot Task](Subtask D) Topical Classification Task 1

—search result classification, eg.using clustering

### Data set:

- ‘NW100G-01’ (100GB Web page data crawled in 2001 from “\*.jp”) for Subtasks A and B
- ‘Target data’ (a comparatively small size of subsets of the NW100G-01) for Subtasks C and D.

ntcir4-ov 2004-06-02

63

## NTCIR-4 WEB

### Informational Retrieval Task2

- **Goal:** To assess effectiveness of “Subject Search”.
- **Topics:**
  - Delivered 300, Mandatory runs: TITLE, ALT(0-3) (alternate queries) or DESC -only
  - Assessed 80 topics for ‘Target’,and 35 of them for ‘Survey’
  - Remaining 48 are now under assessment,
- **Active participants:** 5 (+2 invited groups +2 organizers’ systems)
- **Submitted runs:** 135
  - Almost used content info only, some used anchor text, several used link info.
- **Pooling:** Top 100 docs for ‘Survey and top 20 docs for ‘Target’- (their possible out-linked pages were also assessed.)
- **Rel. judgment:** *highly, fairly & partially relevant, irrelevant and best 3 relevant.* “Content duplication” was judged using automatically detected candidates.

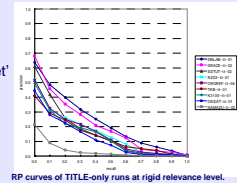
ntcir4-ov 2004-06-02

64

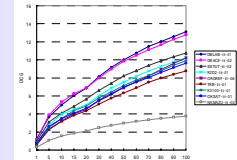
## NTCIR-4 WEB

### Informational Retrieval Task2

- **Evaluation measures:**
  - Avg.prec. and DCG for ‘Survey’.
  - Prec., DCG and WRR(MRR) @10 for ‘Target’
- **Alternative evaluation as trial:**
  - Investigated on evaluation measures reflecting users’ intuition as ‘user-oriented evaluation measures’. (see Ohtsuka in working-notes.)
- **Future work:**
  - Analyze the evaluation results further.
  - Tuning parameters of WRR (‘Weighted Reciprocal Rank’)
  - Verify stability of evaluation measures
  - Check comprehensiveness of assessment results
  - Analyze topic-by-topic behavior of each



RP curves of TITLE-only runs at rigid relevance level.



DCG curves of TITLE-only runs at rigid relevance level.

ntcir4-ov 2004-06-02

## NTCIR-4 WEB

### Navigational Retrieval Task 1

- **Goal:** To assess retrieval effectiveness of “Known Item Search”.
- **Topics:**
  - Delivered 300, Mandatory run : using TITLE only.
  - Assessed 144 and used 87 or 72 for evaluation depending on document set definition. Remaining 156 are now under assessment.
- **Active participants:** 5 (+organizers)
- **Submitted runs:** 16 (+68 by organizers).
  - 8 (+24) runs used anchor text, 4 (+40) runs used link info, 4 (+4) runs used content info only.
- **Pooling:** Top 10 docs (their possible out-linked pages were also assessed.)
- **Rel. judgment:** *relevant, partially relevant, non-relevant.* “Representativeness” was judged based on every available information, e.g., provider of the page, content (text, images, etc.), URL, out-linked pages.

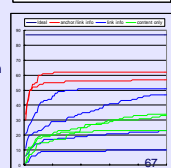
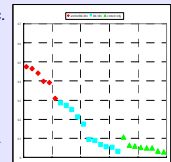
ntcir4-ov 2004-06-02

66

## NTCIR-4 WEB

### Navigational Retrieval Task 1

- **Evaluation measures:** DCG and MRR at top-10 doc. level
- **Evaluation result:**
  - Tendency on MRR & DCG
    - Several anchor-base systems performed best.
    - Several link-base systems performed fairly.
    - Content-base systems performed poorly.
  - Tendency on ranks of found rel. docs
    - Anchor-base systems tend to be able to pinpoint rel. docs.
    - Link-base and content-base systems tend to return noisy results.
- **Future work:**
  - Verify stability of evaluation measures
  - Check comprehensiveness of assessment results
  - Study on evaluation measures reflecting users’ intuition
  - Analyze topic-by-topic behavior of each systems
- **WEB Navigational Retrieval Task at NTCIR-5:**
  - Use larger Web page data set — 300GB–1TB?
  - Crawl each site more deeply — >10k pages/site?
  - Reinforce “Open Laboratory”



ntcir4-ov 2004-06-02

67

## Challenges in Information Access

Scaling-up

Beyond the Heterogeneity

Language, media, document genres, etc.

Appreciate each difference

Beyond "Document" Retrieval

Answer/info in documents

"Needs" Behind the Queries

User's situation, task, problem

Beyond "topic" and "fact"

\*\*\* Evaluation methodology and metrics must reflect the social needs for the technologies.\*\*\*

ntcir4-ov

68

## Contact Info & Online Proceedings

Documents used are Asian Languages but participation from all over the world is more than welcome!!

Links to freely available resources are available. Many of them have English manual and have been used by non-Asian active participants.

Inquiries: Noriko Kando at [kando@nii.ac.jp](mailto:kando@nii.ac.jp), or NTCIR-Admin Group at [ntcdam@nii.ac.jp](mailto:ntcdam@nii.ac.jp)

Online Proceedings & other info:

<http://research.nii.ac.jp/ntcir/>

ntcir4-ov 2004-06-02

69

Thanks      Merci  
Danke schön      Gracie  
Gracias      Tack  
Köszönöm      Kiitos  
Terima Kasih      Khap Khun  
Ahsante      Tak  
謝謝      ありがとう

<http://research.nii.ac.jp/ntcir/>

ntcir4-ov 20

70