Lessons from NTCIR-4: Focusing on Evaluation of CLIR on East Asian Languages, Patent and QA

Noriko Kando*1, Kazuaki Kishida *1,*2

*1:National Institute of Informatics

*2: Surugadai University

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

{kando, kkishida} (at) nii. ac. jp

NTCIR Workshop is :

A series of evaluation workshops designed to enhance research in **information access** technologies by providing infrastructure of large-scale evaluation.

Project started late 1997, Once per 1½ years

1st : Nov.1,1998- Sept.1,1999

2nd : June,2000– March,2001

3rd : Sept 2001- Oct 2002

4th: Apr 2003 – June 2004

5th: Oct 2004 – Dec 2005

* **N**ii **T**est **C**ollection for **I**nformation **R**etrieval systems

* Co-sponsored by NII and MEXT Grant-in-Aid on Informatics

Focus of NTCIR

Lab-type IR Test

Asian Languages/cross-language

Variety of Genre

Parallel/comparable Corpus

New Challenges

Intersection of IR + NLP

To make information in the documents more usable for users!
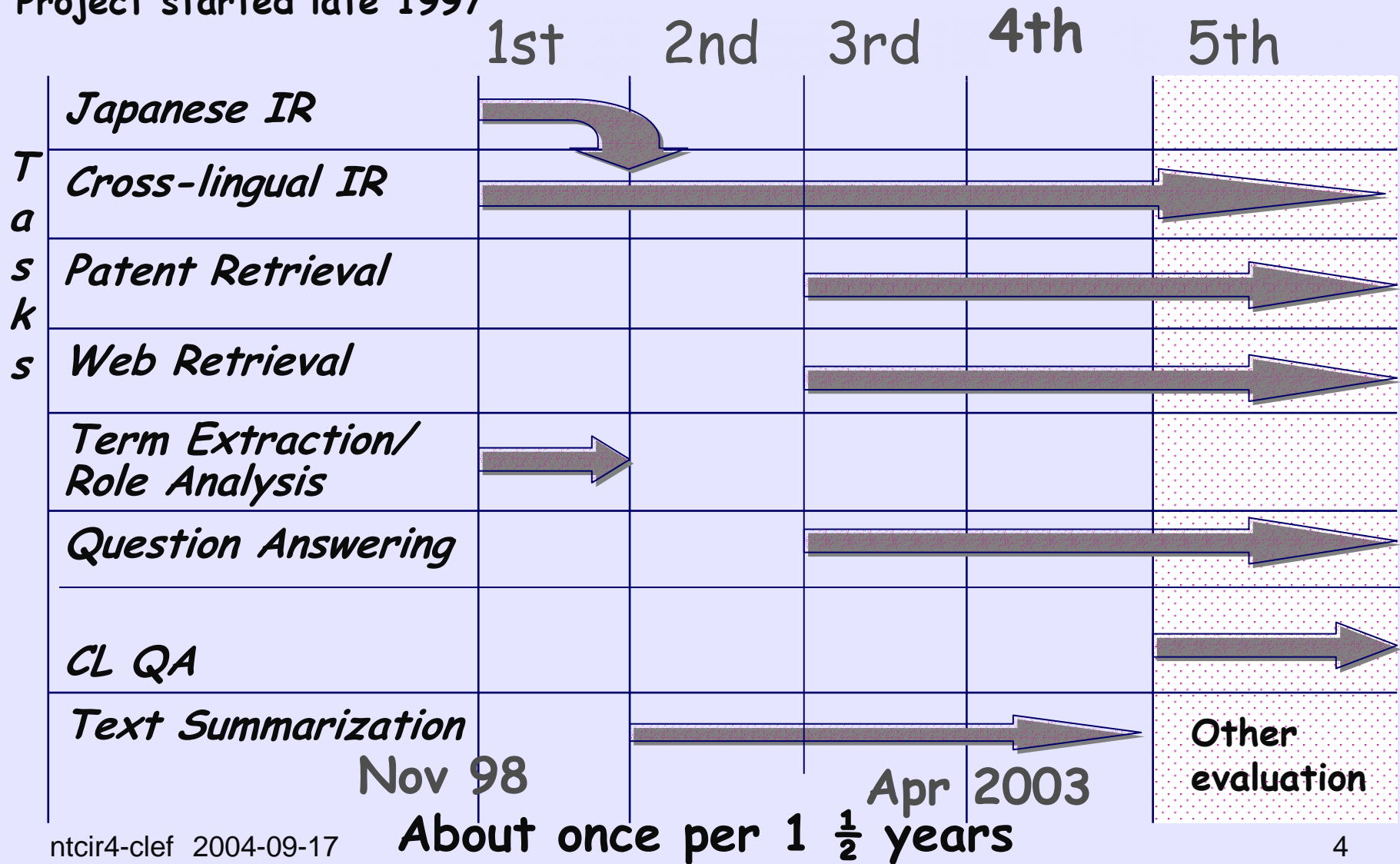
Realistic eval/user task

Forum for Researchers

Idea Exchange

Discussion/Investigation on Evaluation methods/metrics

Tasks (Research Areas) of NTCIR Workshops

Project started late 1997

Tasks	1st	2nd	3rd	4th	5th
Japanese IR					
Cross-lingual IR					
Patent Retrieval					
Web Retrieval					
Term Extraction/ Role Analysis					
Question Answering					
CL QA					
Text Summarization					Other evaluation

Nov 98   **Apr 2003**

About once per 1 ½ years

NTCIR test collections

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

Ct：Traditional Chinese、Cs：Simplified Chinese、K：Korean、J：Japanese、E：English

NTCIR Workshop 4 (2003-2004) Organizers

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

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

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

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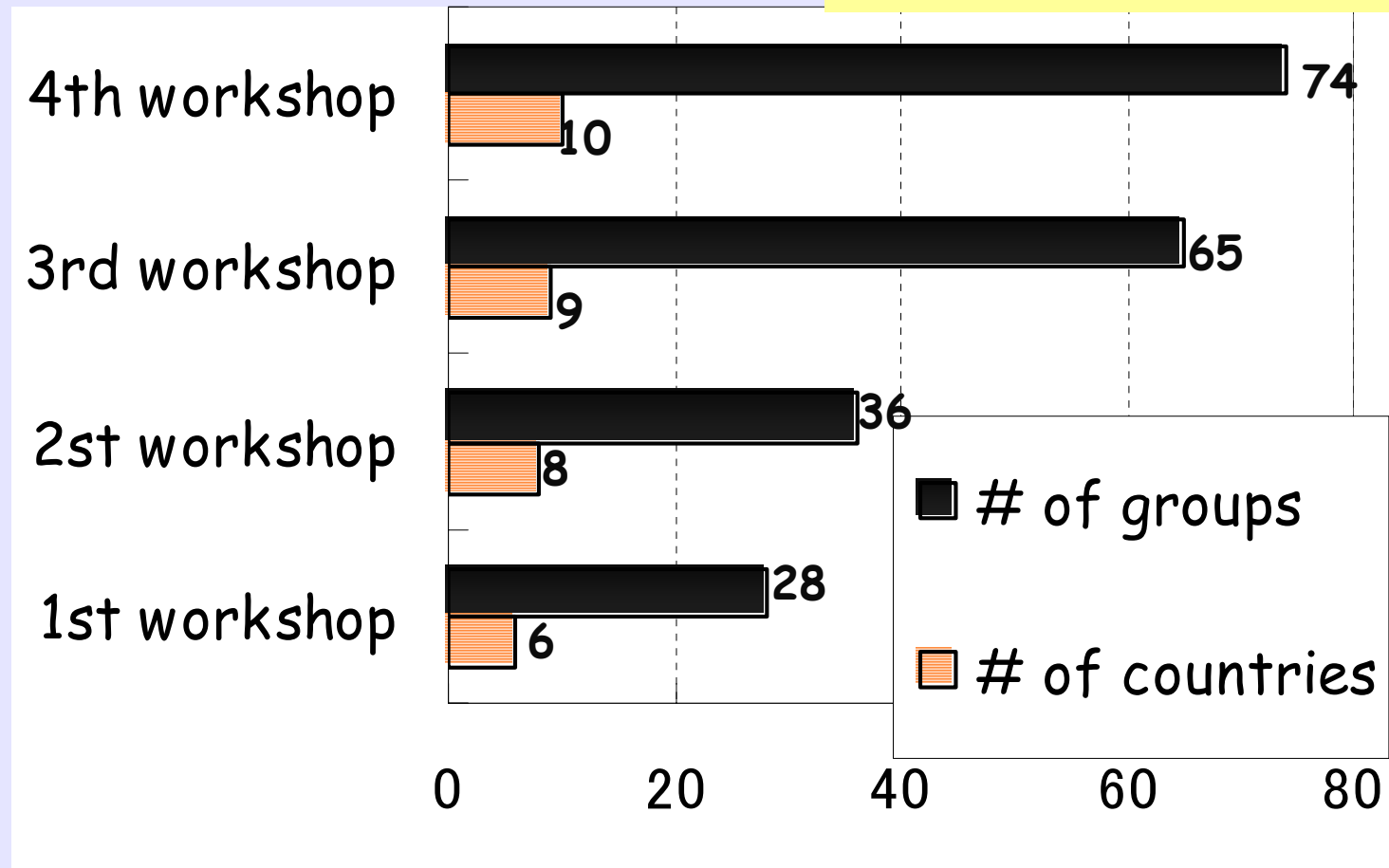Junichi Fukumoto, Ritsumeikan U
Tsuneaki Kato, U Tokyo
Fumito Masui, Mie U

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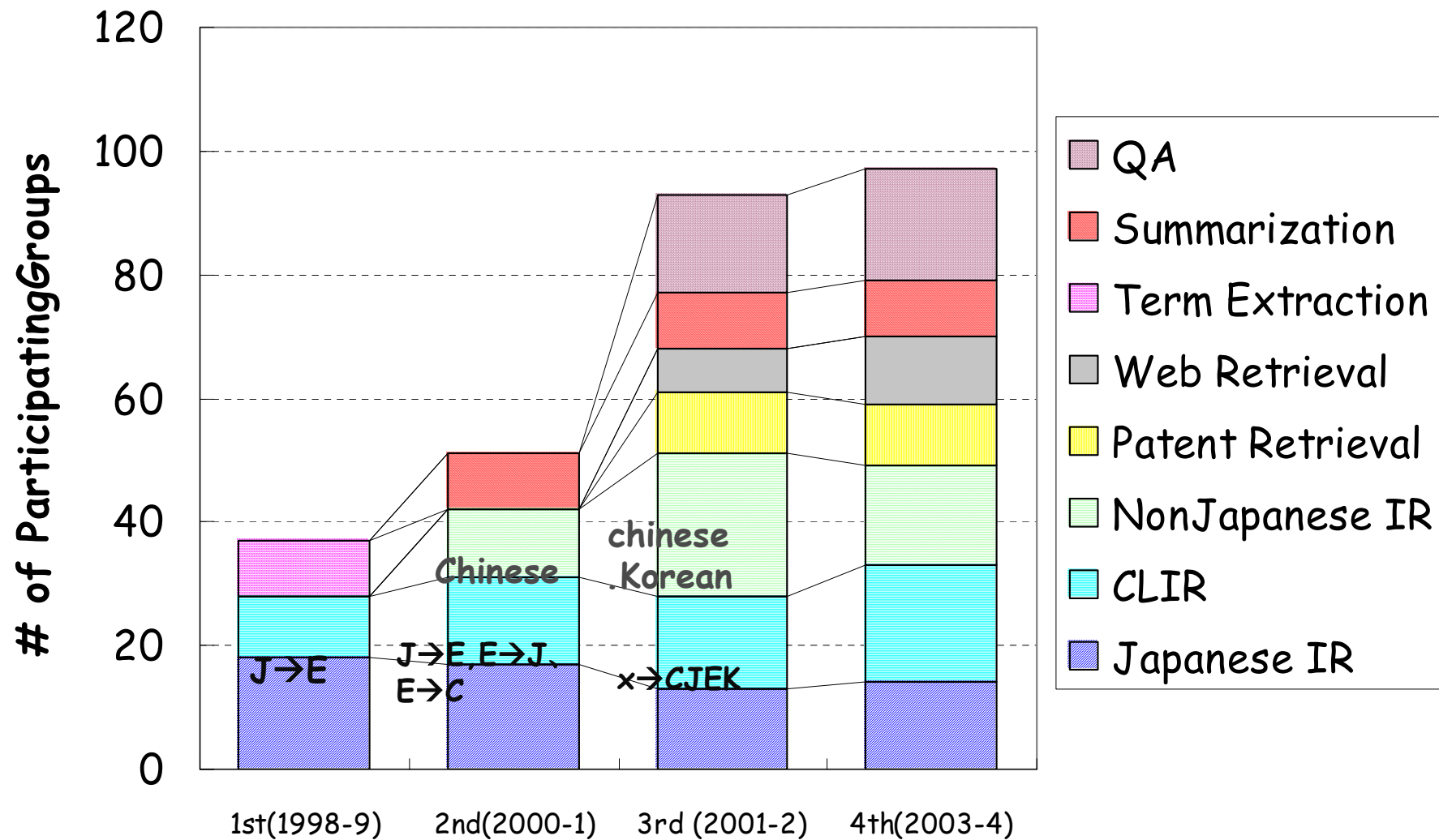Hidetsugu Nanba, Hiroshima C U
Manabu Okumura, TITEC

+WEB
Koji Eguchi, NII
Keizo Oyama, NII

NTCIR workshop:
Number of Participating Groups

74 groups from 10 countries

4th workshop	74
	10
3rd workshop	65
	9
2st workshop	36
	8
1st workshop	28
	6

0   20   40   60   80

■ # of groups

▨ # of countries

Number of Participants by Tasks

[CLIR]
Chinese Academy of Sciences (China PRC)
Clairvoyance Corporation and Justsystem (USA)
Communications Research Laboratory-1 (Japan)
Fu Jen Catholic University (Taiwan ROC)
Hong Kong Polytechnic University (Hong Kong, China PRC)
Hummingbird (Canada)
Institute of Inforcomm Research (Singapore)
Korea University (Korea)
Nara Institute of Science and Technology-1(Japan)
National Institute of Informatics-1 (Japan)
National Taiwan University (Taiwan ROC)
Oki Electric-1 (Japan)
PATOLIS (Japan)
Pohang University of Science and Technology (Korea)
Queens College City Univiersity of New York (USA)
Ricoh-1 (Japan)
Royal Melbourn Intitute of Technology (Australia)
Thomson Legal and Regulatory (USA)
Tianjin University (China PRC)
Toshiba (Japan)
University of Arizona (USA)
University of California Berkeley (USA)
University of Chicago (USA)
University of Neuchatel (Switzerland)
University of Tsukuba (Japan)
Yokohama National University (Japan)

[PATENT]
Fujitsu Laboratories (Japan)
IBM Research (Japan)
Japan Patent Information Organization / Hitachi (Japan)
Nagaoka University of Technology (Japan)
NTT DATA (Japan)
Osaka Kyoiku University (Japan)
PATOLIS (Japan)
Ricoh-2 (Japan)
Tokyo Institute of Technology (Japan)
University of Tsukuba (Japan)

[QAC]
AIST/University of Nagoya/Univeristy of Tsukuba (Japan)
Communications Research Laboratory-1 (Japan)
Iwate Prefectural University (Japan)
Keio University (Japan)
Matsushita Electoric Industiral-1 (Japan)
Mie University (Japan)
Nagaoka University of Technology (Japan)
Nara Institute of Science and Technology-2 (Japan)
New York University (USA)/Communication Research Lobaratory-2 (Japan)
NTT Communication Science Laboratories-1 (Japan)
NTT DATA (Japan)
Oki Electric-2(Japan)
Pohang University of Science and Technology (Korea)
Ritsumeikan University (Japan)
Toshiba (Japan)
Toyohashi University of Technology-1 (Japan)
University of Tokyo-1 (Japan)
Yokohama National University (Japan)

[TSC]
Communications Research Laboratory-2 (Japan) / New York University (USA)
Graduate University for Advanced Studies (Japan)
Hokkaido University (Japan)
Pohang University of Science and Technology (Korea)
Ritsumeikan University (Japan)
Toyohashi University of Technology-1 (Japan)
University of Electro-Communications (Japan)
University of Tokyo-1 (Japan)
Yokohama National University (Japan)

[WEB]
Hokkaido University (Japan)
Ibaraki University (Japan)
Matsushita Electoric Industiral-2 (Japan)
NEC (Japan)
NII-2/Univ. of Tokyo-2/KYA Group (Japan)
NTT Communication Science Laboratories-2 (Japan)
Osaka Kyoiku University (Japan)
Tokyo Metropolitan University (Japan)
Toyohashi University of Technology-1 (Japan)
Toyohashi University of Technology-2 (Japan)
University of Tsukuba/University of Nagoya

Participants

74 groups from 10 countries

CLIR Participants

Chinese Academy of Sciences (China PRC)

Clairvoyance Corp and Justsystem (USA)

Communications Research Laboratory-1 (Japan)

Fu Jen Catholic U (Taiwan ROC)

Hong Kong Polytechnic U (Hong Kong, China PRC)

Hummingbird (Canada)

Institute of Inforcomm Research (Singapore)

Korea U (Korea)

Nara Institute of Science and Technology-1(Japan)

National Institute of Informatics-1 (Japan)

National Taiwan U (Taiwan ROC)

Oki Electric-1 (Japan)

PATOLIS (Japan)

Pohang U of Science and Technology (Korea)

Queens College City U of New York (USA)

Ricoh-1 (Japan)

Royal Melbourn Intitute of Technology (Australia)

Thomson Legal and Regulatory (USA)

Tianjin U (China PRC)

Toshiba (Japan)

U of Arizona (USA)

U of California Berkeley (USA)

U of Chicago (USA)

U of Neuchatel (Switzerland)

U of Tsukuba (Japan)

Yokohama National U (Japan)

Schedule for NTCIR-4

April 2003:   Document Release

June-Sept, 2003: Dry Run

Oct-Dec, 2003:   Formal Run

20 Feb 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 Submission Due

31 Oct 2004: Formal Proceedings incl Open Session

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

Acknowledgment

- Japan Intellectual Property Association

  Korea Institute of Science and Technology Information (KISTI).
  National Taiwan University

- Central Daily News
- China Daily News
- China Times Inc.
- Chosunilbo
- Hankooki.com
- Industrial Property Cooperation Center
- Japan Parent Office
- Japan Patent Information Organization

Korea Economic Daily
Linguistic Data Consortium
Mainichi Newspaper
Nippon Database Kaihatsu, Co. Ltd.
NRI Cyber Patent
PATOLIS
the Sing Tao Group
Taiwan News
UDN.COM
Wisers Information Ltd.
Yomiuri Shinbun

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

- ## 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

- ## Languages

  - Chinese (C), Japanese (J), Korean (K), English (E)

Test Collection

- Document sets – News articles (1998-99)
  - Chinese: 381,681 docs   - 6 sources
  - Japanese: 596,058 docs – 2 sources
  - Korean: 254,438 docs   - 2 sources
  - English: 347,550 docs   - 2 large + 5 small sources
- Queries – 60 topics
- Relevance Judgments – 4 grades
  - Highly Relevant (S), Relevant (A), Partial Relevant (B), Non-Relevant (C)
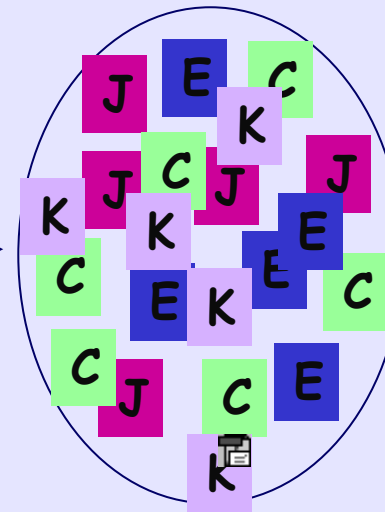- Mandatory Runs
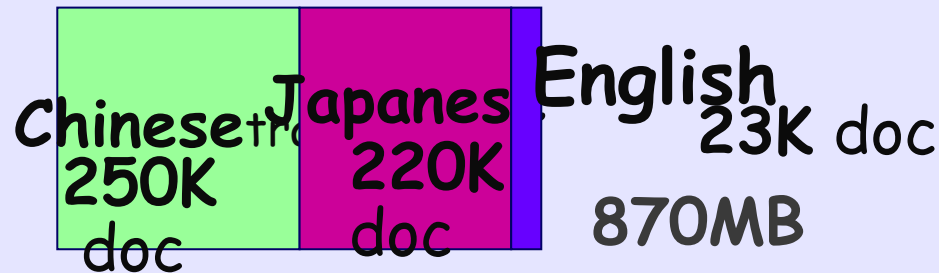 - TITLE-only run, DESC-only run

NTCIR-4 CLIR

60 topics

Chinesetrad
Korean
English
Japanese

Documents

1.6 M Docs
3.3 GB

Published in
1998-1999

- **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

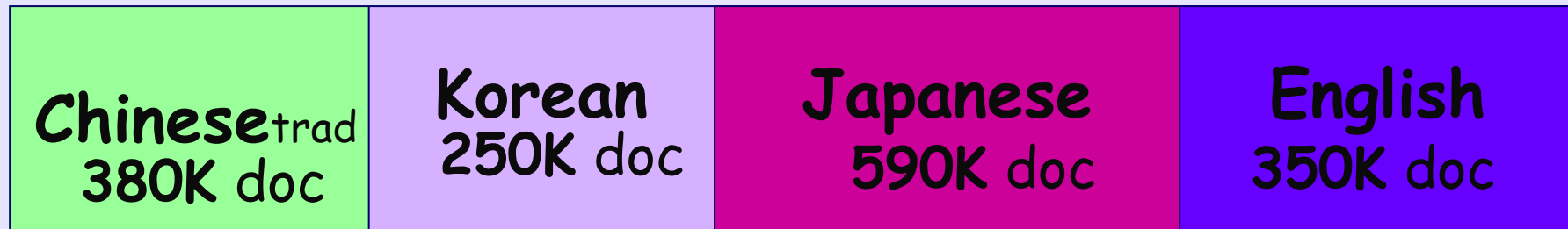# Documents for CLIR at NTCIR

NTCIR-3

Published in 1998-1999

Chinese trad
250K doc

Japanese
220K doc

English
23K doc

870MB

Published in 1994

Korean
66K doc

NTCIR-4

Published in 1998-1999

| Chinese trad 380K doc | Korean 250K doc | Japanese 590K doc | English 350K doc |

3.3GB

Good balance btw 4 languages.
Every language is multi-sources.

Result Submission

- ## 26 groups submitted results
  - From Australia, Canada, China, Hong Kong, Japan, Korea, Singapore, Switzerland, Taiwan, USA (10 countries and areas)
- ## Number of submitted runs
  - SLIR: 182 runs from 19 groups
  - BLIR (or PLIR): 149 runs from 17 groups
  - MLIR: 37 runs from 5 groups
  - TOTAL: 368 runs

Techniques Used

- Indexing, Stop Words, Decompounding
- Mostly "Query Trans", but one "Bi-Directional"
- Query and Document translation
  - MT, MRD, Parallel corpora
- Translation disambiguation
- Out-of-vocabulary (OOV) problem
  - Use of Web resources
  - Transliteration   - Cognate
- Query expansion techniques
  - Pseudo-relevance feedback, FPRF
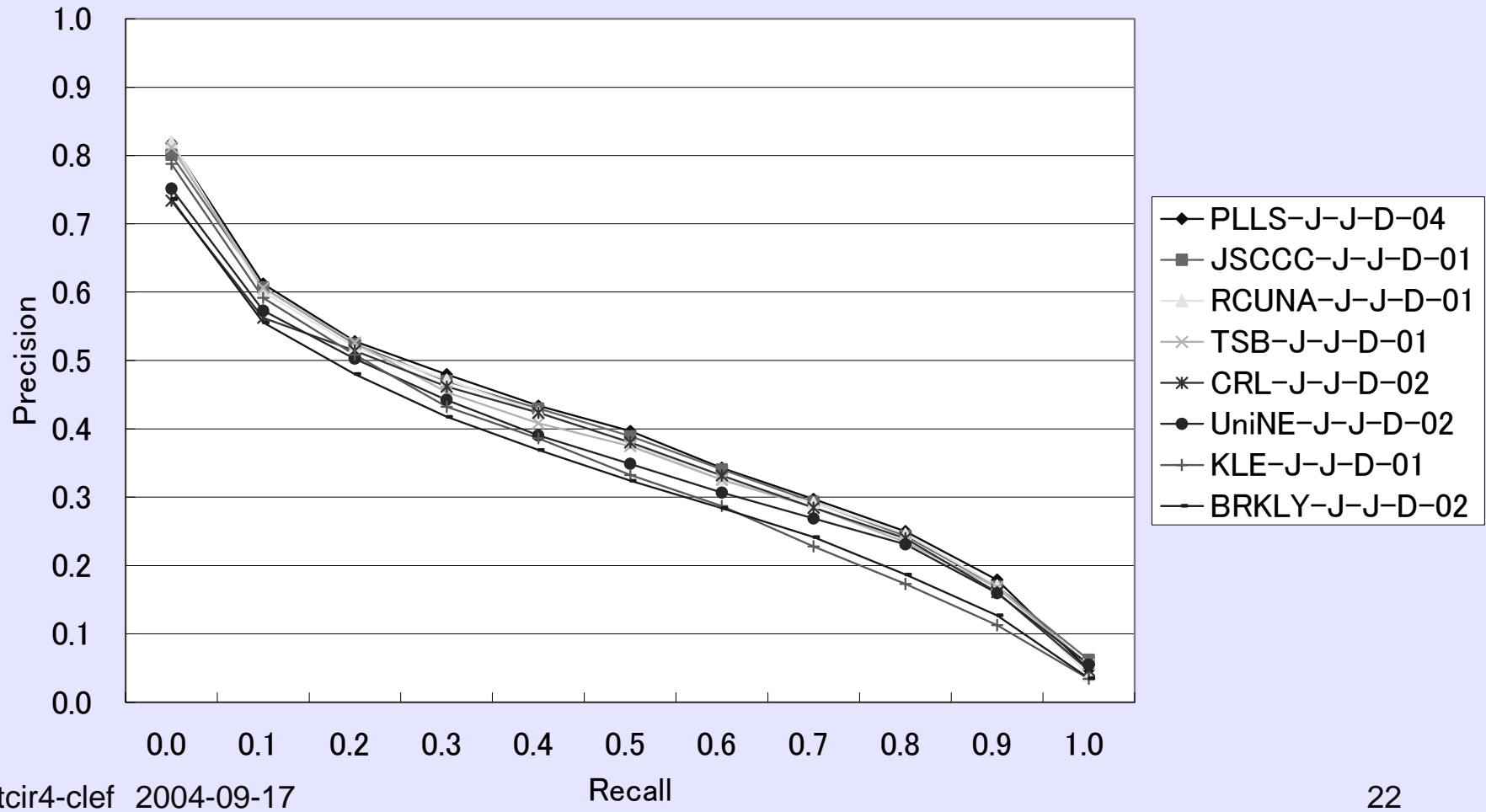 - Use of Knowledge ontology
- Merging strategies
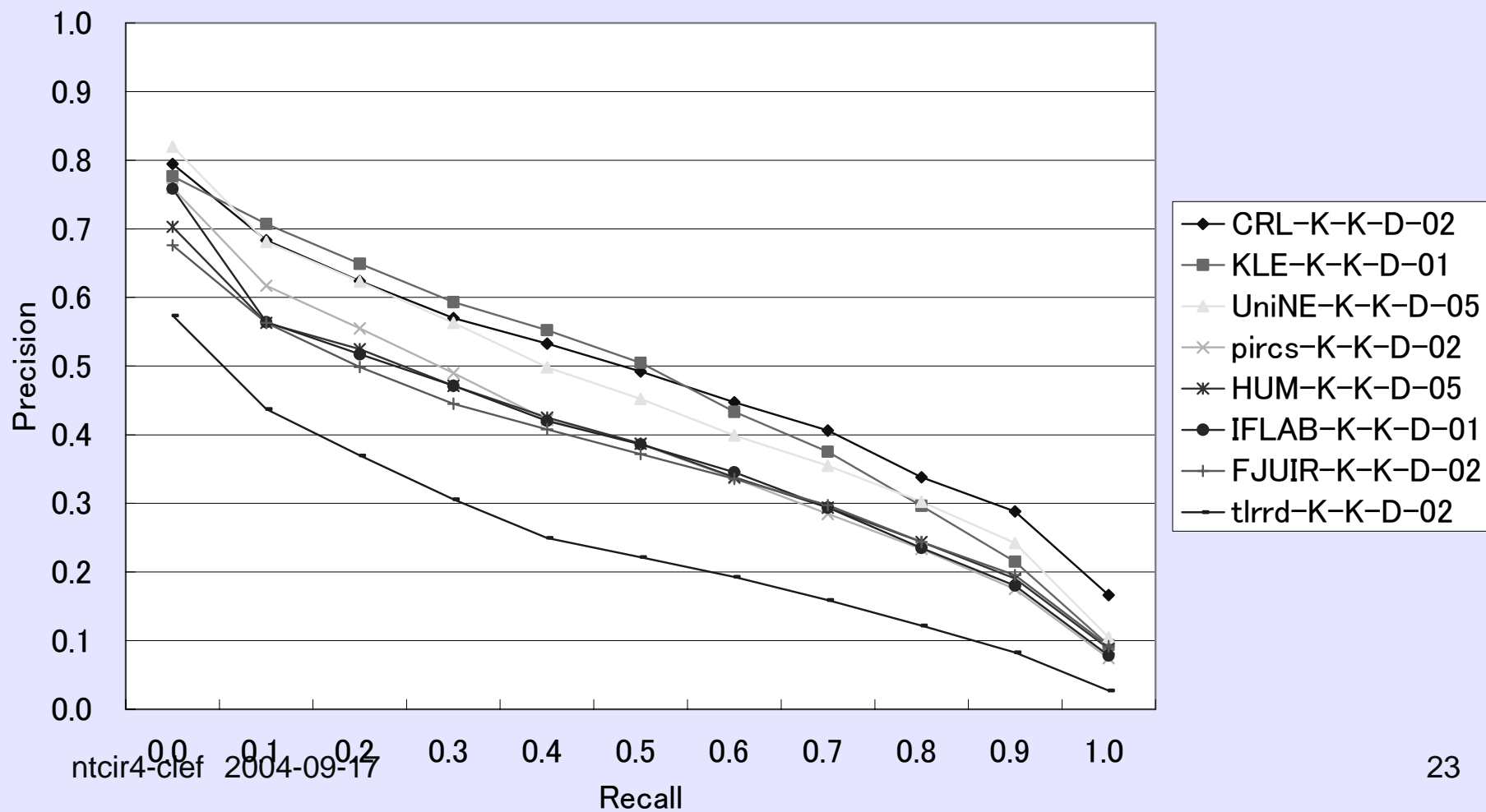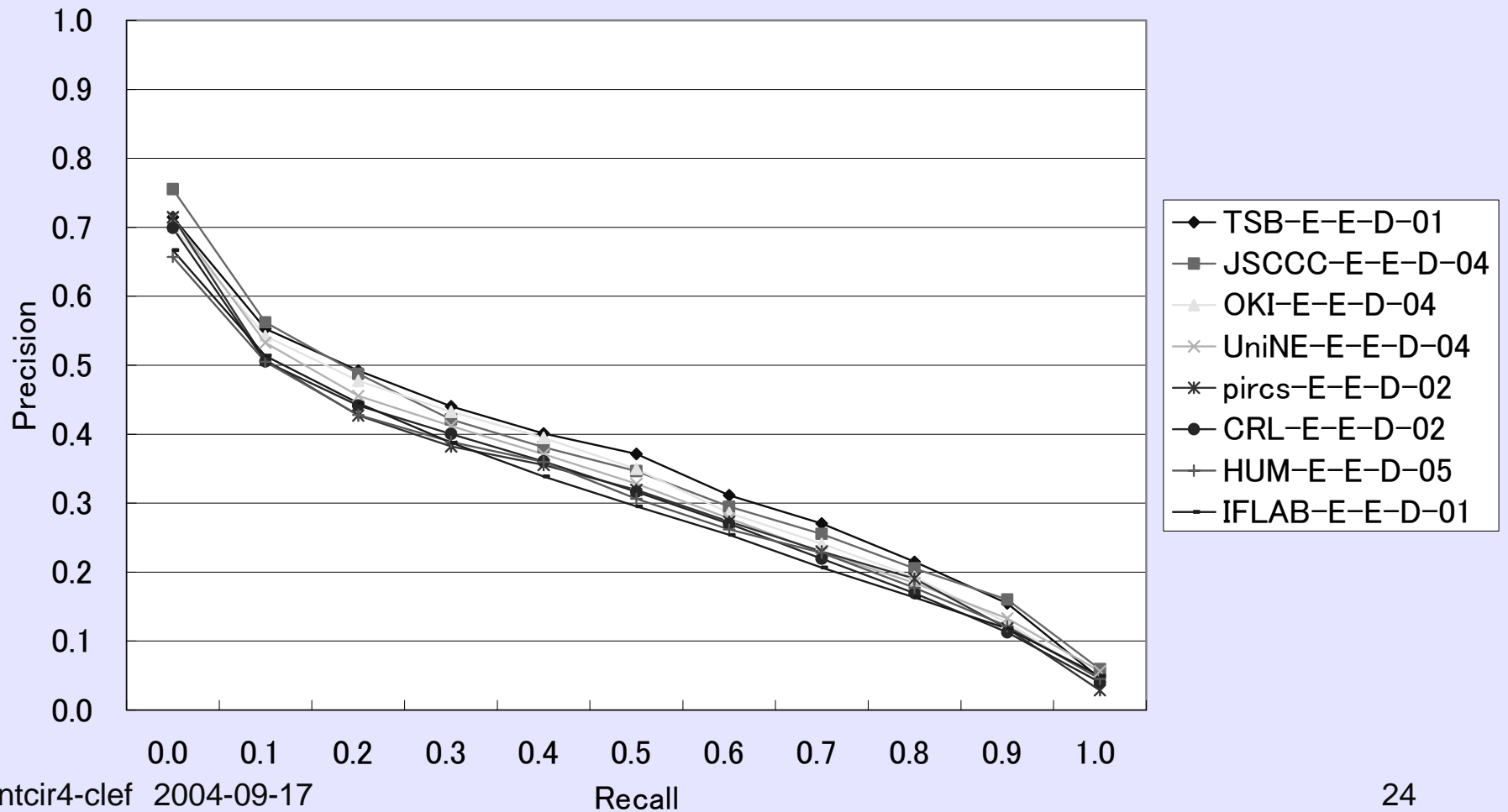
SLIR: C-C-D (Rigid)

C-C-D(Rigid)

Legend:
- I2R-C-C-D-01
- OKI-C-C-D-04
- pircs-C-C-D-02
- RCUNA-C-C-D-01
- UniNE-C-C-D-03
- KLE-C-C-D-01
- IFLAB-C-C-D-01
- JSCCC-C-C-D-03

I2R: Ontology

Precision (y-axis), Recall (x-axis)
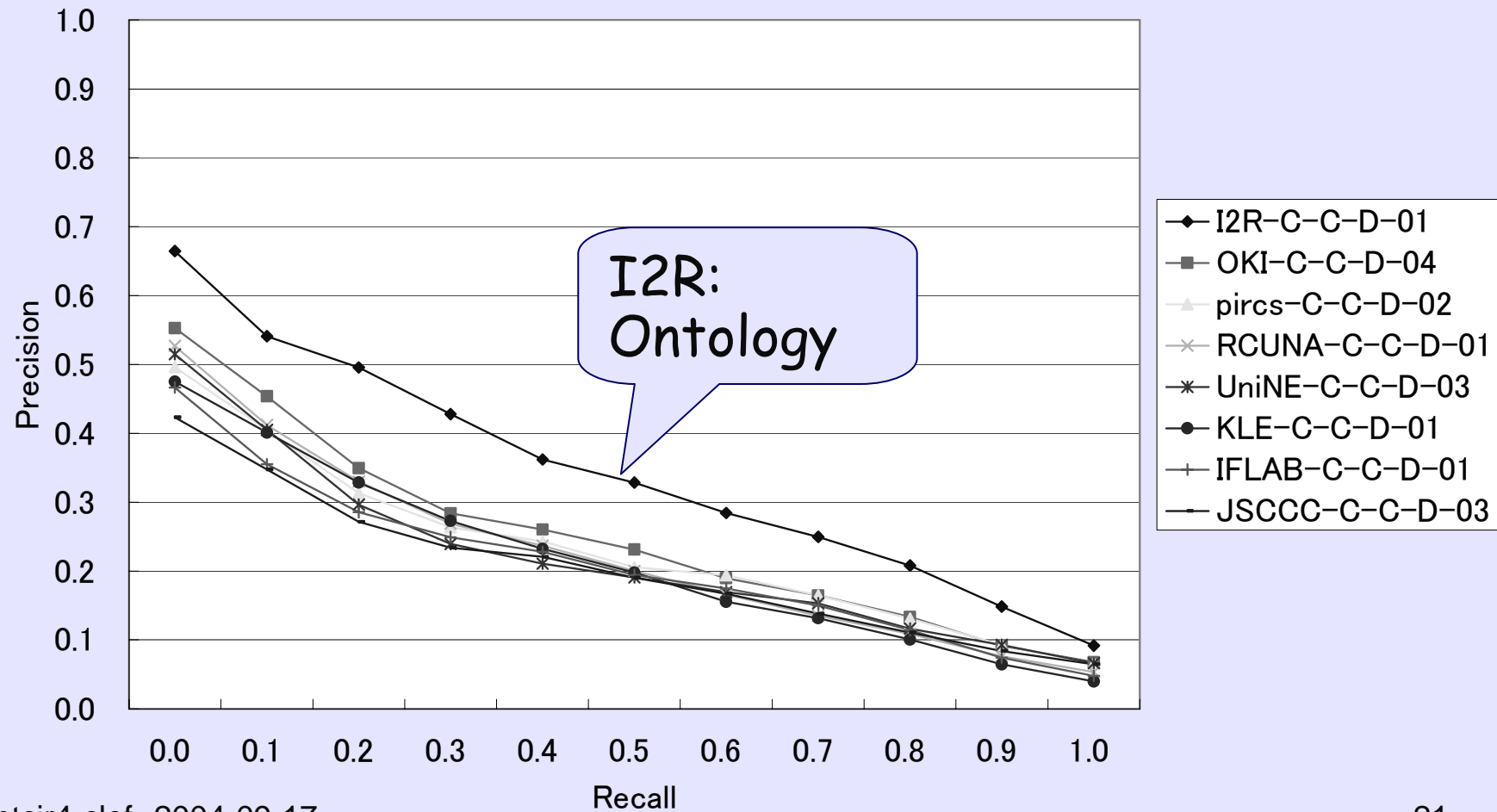
SLIR: J-J-D (Rigid)

J-J-D(Rigid)

SLIR: K-K-D (Rigid)

K-K-D(Rigid)

Legend:
- CRL-K-K-D-02
- KLE-K-K-D-01
- UniNE-K-K-D-05
- pircs-K-K-D-02
- HUM-K-K-D-05
- IFLAB-K-K-D-01
- FJUIR-K-K-D-02
- tlrrd-K-K-D-02

ntcir4-clef 2004-09-17

23

SLIR: E-E-D (Rigid)

E-E-D(Rigid)

Legend:
- TSB-E-E-D-01
- JSCCC-E-E-D-04
- OKI-E-E-D-04
- UniNE-E-E-D-04
- pircs-E-E-D-02
- CRL-E-E-D-02
- HUM-E-E-D-05
- IFLAB-E-E-D-01

ntcir4-clef 2004-09-17

BLIR: C-J-D (Rigid)

C-J-D(Rigid)

Legend:
- TSB-C-J-D-03
- BRKLY-C-J-D-03
- NII-C-J-D-02
- OKI-C-J-D-04
- tlrrd-C-J-D-02
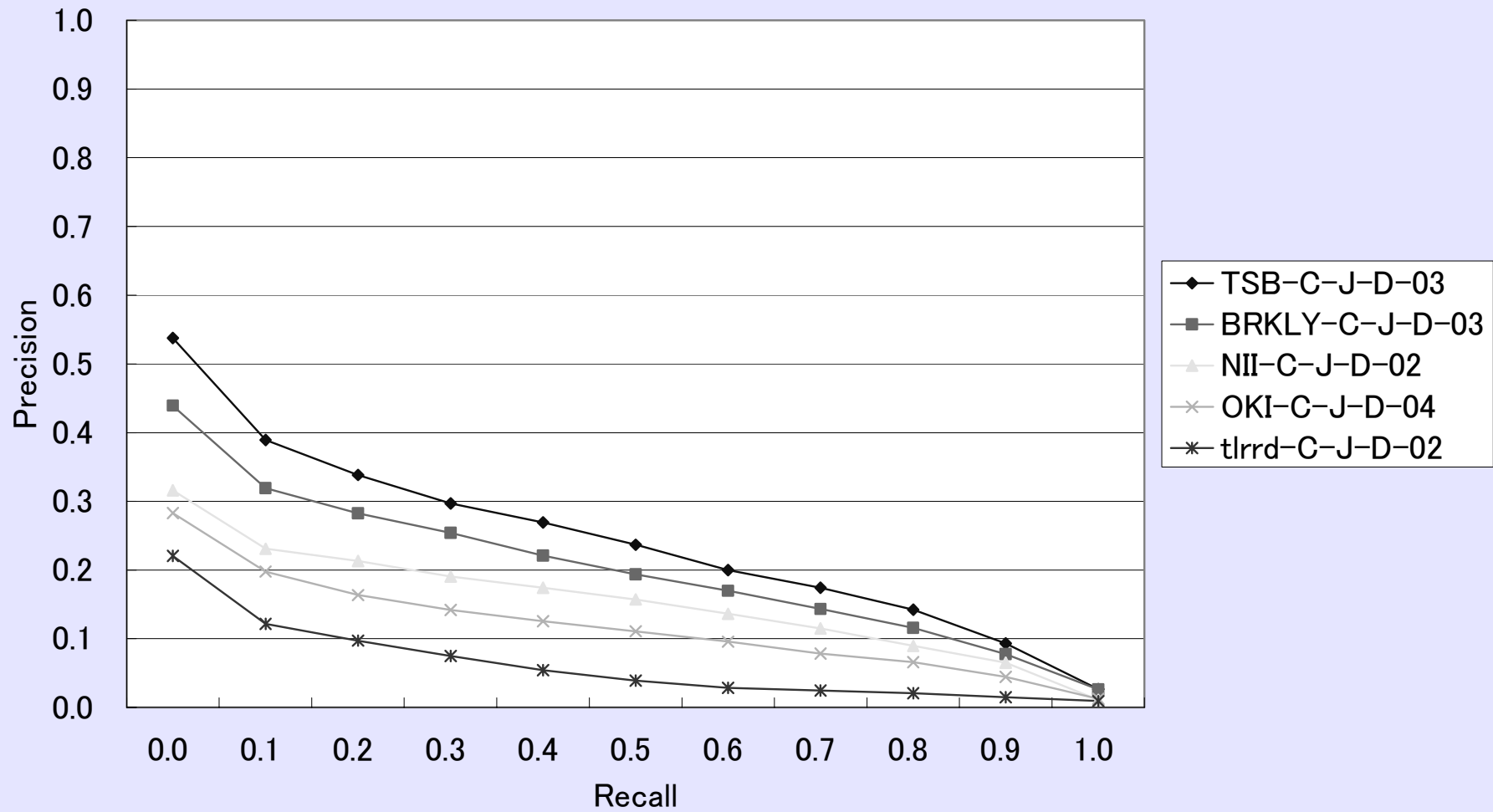
Best SLIR and BLIR runs
(D-run, Rigid) MAP and % to Monolingual

	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%

Extremely high!

NTCIR-4 CLIR summary

- Various techniques for improving search performance were used.
- BLIR on Korean doc >> on Chinese doc -- why?
- Non-pivot > pivot
- Performance of MLIR was low. More space for further investigation.

Patent Retrieval Task

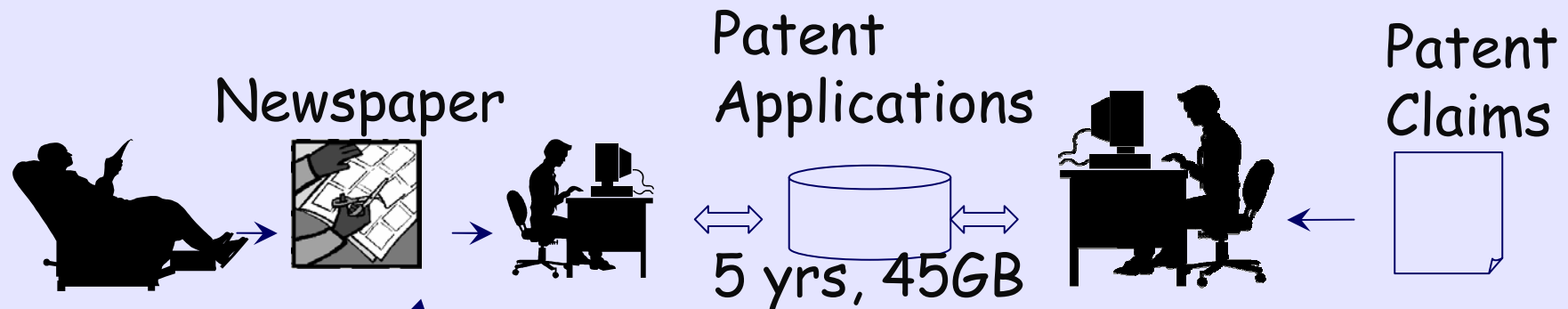## Task Organizers
Atsushi Fujii (Univ of Tsukuba)
Makoto Iwayama (TIT/Hitachi)
Noriko Kando (NII)

In Cooperation with: Japan Intellectual Property Association (JIPA)

Patent Retrieval Tasks
situation & users' information seeking task

Newspaper

Patent
Applications

Patent
Claims

5 yrs, 45GB

**NTCIR-3 PATENT
(2001-2002)**

**NTCIR-4 PATENT
(2003-2004)**

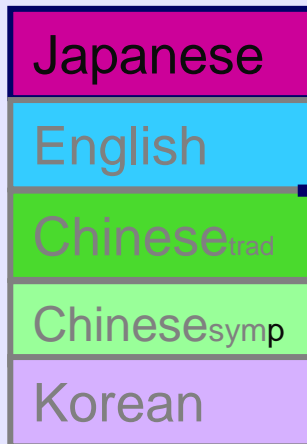Technological Survey:
Search patents by newspaper
End user: non-experts (ex.
Business manager)

From a claim of a new
patent application, search
patents that can
invalidate the new patent
application.
User: patent experts

NTCIR-4 Patent (2003-2004)

TOPICS
(34 manual +
69 automatic)

Japanese
English
Chinese_{trad}
Chinese_{symp}
Korean

**Patents
(claims)**

DOCUMENTS
Ca.3.5 M docs
Ca. 45GB

(1993-1997)
Full text with
author's abstract
(in Japanese)

By professional
abstractors

Translation

(1993-1997)
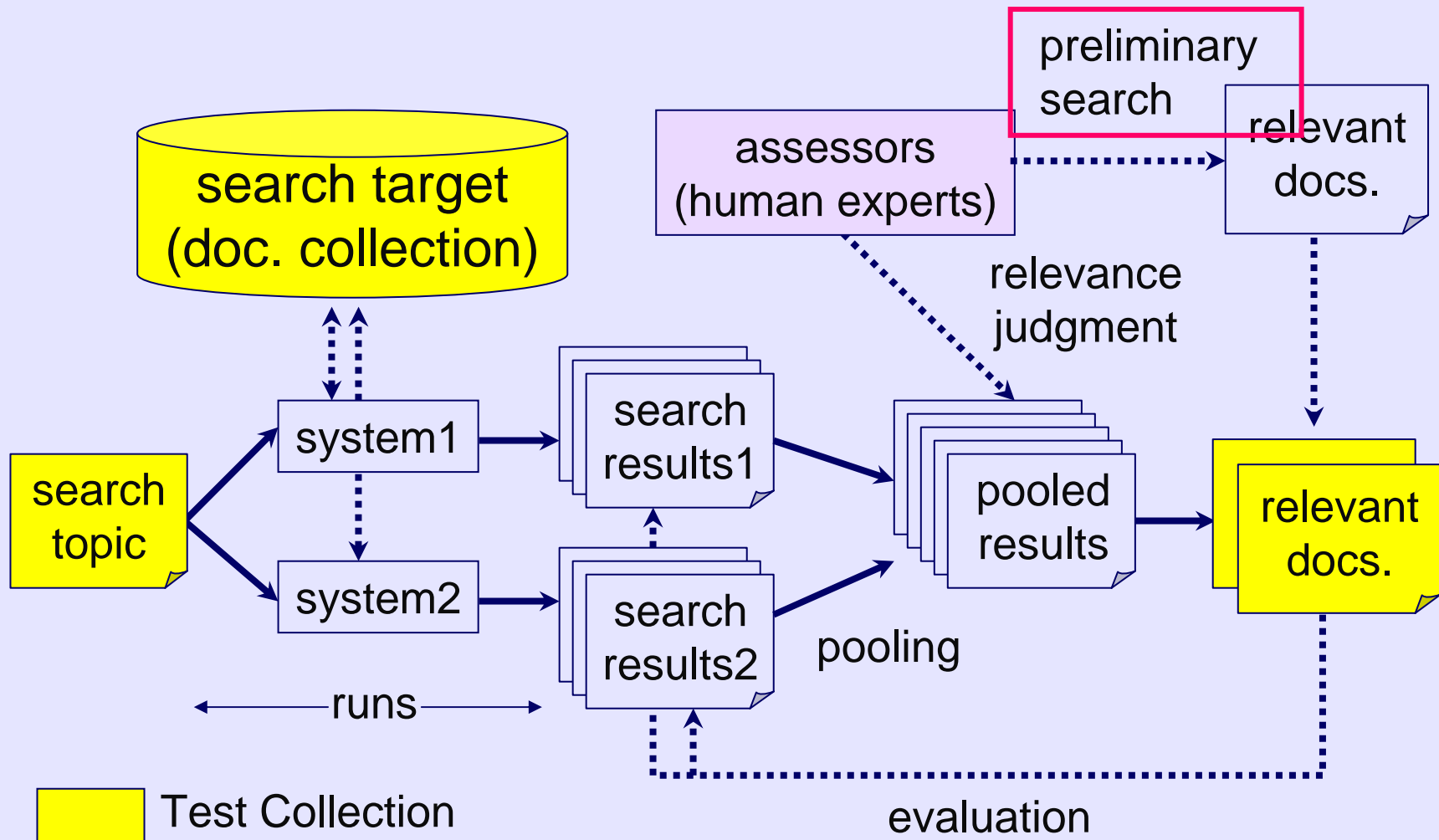Abstract
(in English)
3.5 million docs.

Main: Search patents by patent
 – text retrieval + relevant passage pinpointing

Feasibility: patent map automatic creation
 – make a table from a set of relevant patents on a topic (more than 100 patents), to see the tech trends. text mining, 3 year task

1993-97 are used for evaluation

Test Collection Creation Procedure

Search topics

- Japanese patent application rejected by Japanese Patent Office (JPO)
- 34 main topics: selected and judged by human patent experts of "Japan Intellectual Property Association" (JIPA)
- 69 additional topics: applications rejected by JPO/ evaluate by using the citations only
- Quite few relevant documents
- English, Korean, and simplified/traditional Chinese translations topics for CL patent IR

Example search topic

Date of filing

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

```
<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>
```
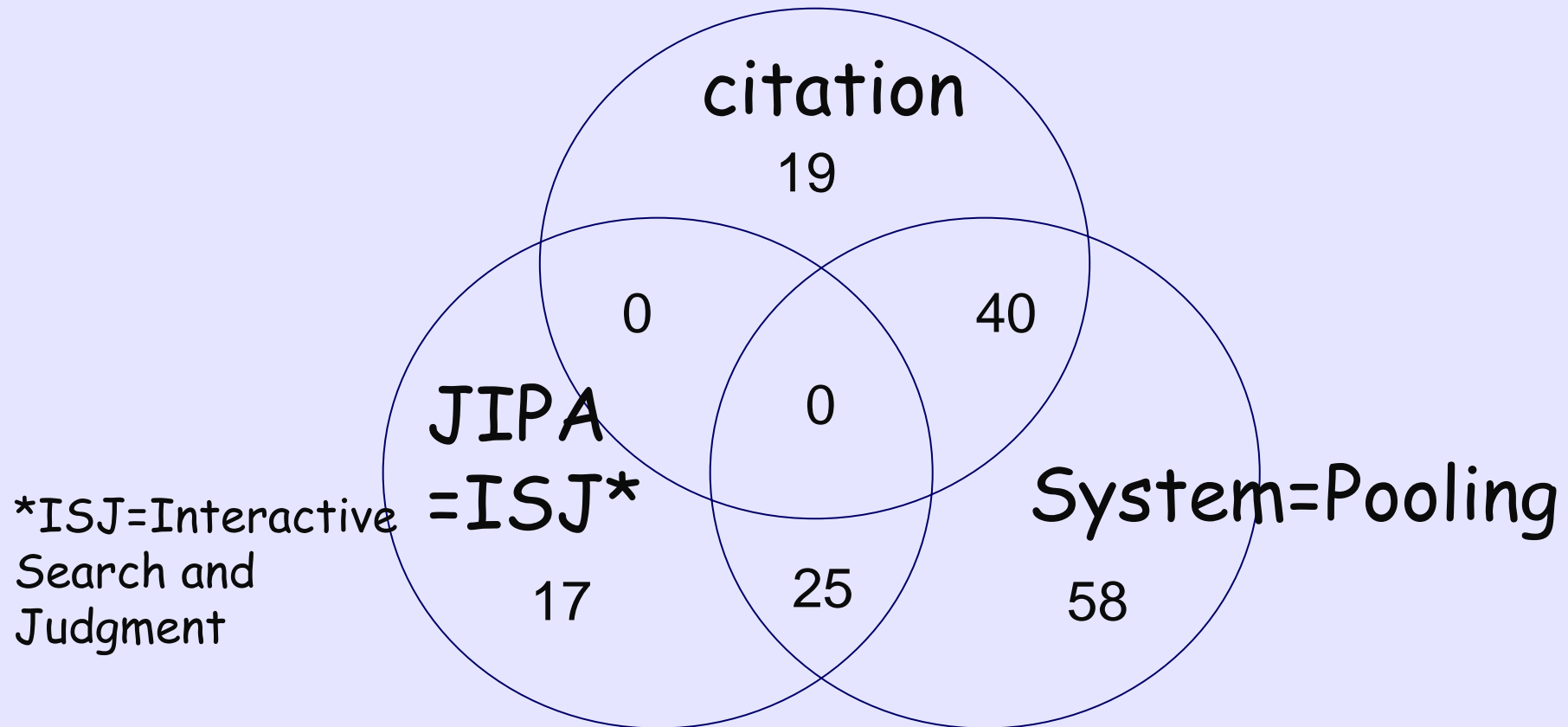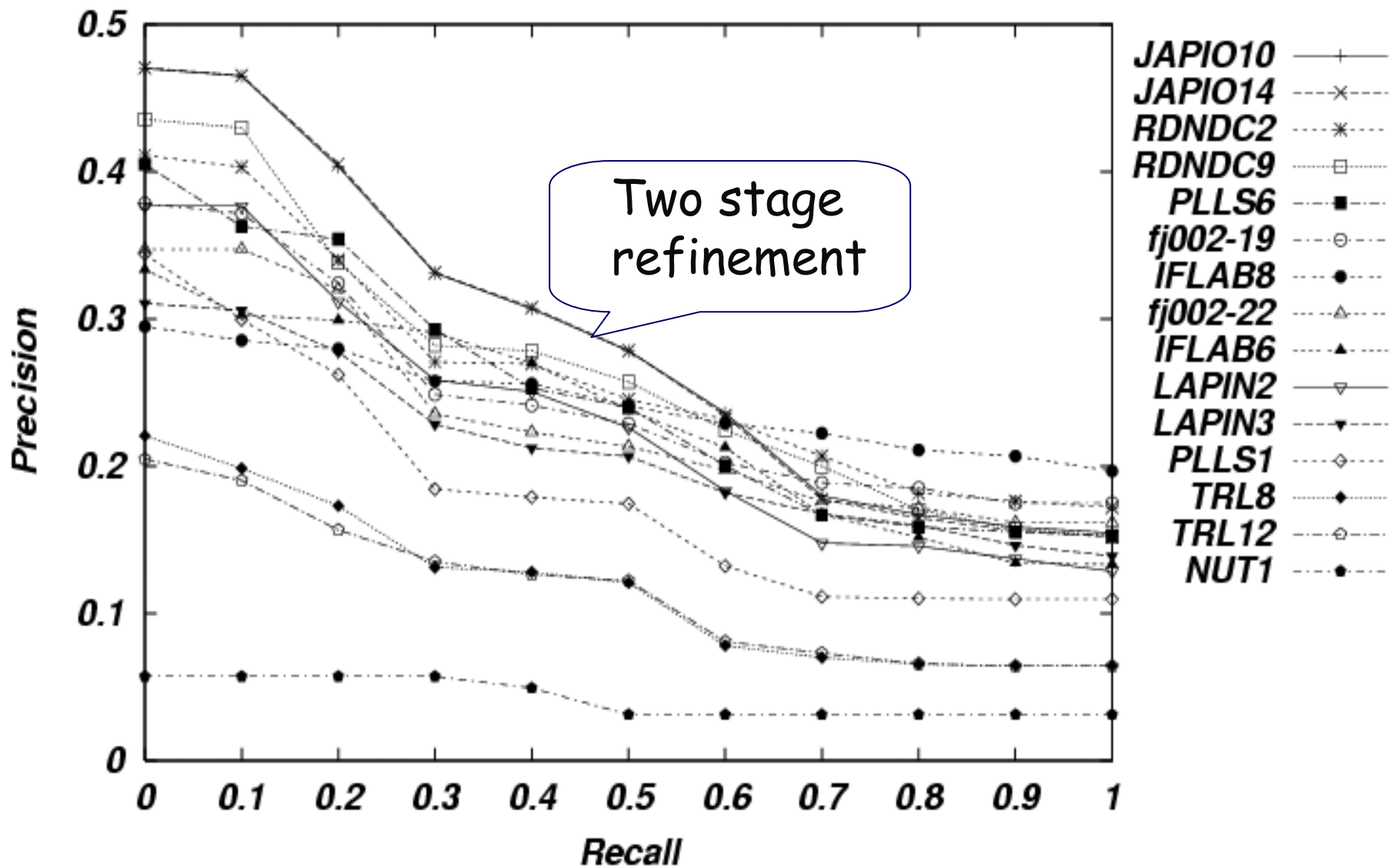
Target for invalidation

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)

Details of relevant documents
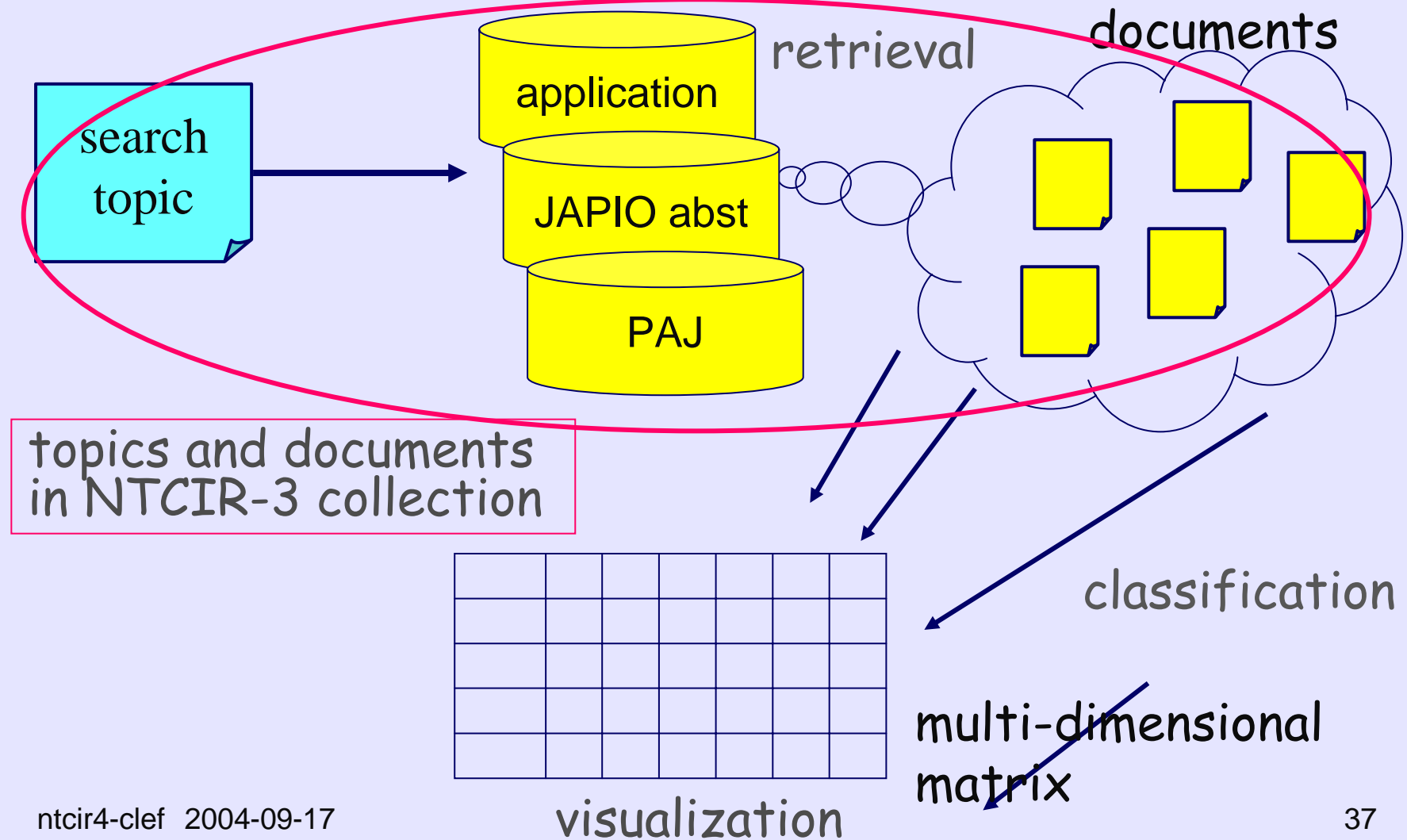## (A: rigid relevant)

citation

19

0   40

JIPA
=ISJ*

0

*ISJ=Interactive
Search and
Judgment

System=Pooling

17   25   58

total number of A-rel documents is 159

Main, Rigid

Two stage refinement

Feasibility Study: automatic patent map generation

search topic

application

JAPIO abst

PAJ

retrieval

documents

topics and documents in NTCIR-3 collection

classification

multi-dimensional matrix

visualization

Example (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 38

Question Answering Challenge

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)

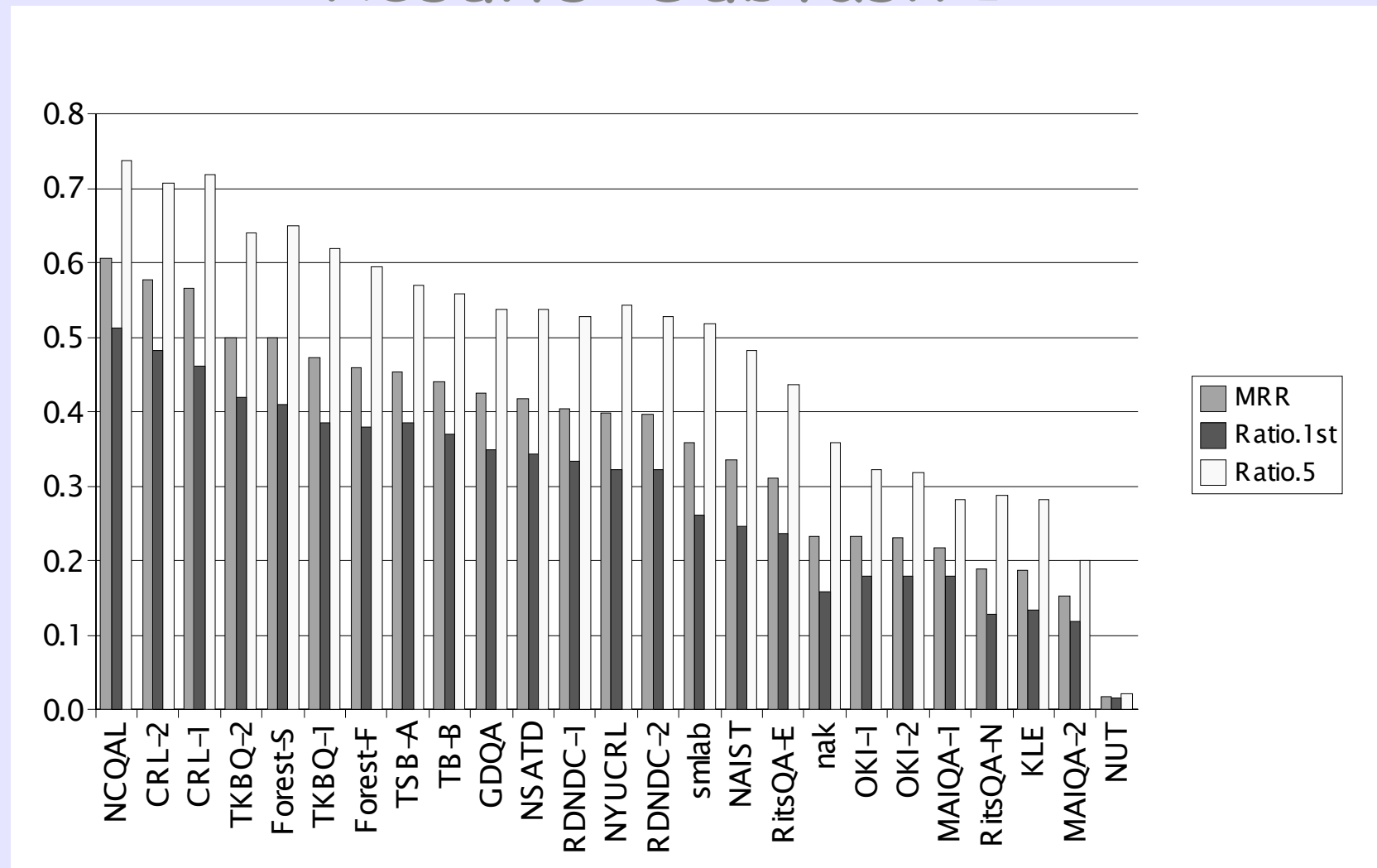Report writing task: topic centered vs browsing, Eval by F-measure

-Exact Answers   - Return in 48 hours

-Doc IDs are required as support information
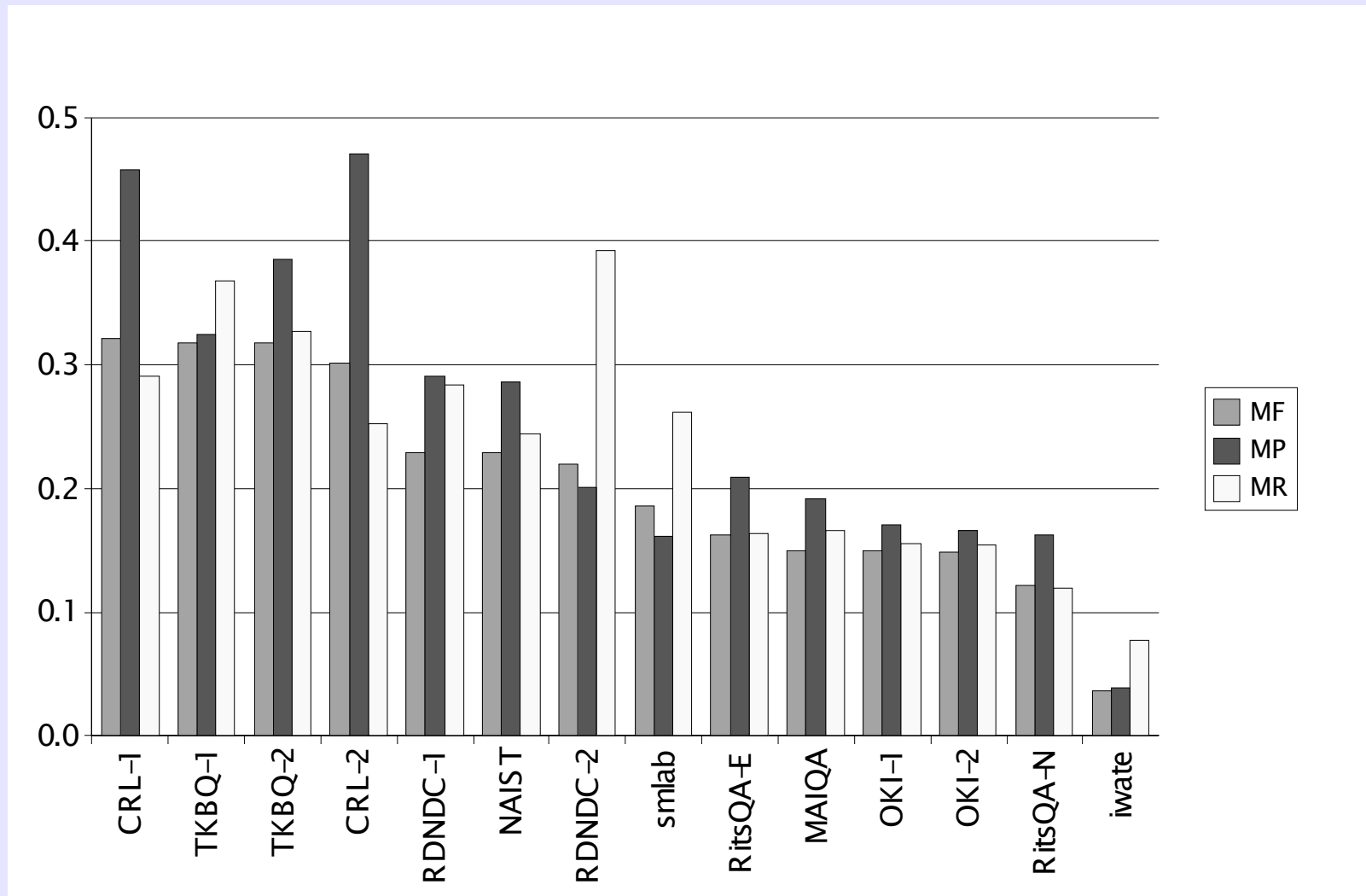
Task design was improved drastically from NTCIR-3

Results: Subtask 1

MRR of correct ratio of 1st ranked answer and among 5th ranked ones

Results: Subtask 2

Average F-measure, Precision, and Recall over all Qs

Subtask 3: Series of Question Situation Settings (User's Task)

1. **Collecting information about a particular topic**
   - One (hidden) global topic and series of Qs on subtopics of the global topic

2. **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

Relation to Multi-Doc Summarization

Answering a series of Qs has a **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
(hidden global Q= Seiji Ozawa)

- 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

Example of Series of Questions
## (Browsing type Q= topics shifting)

- 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

Evaluation by MMF
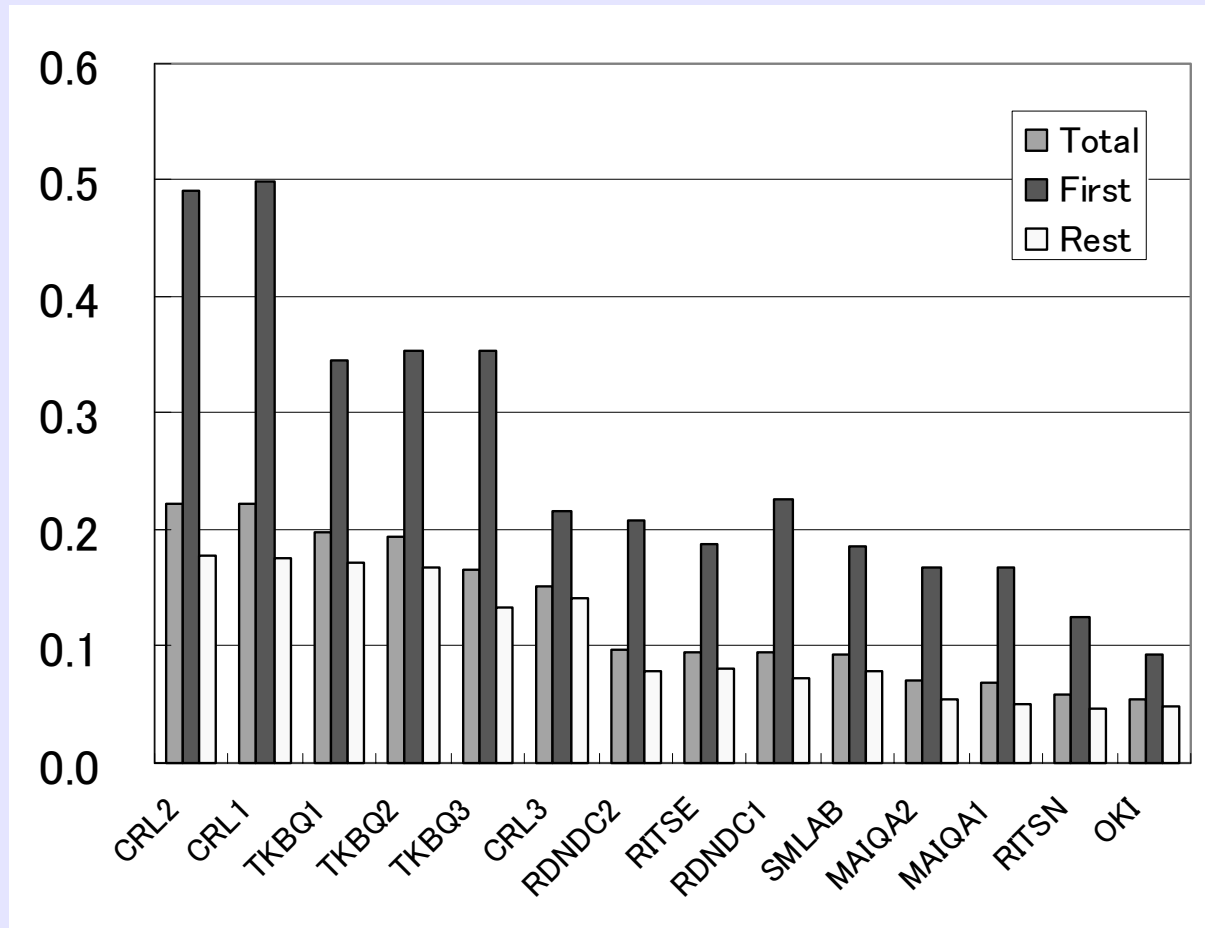
Differences on Series Type
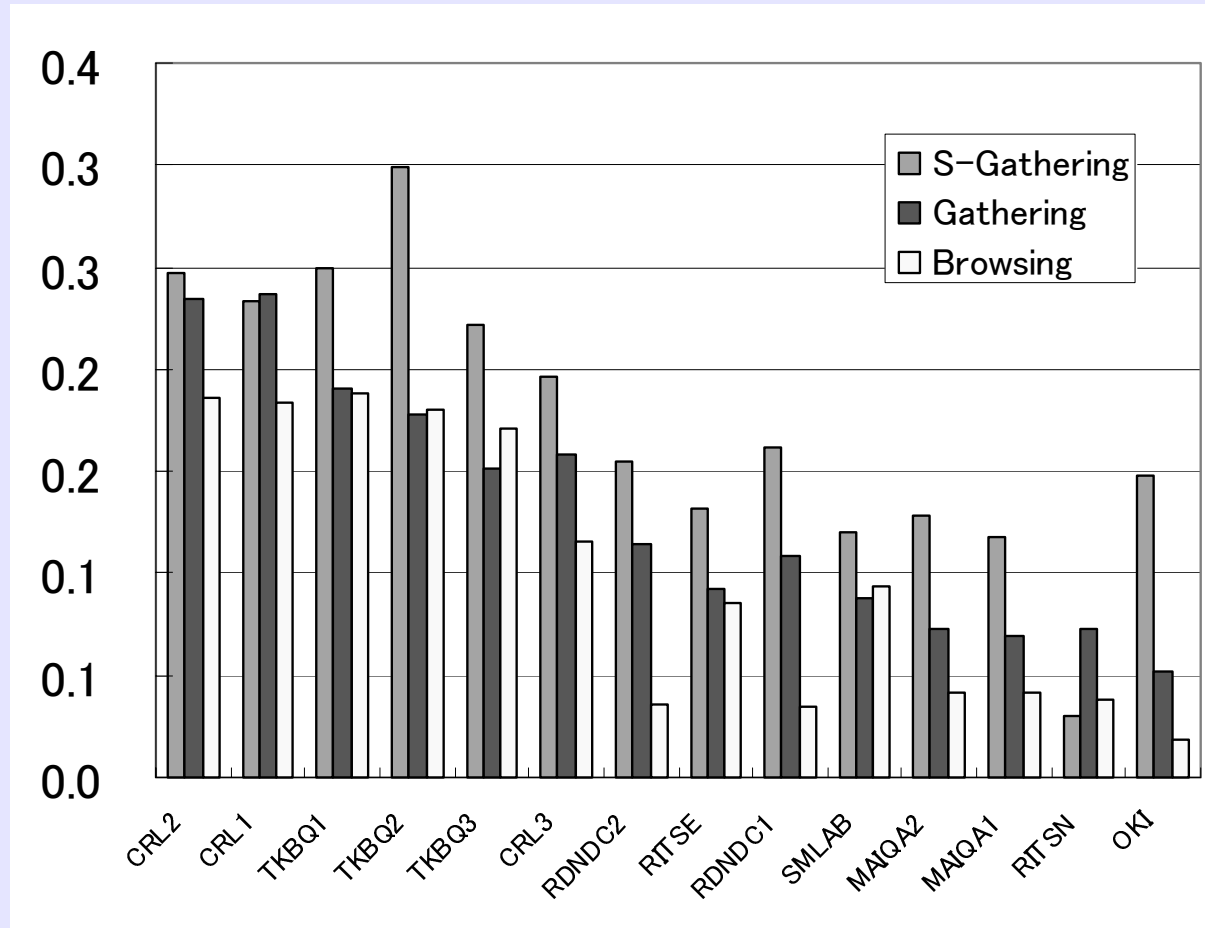
Problems on Evaluation

One set of all the answers == F-measure

- ## Multiple answers and context
  Ex.
    - Q1-Countries in East Asia? Ans-PRC, ROC, N Korea, S Korea, UK
    - Q2-Capitals of these countries? Ans- Beijing, Taipei, Pyongyang, Soul, Tokyo

- ## Expression diversity and identification of the same answers
  Ex. A and B are the same or not. # of total correct answers and recall value depends on such decision

- ## Major and minor answers

> Wrong answer

> Tokyo is not capital of UK. Correct answer for Q2 but this system produced wrong answer for Q1.

Text **S**ummarization **C**hallenge
- Two types of summarization -

- Extraction
  - Extracting important sentences from document sets
    length: # of sentences

- Abstraction
  - Producing summaries from document sets
    length: # of characters

Two lengths:
short, long

Automatic Extract Evaluation → Reusable Summarization Test Collection

See. Hirao (COLING 2004)

NTCIR-4 WEB

(A)Informational Retrieval Task

(B)Navigational Retrieval Task

[*Pilot*]**(C) Geographical Task**

[*Pilot*]**(D) Topical Classification Task**
retrieval result classification, eg.using clustering

Documents:

- 'NW100G-01' (100GB Web pages crawled in 2001 from "*.jp") for Subtasks A and B

- 'Target data' (subset of the NW100G-01) for Subtasks C and D.

Challenges in Information Access

Scaling-up

Beyond "Document" Retrieval

Answer/info in documents

Beyond the Heterogeneity

"Needs" Behind the Queries

Language, media, document genres, etc.

Appreciate each difference

User's situation, task, problem

Beyond "topic" and "fact"

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

NTCIR-5 (Mtg: Dec.6-9, 2005)

- **CLIR:** focus on NE, OOV, new docs 2000-2001.
- **Patent Retrieval:**
  - Invalidity Search, 10 year patent fulltext
  - Text Categorization to F-terms (good fine granularity for columns for a patent map axis)
- **QAC:**
  - Series of Questions (J-J)
- **WEB:**
  - Navigational Retrieval, New 0.5TB data

You are most welcome!

- **Pilot CLQA:** E-C, C-C, E-J

Application form is available on the WEB.

TSC & QA visual are held as different evaluations.

Contact Info & Online Proceedings

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

Inquiries: Noriko Kando at kando (at) nii. ac.jp

Online proceedings, application & other info: http://research.nii.ac.jp./ntcir/

NTCIR-4 Online working notes, slides, posters-:
http://research.nii.ac.jp/ntcir-ws4/NTCIR4-WN/

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

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