

# CLIR System Evaluation at NTICR Workshops

Noriko Kando

National Institute of Informatics (NII), Japan  
(formerly NACSIS)

*<http://www.nii.ac.jp/ntcir/>*

*CLEF 2*

*Sept.4, 2001*

# NTCIR Workshop is :

- NII-NACSIS Test Collection for Information Retrieval and Text Processing
- A series of evaluation workshops designed to enhance research in information retrieval and related areas such as text summarization by providing large-scale test collections and a forum of research groups.

# Brief History

Project start on 1997

NTCIR WS 1 (Nov., 1998~Sept., 1999)

Sept. 1999: IREX joined: Summarization, QA

Nov. 1999: Int'l collaboration: Asian languages  
CLIR

Apr. 2000: RCIR/NII: permanent host

NTCIR WS 2 (June, 2000~March, 2001)

NTCIR WS 3 (Oct., 2001~Oct., 2002)

# Tasks

## NTCIR WS 1

- (1) Ad Hoc Retrieval (J-JE)
- (2) CLIR (J-E)
- (3) Term Extraction & Role Analysis

## NTCIR WS 2

- (1) Chinese Text Retrieval (C-C, E-C)
- (2) Japanese-English IR (J-J, E-E, J-E, E-J, J-JE, E-JE)
- (3) Text Summarization
  - A.intrinsic: A1-extract, A2-abstract
  - B.extrinsic: IR task-based

# Participants (NTCI R WS2)

ATT Labs & Duke Univ., Communications Research Laboratory (Japan), Fuji Xerox, Fujitsu Laboratories, Fujitsu R&D Center, Hitachi Co., Hong Kong Polytechnic, IoS-Chinese Academy, Johns Hopkins Univ., JUSTSYSTEM, Kanagawa University, KAI ST/KORTERM, Matsushita Electric Industrial, Natl. TsinHua Univ., NEC, NII, NTT & NAI ST, OASIS (Aizu Univ.), Osaka Kyoiku Univ., Queen College-City University of New York, Ricoh Co. (2), Surugadai Univ., Trans EZ, Toyohashi Univ. of Technology (2), Univ. of California Berkeley, Univ. of Cambridge/Toshiba/Microsoft, Univ. of Electro-Communication (Japan) (2), Univ. of Lib. and Inf. Science (Japan), Univ. of Maryland, Univ. of Tokyo (2), Yokohama National Univ., Waseda Univ. **36 groups from 8 countries**

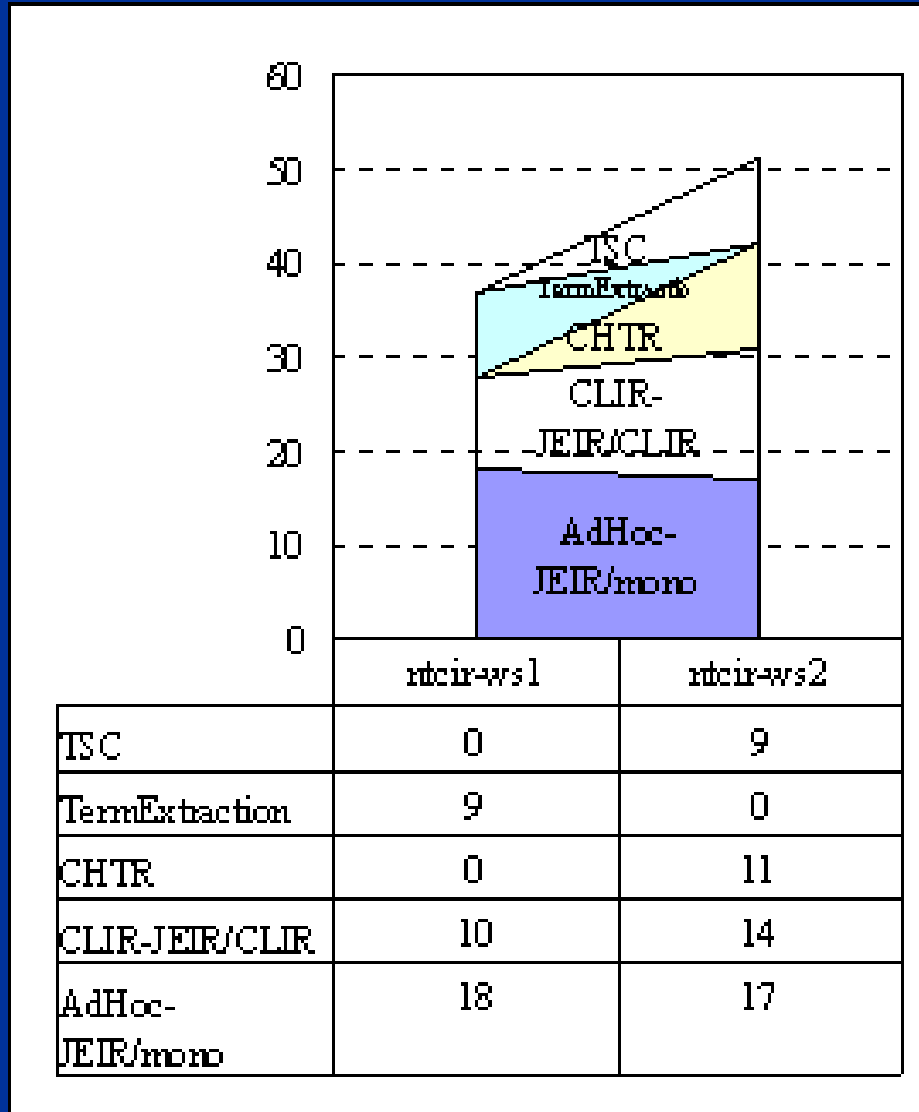
<b>Task</b>	Enrolled	Submitted	
<b>CHTR</b>	CHIR	14	10
	ECIR	13	7
	total	16	11
<b>JEIR</b>	J-J	22	17
	E-E	11	7
Mono	total	22	17
	J-E	16	12
	E-J	14	10
	J-JE	11	6
	E-JE	11	4
CLIR	total	18	14
	total	31	25
<b>TSC</b>	A extrinsic		7
	B intrinsic		5
	total	15	9
total		45	36

Number of participants of NTCIR Workshop 2 (by tasks)

# Number of participant (by attribute) (NTCIR WS 2)

	Univ.	Nat.Inst.	Company
CHTR	7	2	2
JEIR	15	3	7
TSC	3	1	5
total	20	4	12

# NTCIR Workshop 1 vs 2



New comers:  
 CHTR:7,  
 JEIR:12, and  
 TSC:4



# IR tasks at NTCIR WS 2

## Chinese Text Retrieval Task(CHTR) :

(a) Chinese monolingual IR (CHIR) C-C

(b) English-Chinese CLIR (ECIR) E-C

CIRB010: 5 newspapers published in Taiwan in 1998-1999, ca.137,000 C docs, 50 topics (C, E)

## Japanese-English IR Task(JEIR) :

(a) monolingual : J-J, E-E

(b) CLIR : J-E, E-J, J-JE, E-JE

NTCIR-1 &-2: scientific abstracts & extended summaries, ca.735,000 J docs + 418 E docs , 49 topics (J, E)

# Chinese Text Retrieval Task

CHIR;C-C

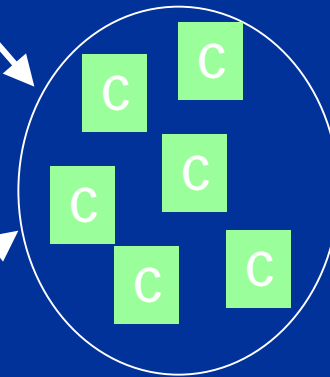
Topics

Chinese

Documents

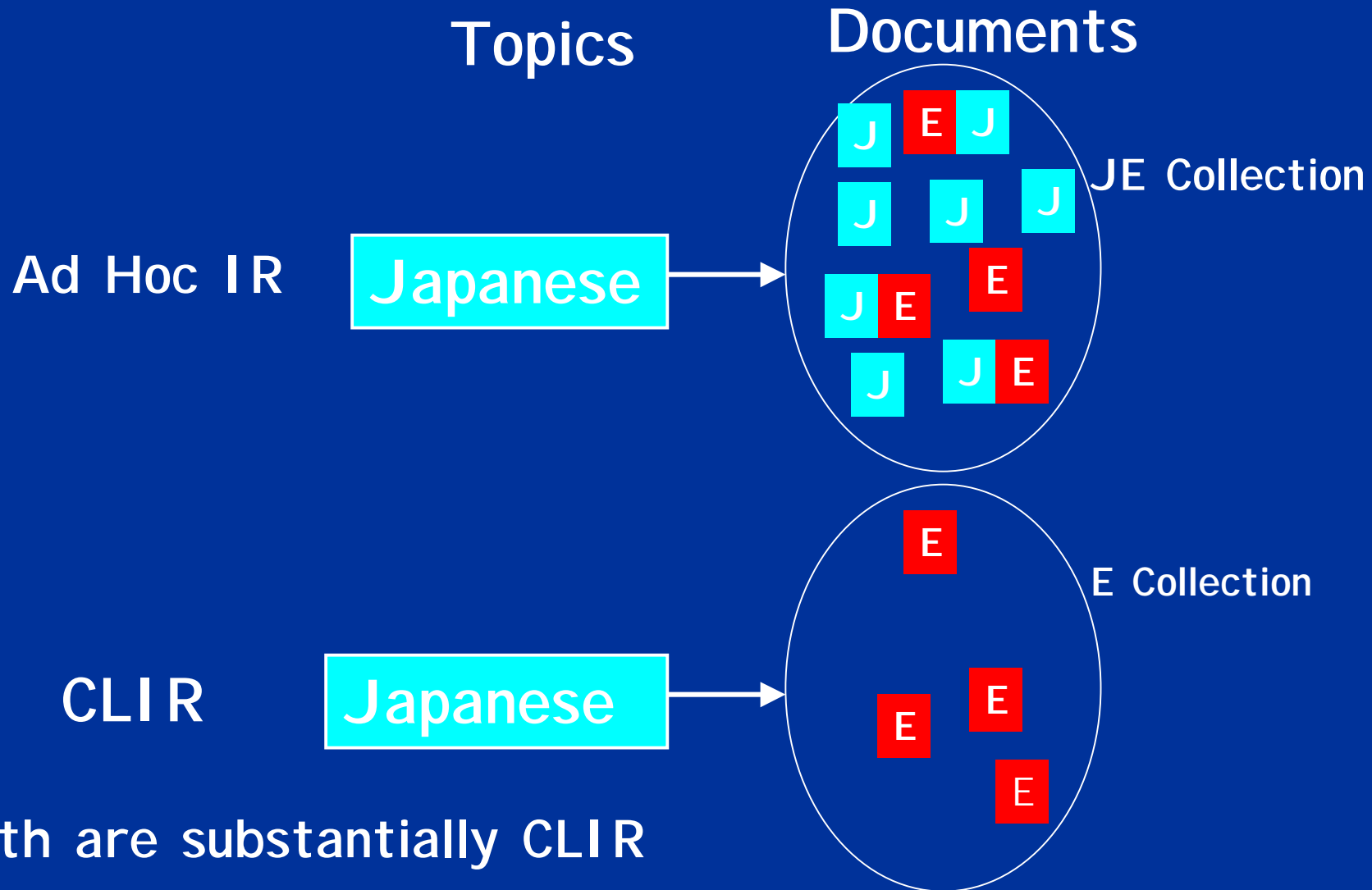
ECIR;E-C

English



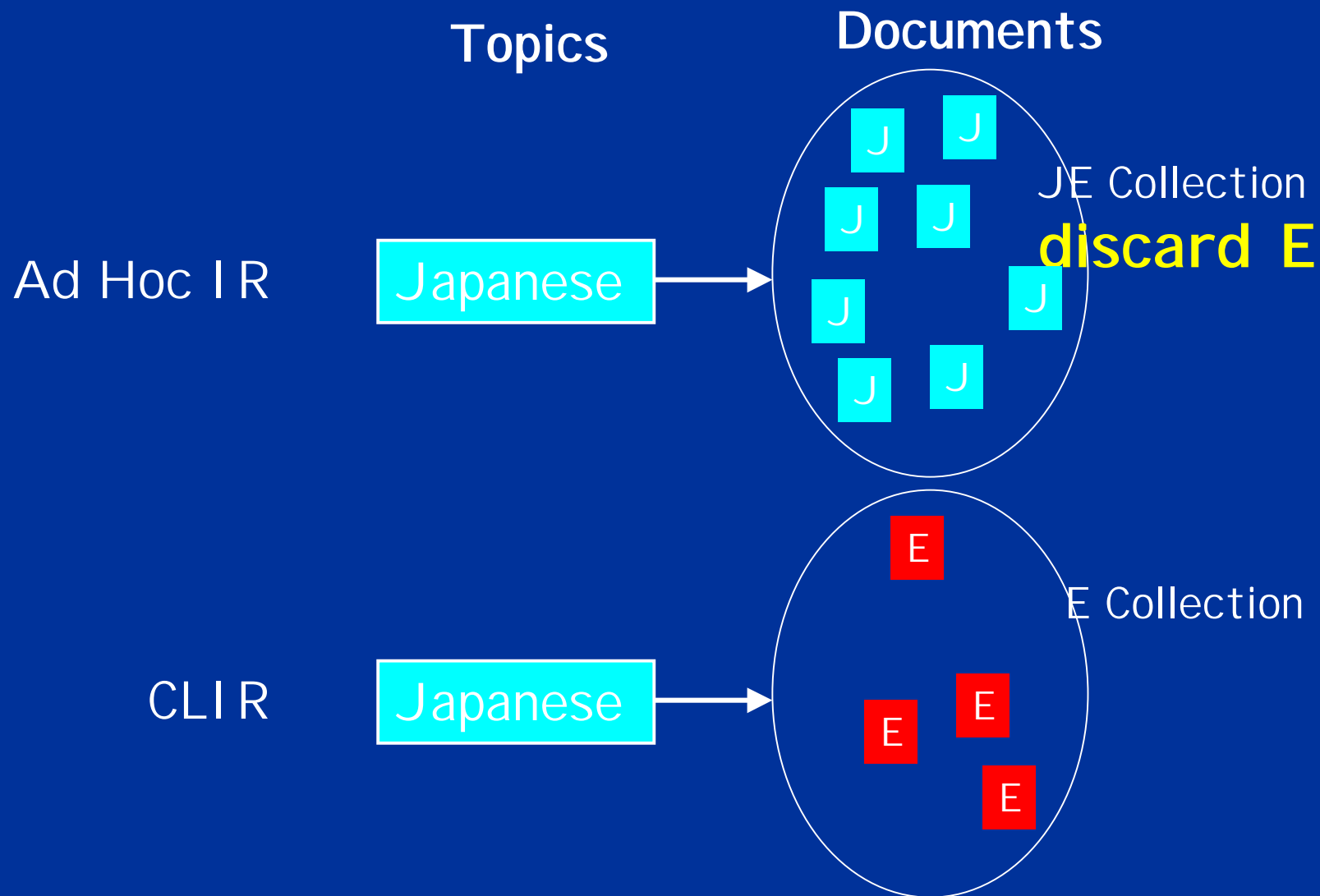
# Japanese-English IR at NTCIR WS1

-- planned



# Japanese-English IR at NTCIR WS1

-- actually done by several groups

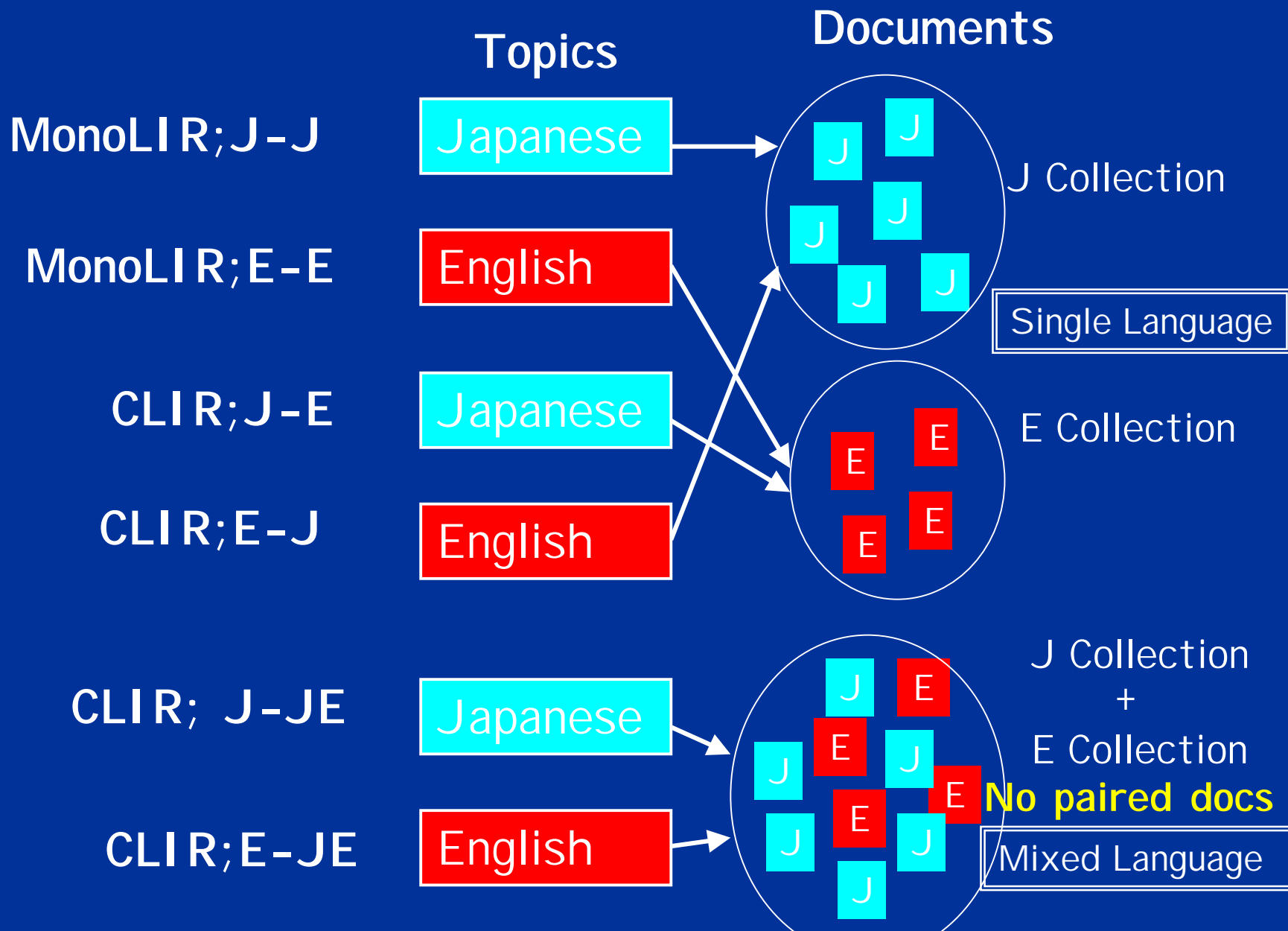


# JEIR

at

# NTCIR

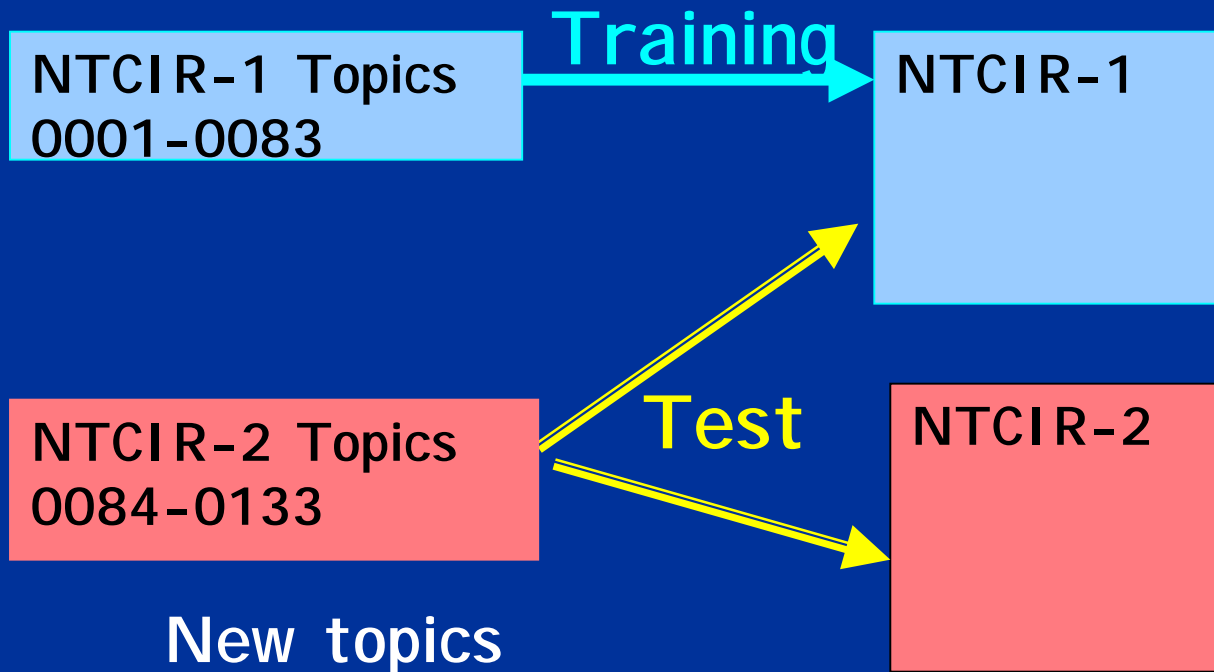
# WS2



# NTCIR Workshop 2

Queries

Documents



# Test Collections

		doc		topic	judgment	additional data
		#	size			
CIRB010	C	137K	200MB	C50 E50	4 grades	
NTCIR-1	J	332K	312MB	83	3 grades	tagged corpus
	E	187K	218MB			
NTCIR-2	J	403K	600MB	49	4 grades	segmented data
	E	135K	200MB	49		

NTCIR-1: Conference paper abstracts

NTCIR-2: Conference paper abstracts + extended summary of grant reports

# Sample Document (CIRB010)

<DOC>

<DOCNO>chinatimes\_focus\_0005660</DOCNO>

<LANG>CH</LANG>

<DATE>05071999</DATE>

<HEADLINE>解決高鐵融資 尋求第三管道</HEADLINE>

<TEXT>

<P>【記者羅兩莎台北報導】據負責台灣高速鐵路聯合貸款的主辦銀行表示，高鐵融資問題目前仍卡在銀行團、交通部高鐵路以及台灣高鐵路公司「三方合約」內容的訂定。在銀行團和交通部一直未能就相關歧見達成共識之下，三大主辦銀行原則決定，將尋求行政院經建會等第三管道與交通部協調，以儘早解決銀行團和交通部之間對融資問題的歧見。</P>

<P>高鐵路案將向國內銀行融資二千八百多億元，這項聯貸案確定由交銀、台銀和中國國際商業銀行共同主辦。不過，由於高鐵路是國內首宗BOT案，潛在風險究竟有多高，銀行無從評估。三大主辦銀行與交通部和台灣高鐵路公司訂定貸款合約時，重點亦著重在風險控制以及債權確保。</P>

<P>據主辦銀行主管表示，銀行當然希望債權確保不會有問題，譬如，在三方合約中訂定，由政府出面保證萬一將來台灣高鐵路公司蓋不下去時，政府可以出面買下，負責把工程完成等。

</P>

</TEXT>

</DOC>



# Sample Document (NTCIR-2: J Collection)

```
<REC>
<ACCN>kaken-j-0924516300</ACCN>
<YEAR>1992</YEAR>
<SBJ1 TYPE="kanji">802: 情報学</SBJ1>
<PJNM TYPE="kanji">文献の論理構造に基づく全文データベース検索システムの開発研究</PJNM>
<ABST TYPE="kanji"><ABST.P>本研究は、学術文献などの文書の全文を収容する全文データベースについて、それらの文書の論理構造に即した検索を可能とするシステムを研究・開発しようとするものである。3年次にわたって下記の項目について研究および開発を行なった。</ABST.P><ABST.P>1.全文データベースに対する検索要求の詳細分析を行ない、SGMLの文書型定義に基づいて検索・表示要求を効率的に記述するための表記形式DQL(Document Query Language)の詳細設計を行なった。SQLを拡張し、文書構造を扱うための記述を可能にした。</ABST.P><ABST.P>2.文献の文書構造を図形的に表示し、要素をポインティングデバイスで指定して検索条件・表示指示を行なうユーザ系のソフトウェアを設計し、ワークステーション上でグラフィカルユーザインタフェース(GUI)を用いて開発した。</ABST.P><ABST.P>3.文書構造を各構成要素間の二項関係で関係データベース管理システム上に表現し、DQLで記述された検索要求を処理するサーバ系のソフトウェアを汎用大型計算機上に開発した。</ABST.P><ABST.P>4.サーバ系とユーザ系の接続方式を開発し、LANおよびISDNを介して連動させて動作を確認した。</ABST.P><ABST.P>5.全体的な処理性能、使い心地、検索精度などについて評価を行ない、実用に向けての課題と解決方式を検討した。</ABST.P><ABST.P>現在のシステムにはサーバ系の性能に改善の必要性が認められ、検討の結果、二項関係に参照先のレコードIDを含めることが有効であることがわかったので、今後これを実現することが課題となる。また、ユーザ系においては指定した検索要求をよりわかりやすく表示する必要があることが明かとなり、考案したいくつかの方式について実験により検討することが課題となる。</ABST.P><ABST.P>以上の結果、本システムの設計概念、および実現方式の妥当性が確認でき、課題への対処の方針も示すことができたことにより、実用化の可能性が示された。</ABST.P></ABST>
<KYWD TYPE="kanji">全文データベース / 情報検索 / 文書構造 / SGML / GUI / 分散処理</KYWD>
</REC>
```

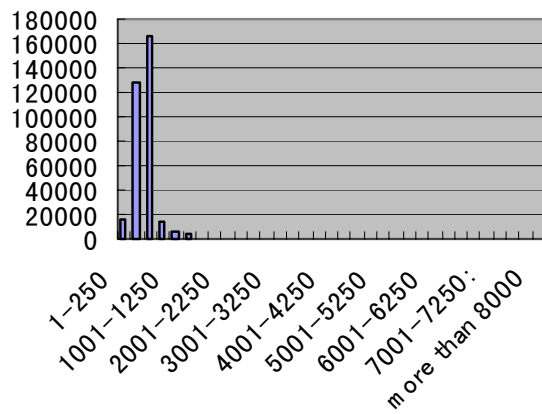
# Sample Document (NTCIR-2:E Collection)

```
<REC>
<ACCN>kaken-e-2469487463</ACCN>
<YEAR>1992</YEAR>
<SBE1 TYPE="alpha">802: *</SBE1>
<PJNE TYPE="alpha">Development of a Full Text Retrieval System based on Logical Structure of Documents</PJNE>
<ABSE TYPE="alpha"><ABSE.P>The investigators have conducted research and development of a system with which users can retrieve documents from a full text database containing full documents such as academic papers. Results obtained were as follows: </ABSE.P><ABSE.P>1. Based on detailed analysis of retrieval requests for full text databases, they made detailed design of a formal notation DQL (Document Query Language) for describing retrieval and display requests efficiently according to the document type definition of SGML. </ABSE.P><ABSE.P>2. They designed a user system software which displays document structures and let users select elements with pointing devices and specify retrieval conditions and display instructions. The system was developed on work-stations using graphical user interface systems. </ABSE.P><ABSE.P>3. They developed a server system software on a main frame computer which stores documents in a relational database management system by expressing the structure as binomial relations among elements and processes retrieval requests described in DQL. </ABSE.P><ABSE.P>4. They developed a communication method between sever and user systems and confirmed the functionality by experiments using LAN and ISDN. </ABSE.P><ABSE.P>5. They evaluated the overall performance, usability, retrieval precision and so on and considered on the problems and their solutions toward the practical use.</ABSE.P><ABSE.P>Through evaluation, the necessity of performance improvement of the server system was revealed. Further investigation has made it clear that including referenced record identifiers within the binomial relation records is effective. It is the future issue to implement this method. The necessity for the user system to represent specified retrieval requests more understandably was also revealed. Several methods already proposed should be studied by experiments.</ABSE.P><ABSE.P>As the result of the research, feasibility of the design concept and the implementation method of this system was confirmed, the approach to the existing problems was presented, and the reality of the full-fledged system was shown.</ABSE.P></ABSE>
<KYWE TYPE="alpha">Full Text Database / Information Retrieval / Document Structure / SGML / GUI / Distributed Processing</KYWE>
</REC>
```

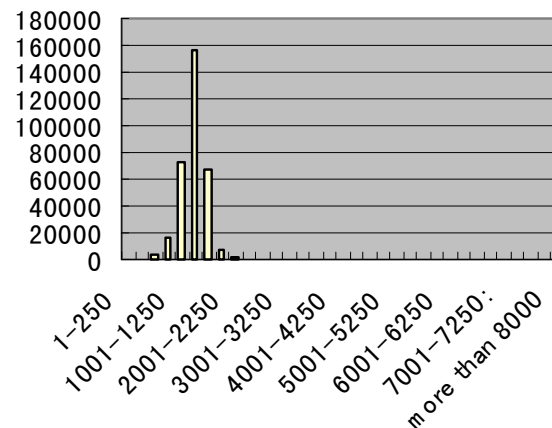
# Length of free text part of subfiles (bytes)

	min	max	average	stdev
NTCI R-1 Conference Paper Abstracts				
J Collection	0	8,003	521.7	224.0
E Collection	1	4,363	531.9	254.9
NTCI R-2 Grant-in-aid Report Subfile				
J Collection	5	17,111	1,366.8	291.4
E Collection	15	20,396	1,614.6	492.7
Newspaper articles (Japanese only)				
I REX	20	15,770	895.9	977.8

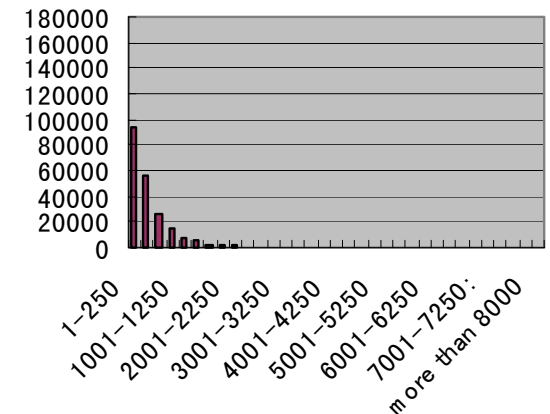
GAKKA\_IJ



KAKEN\_J  
Grand-in-Aid report



IREX



# Topics: mandatory run=D only

<TOPIC q=0005>

<TITLE>特徴次元リダクション</TITLE>

<DESCRIPTION>クラスタリングにおける特徴次元リダクション  
</DESCRIPTION>

<NARRATIVE>オブジェクトのクラスタリングを行なうとき、オブジェクトを特徴ベクトルで表現することが望まれる。アプリケーションによっては、オブジェクトの次元は数千、数万となることがある。このような場合、事前に次元を落とすことが必要になる。正解文書は、特徴次元リダクション方法について、理論面から、または実験によって、提案、比較などを行なっているもの。画像処理などの実験の操作の一部として特徴次元リダクションを用いているだけでは要求を満たさない。</NARRATIVE>

<CONCEPT>特徴選択, 主成分分析, 情報の粒度, 幾何クラスタリング</CONCEPT>

<FIELD>1.電子・情報・制御</FIELD>

</TOPIC>

# Topic Preparation: Guideline (for JE collection)

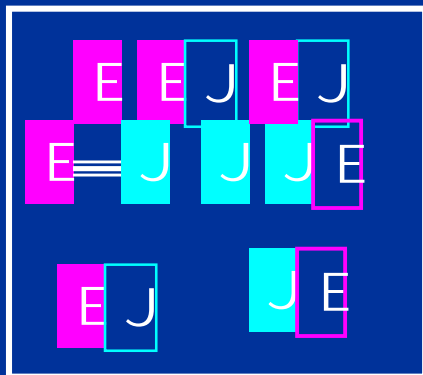
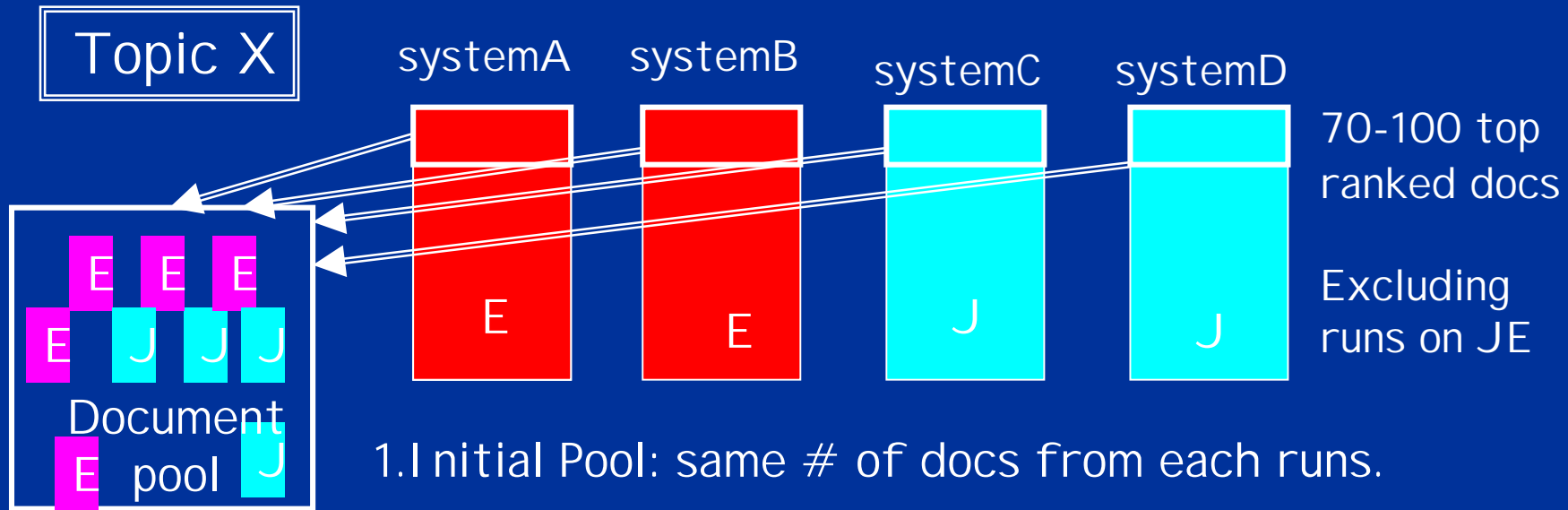
- Statements of “user need” rather than “queries”
- **Not too easy : polysemy + synonymy**
  - Simple word matching of <DESCRIPTION> terms can not retrieve every relevant docs.
  - A document containing <DESCRIPTION> terms can be non-relevant. Because in the real world documents, a concept can be represented by different terms and a term can represent different concepts.
- **5+ relevant documents at initial search**
- **Estimate less than 333 relevant docs/topic** (it is to draw R-P graph appropriately using 1000 documents submitted.)

# Relevance Judgments

- Exhaustive list of the relevant documents for each topic
- Pooling
- 4 grades:
  - Highly Relevant (S)
  - Relevant(A)
  - Partial Relevant(B)
  - Non-relevant(C)
- 2 analysts: Cross-checked
- Contains extracted phrases/passages showing the reason that the analyst judged it as “relevant”

# Cross-lingual Pooling

Retrieval results for topic X (ranked lists)



2. Find the pairs and add them, if not included in the initial pool.

3. Relevance judgments by human assessors

4. Additional manual search for 16 topics

rel docs > 100 or initial pool=top 70 docs

# Pool size

average, minimum and maximum number of documents per topic, and the total size.

topic	1st pool+presearch			additional manual search		final judgments	
	J	E	total	J	E	J	E
total	61881	50369	103234	4513	1920	106626	66729
ave	1262.9	1027.9	2106.8	282.1	120.0	2176.0	1361.8
min	884	689	1455	3	0	1449	963
max	1702	1483	2824	753	497	3331	2144



# Number of relevant docs

average, minimum and maximum number of relevant documents per topic, and the total

topic	initial judgments				additional manual search*				final judgments			
	J		E		J		E		J		E	
	S+A	S+A+ B	S+A	S+A+ B	S+A	S+A+ B	S+A	S+A+ B	S+A	S+A+B	S+A	S+A+ B
total	3037	4760	1401	2034	222	362	1	111	3259	5122	1402	2145
ave	62.0	97.1	28.6	41.5	13.9	22.6	0.1	6.9	66.5	104.5	28.6	43.8
min	6	12	3	5	0	0	0	0	6	12	3	5
max	201	302	170	197	84	101	1	45	203	367	170	242

\*16 topics

# Additional Data -- Segmented Data

<PJNM>誤り の 多い 文字列 から の 原  
文\_復元 に 関する 研究</PJNM>

<ABST>特定\_分野 の 文献 に おける 単語  
\_出現\_状況 の 調査\_結果 から、文章\_中  
で 一意 に 定まり、かつ 高頻度 で  
出現 する 特定 で ...

**Hard Segmentation:** Space, longer units  
(~compound terms)

**Soft Segmentation :** \_ Under score,  
components in a term

# Evaluation

## trec\_eval

Mean Average Precision

R-P Graph (11pt Recall-Level Precision)

Doc Level Precision

R-Precision

By topic: difference from MEAN

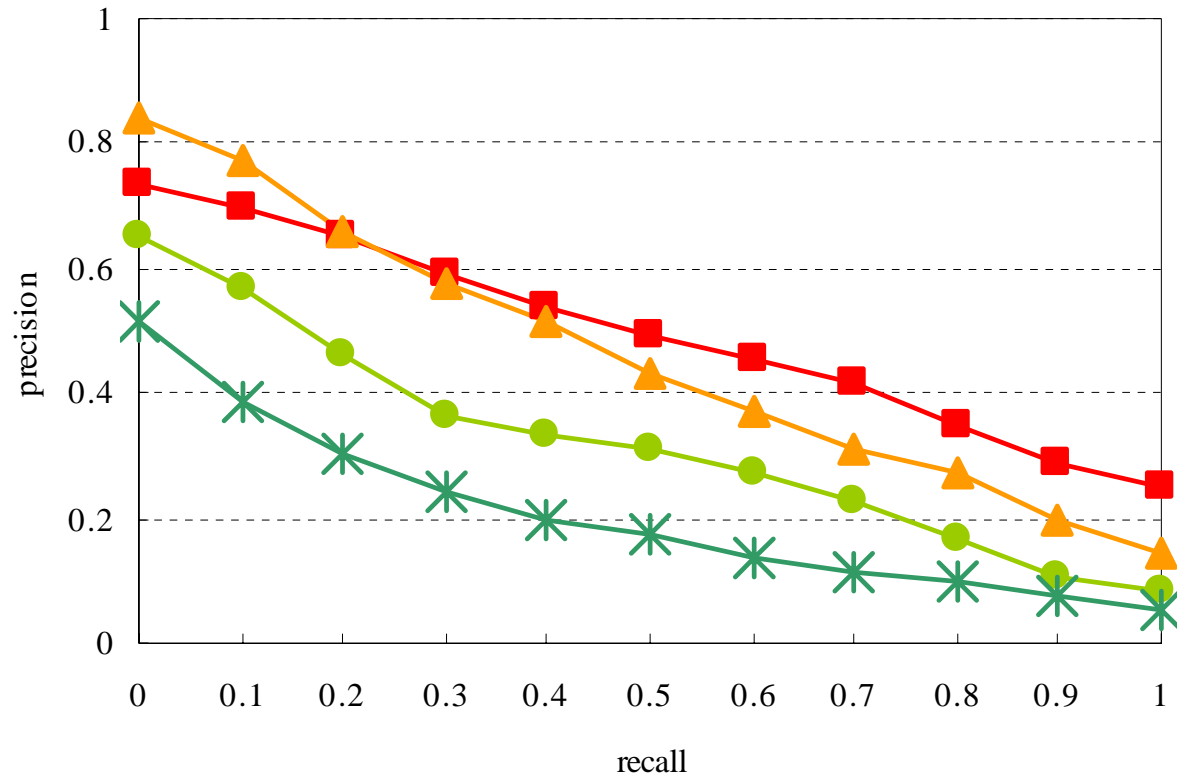
## Multigrade judgments 1: Two threshold

Level 1 (Rigid Relevant) 3+2

Level 2 (Relaxed Relevant) 3+2+1

## Multigrade judgments 2: Weighted Average Precision, Weighted R Precision, single measure, averageable

## ECIR (E>C) all Rigid Relevance



PIRCS-ECIR-SO

sst ut-ECIR-LO-01

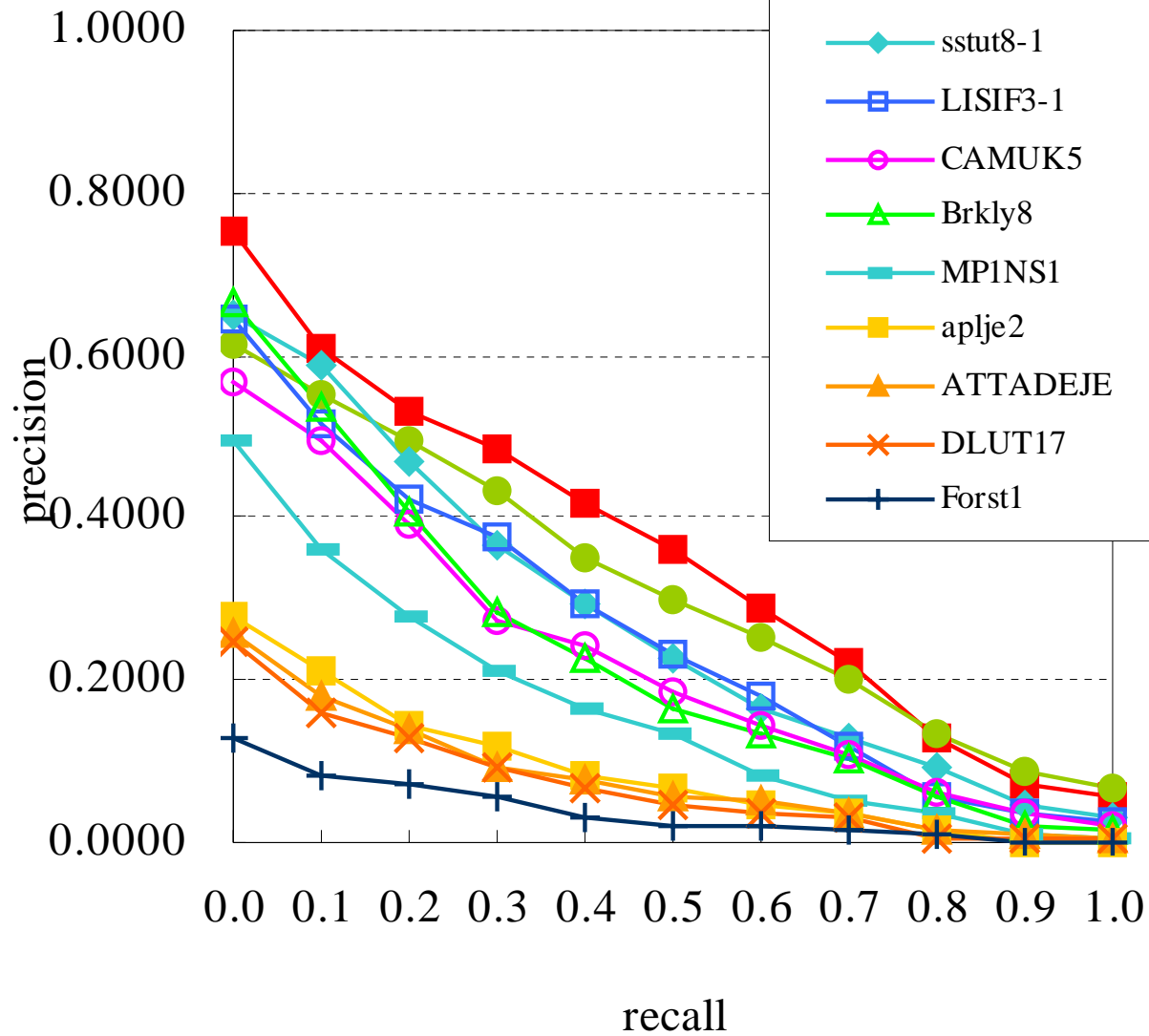
UMD-ECIR-LO-03

BRKLY-ECIR-LO

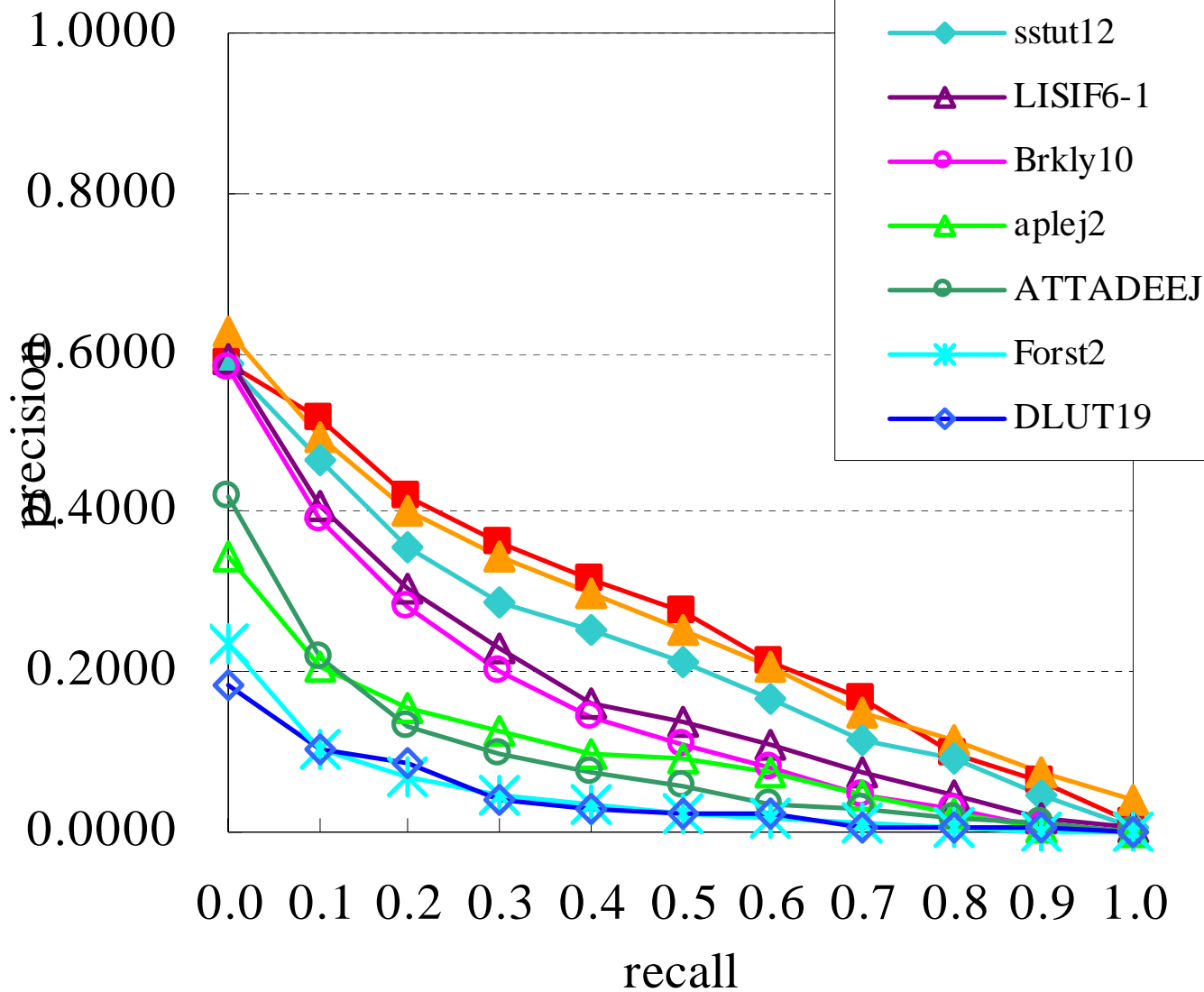
# ECIR (E-C CLIR)

- 17 results submitted from 7 groups
- QE is effective
- Probabilistic model showed better performance
- MRD: Select-all better than select-X, if no further approaches were adapted
- MT: PI RCS used MT and out-performed

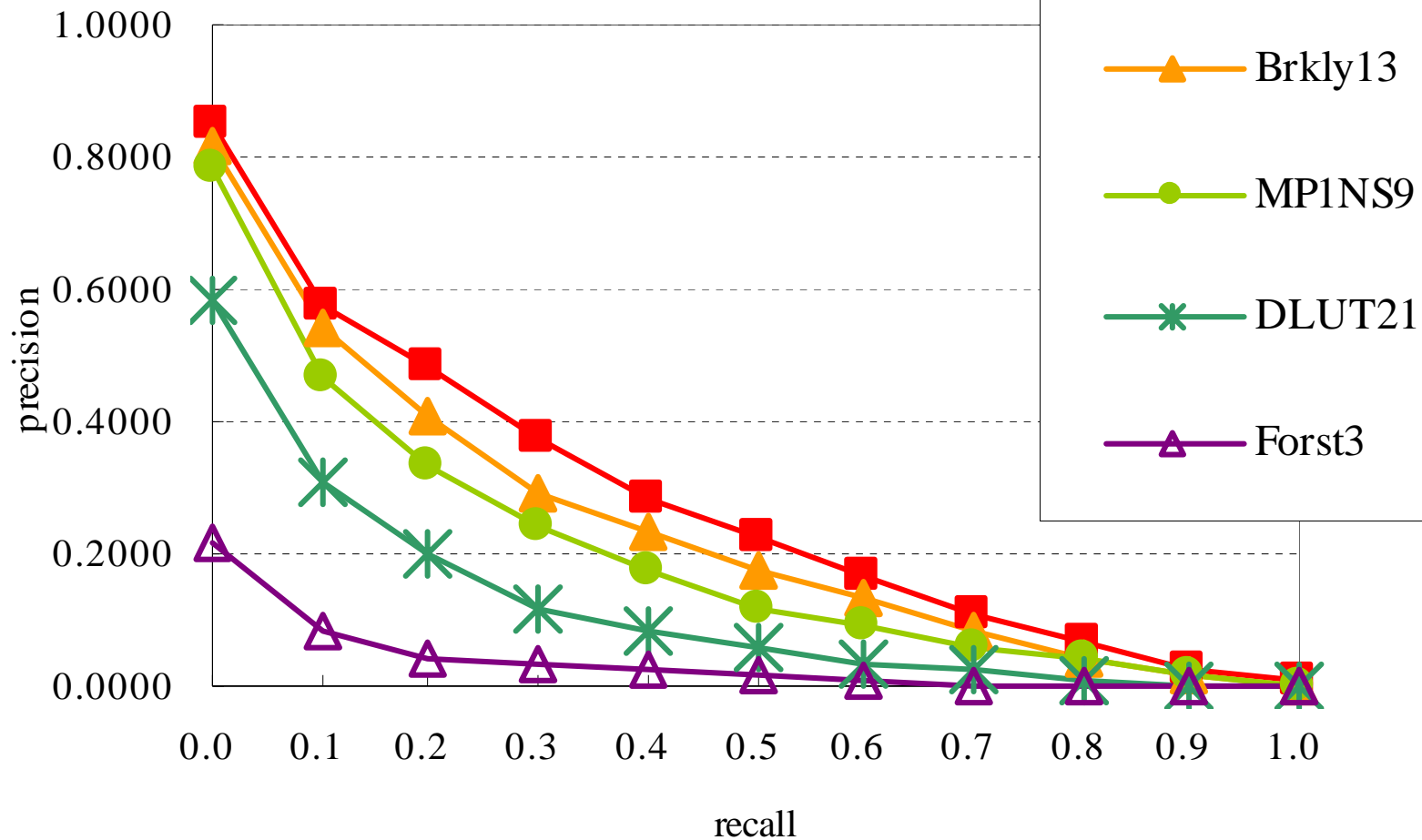
### J-ED auto Level 1



# E-J D auto Level 1



J-JE D auto Level 1





# JE CLIR

## Number of Submitted runs

	Runs	Groups
J-E	40	14
E-J	30	10
J-JE	14	6
E-JE	11	4
Total	95	14

# JE CLIR

- Group from NTCIR WS 1 performed better than new comers
- **Various models:** Vector Space, LSI, probabilistic
  - Segmented LSI for CLIR
  - Flexible Pseudo-relevance feedback, etc. etc.
- Most groups: Query translation
- LIFIS: Combination of query & document translation
- Corpus based (LSI, segmented LSI, approximate dimension equalization, etc.) generally less effective than MRD, MT; better in x-JE
- QE: pretranslation -1, post transl - 4, both- 1
- **Pseudo RF:** worked well (LCA, FPRF)

# Aanalysis of Test Collections/results

- (A) exhaustivity of the document pool; effect of unique contribution
- (B) inter-analyst consistency and its effect for system evaluation
- (C) topic-by-topic evaluation
- (D) estimate difficulty of topics
- (E) effect of segmentation for search effectiveness
- (F) evaluation measures: nature of MAP, correlation of trec\_eval measures with multigrade judgments; significance test as paired samples, measures for multigrade judgments
- (G) effect of reuse of training data for test (under analysis)

# Future Directions

- Languages;
  - Asian languages
  - CLIR & international collaboration
  - Resource sharing
- More realistic evaluation
  - new document genres + user group ; Web, patent, multimedia etc.
  - Interactive system
- Evaluation of technology to make information in the documents
  - intersection of IR and NLP

# NTCIR Workshop 3 (2001/2)

Once per 1 ½ years:

Meeting: Oct.8-10, 2002

## TASKS:

- CLIR – East Asian Languages: CKEJ
- Patent Retrieval
- QA
- Text Summarization
- Web Retrieval

# NTCIR Workshop 3 (2001/2)- CLIR

- Multilingual CLIR (MLIR):  
X98->CJE
- Bilingual CLIR (BLIR):  
X98->C or J, X94->K
- Single Language IR (SLIR):  
C98->C, J98->J, K94->K

Docs: newspaper published in Asia

C: 380,000 docs, 1998-1999 (TaiwanROC)

J: 240,000 docs, 1998-1999

K: 66,000 docs, 1994 (? more 1992-1994)

E: 22,000 docs, 1998-1999 (Taiwan,Japan)

Topics: CKEJ98-50, CKEJ94-30

# NTCIR Workshop 3 - Patent

- **Cross-Language Cross-DB Retrieval:** search J patents by J/E/C newspaper article (30 topics)
  - Submit a list of top 1000 docs
  - Most relevant passage (optional)
  - Topic creation & judgments by real users
- **Optional task:** any research are invited. ex: patent map generation, paraphrasing claims, aligning claims and examples, summarization, clustering patents.

# NTCIR Workshop 3 - Patent cont'd

## Collections

- Japanese patents: 1998-1999 (about 17GB)
- JAPI O patent abstracts, J, : 1995-1999
- PAJ: Patent Abstracts of Japan (English translations): 1995-1999, 1,750,000 docs
  - Above two are exactly translated E-J abstracts: 1995-1997 are usable for associative dictionary in CL-CDIR task
- PATOLIS test collection (34 search topics for Japanese patents and relevance assessment)
- Newspaper articles (J/E /C) (30+)



# NTCIR Workshop 3 - QA

**Task 1:** Return 5 top answers (noun phrase) /w support information (relevant passage less than 100 characters), 100 Q

**Task 2:** Return only one answer. 100 Q.  
Support information is required.

**Task 3:** a series of questions. The related questions are given for the 30 of questions of Task 2.

**Doc:** J newspaper articles, 1998-1999

# NTCIR WS 3 - Text Summarization

**Task A (single document summarization):**

Given the texts to be summarized and summarization lengths.

**Task B (multi-document summarization):**

Given a set of texts. The information which was used to produce the document set, such as queries, as well as summarization lengths are given to the participants.

**Doc:** J newspaper articles, 1998-1999

# NTCIR Workshop 3 - Web

## A. Survey Retrieval (recall and precision)

A1. Topic Retrieval

A2. Similarity Retrieval

## B. Target Retrieval (precision-oriented)

## C. Optional Task

C1. Search Results Classification

C2. Speech-Driven Retrieval

C3. other

**TS:** Web docs collected from .jp domain + well linked docs. Only J+E are used. 50 J topics. Available at Open-Lab at NII.

# NTCIR Workshop 3 - New Features

Multilingual CLIR: First CLEF model in Asia

Optional tasks (patent,web): research projects using the document collections are invited. Expect this will explore the future tasks.

Search by Documents (patent, web)

Passage Retrieval (patent, QA, web)

Multi-doc summarization

Precision-oriented evaluation (QA, web) &  
Multigrade judgments (CLIR, patent, web)

NTCIR Home:

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

- CFP
- Join MLs
- Registration for Participation
- Collections
- Online Proceedings & Publications

Thanks

Merci

Danke schön

Gracie

Gracias

Tack

Dank je wel

Kiitos

Terima Kasih

Khap Khun

Ahsante

Tak

謝謝

ありがとう