

Overview of the Sixth NTCIR Workshop

Noriko Kando

National Institute of Informatics

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

kando (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

6th: April 2006 - June 2007

* **Nii Test Collection for Information Retrieval systems**
ntcir6 2006-05-16 Noriko kando

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 NTCIRs

Tasks (Research Areas) of NTCIR Workshops

		1st	2nd	3rd	4th	5th	6th	
T a s k s	Japanese IR	→						
	Cross-lingual IR	sci →		news →				
	Patent Retrieval map/classif			→	→			
	Web Retrieval Navigational Geo Result Classification			→	→			
	Term Extraction	→						
	Question Answering Info Access Dialog Summ metrics Cross-Lingual			→	→		→	
	Text Summarization		→	→				
	Trend Information					→	→	
	Opinion Analysis						→	

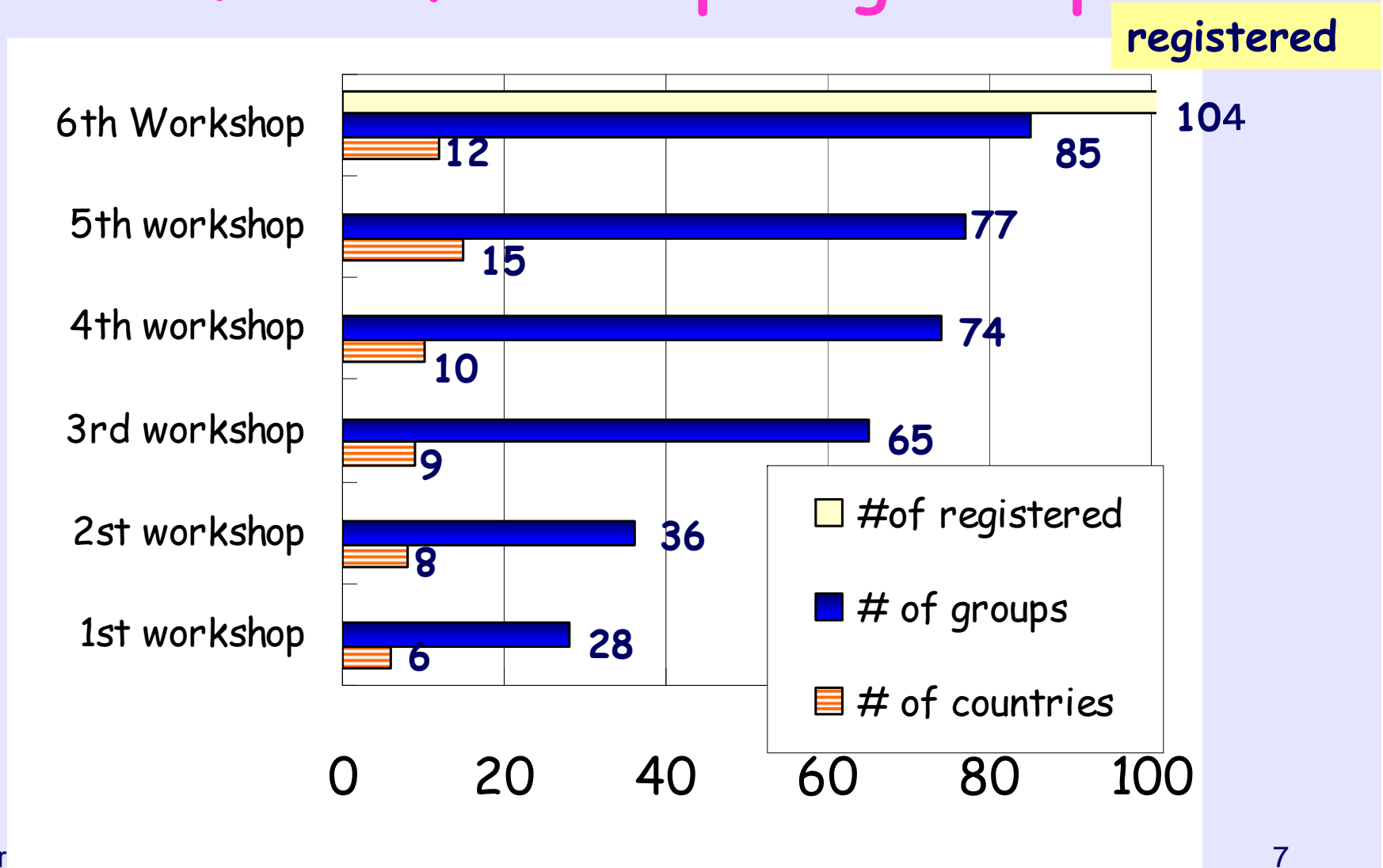
NTCIR-6 (Mtg: May 15-18, 2007)

- **CLIR**: multi-collection. NTC3-5, news docs, CJK
- **CLQA**: E-C, C-C, C-E, E-J, J-J. J-E (factoid)
- **Opinion**: CJE, reuse NTC3-5 CLIR
- **Patent Retrieval**:
 - Invalidity Search, 10 yr patent fulltext ca90GB
 - Text Categorization to F-terms (good granularity for patent map axis)
- **QAC**: Every kind of Qs (J-J), eval by BE
- **[Pilot] Must**: MULTimodal Summarization for Trend information, extract numeric information from a set of documents, and visualize them to show their trends

NTCIR-6 Schedule

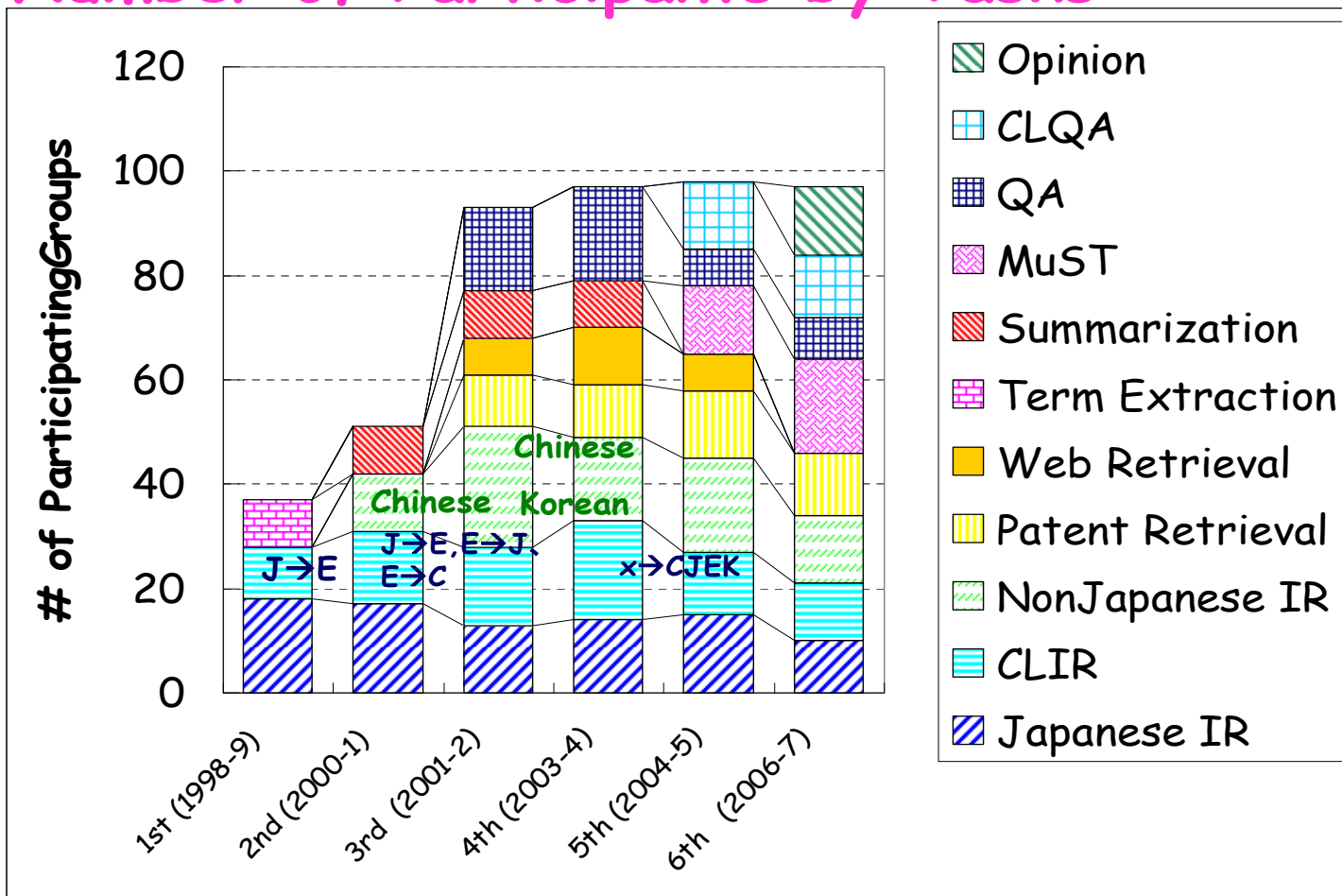
Task	Lang	Formal Run	Meeting
CLIR	CKJ	Done	May 15-18, 2007
CLQA	CJE	Nov 1-7, 2006	
Opinion	CJE	late Dec.	
Patent(IR,CL)	JE	Oct 2006	
QA	J	Sept25-Oct20, 2006	
Trend Info (MuST)	J	Dec 2006	(March 2007)

NTCIR workshop: Number of Participating Groups



Number of Active Participants by Tasks

Number of Participants by Tasks



Active Participants

[CLIR]

Academia Sinica
 Chinese Academy of Sciences (ISCAS)
 Huazhong Normal Univ
 Hummingbird
 Institute for Infocomm Research
 Justsystem Corporation
 National Central Univ
 NICT
 National Taiwan Normal Univ
 Newswatch, Co.
 Osaka Kyoiku Univ
 POSTECH
 Queens College
 Queensland Univ of Technology
 Toshiba / NewsWatch
 Univ of Aizu
 Univ of California; Berkeley
 Univ of Montreal
 Univ of Neuchatel
 Univ of Nottingham
 Yahoo! Japan

[CLQA]

Aoyama Gakuin Univ
 Carnegie Mellon Univ
 Chinese Academy of Sciences (ICT)
 Academia Sinica
 Mount Holyoke College
 National Central Univ
 National Cheng Kung Univ
 Queens College
 State Univ of New York at Albany
 Tokyo Institute of Technology (Furui)
 Toyohashi Univ of Technology (Akiba)
 Yokohama National Univ

[MuST]

Hiroshima City Univ
 Justsystem Corporation
 Keio Univ (saito)
 Mie Univ
 NICT
 NEC (Internet Systems Research Labs)
 Ochanomizu Univ (2 groups)
 Okayama Univ
 Osaka Prefecture Univ (3 groups)
 Ritsumeikan Univ
 Tokyo Denki Univ
 Tokyo Institute of Technology
 Tokyo Metropolitan Univ
 Univ of Tokyo (kato)
 Yokohama National Univ

[OPINION]

Cornell Univ
 Illinois Institute of Technology
 Information and Communications Univ
 Chinese Academy of Sciences (ISCAS)
 National Chiao Tung Univ
 National Institute of Informatics
 NICT
 NEC (Internet Systems Research Labs)
 Chinese Univ of Hong Kong
 Toyohashi Univ of Technology (seki)
 Univ of Maryland
 Univ of Sheffield

[PATENT]

Hiroshima City Univ
 Hitachi; Ltd
 Justsystem Corporation
 Nagaoka Univ of Technology
 NICT
 National Taiwan Normal Univ
 NTT DATA
 NTT-CS
 POSTECH
 Toyohashi Univ of Technology (aono)
 Univ of Sheffield
 Univ of Tsukuba

[QAC]

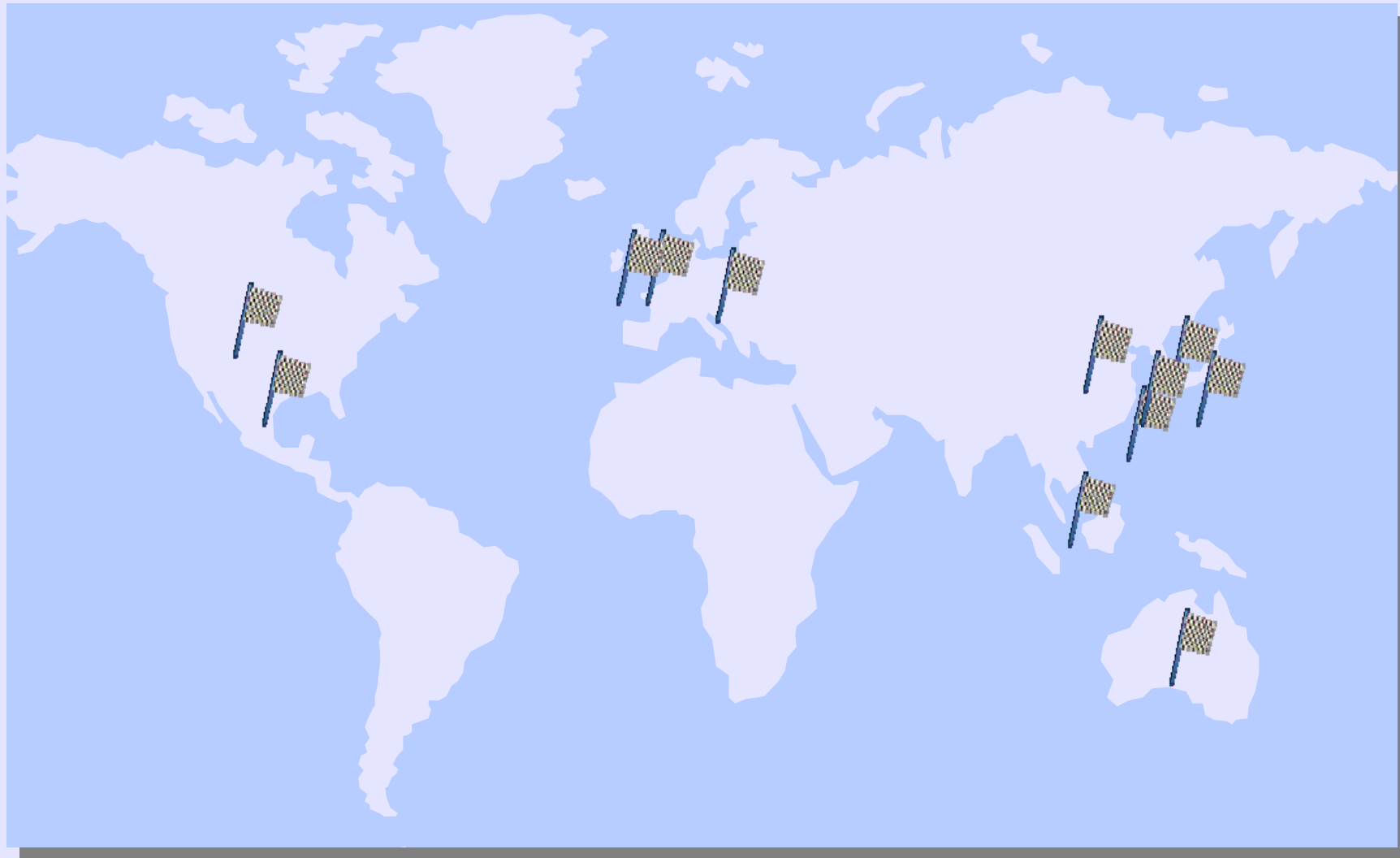
Aoyama Gakuin Univ
 Carnegie Mellon Univ
 Hokkaido University (araki)
 Chinese Academy of Sciences (ISCAS)
 NTT-CS
 Ritsumeikan Univ
 Toyohashi Univ of Technology (akiba)
 Yokohama National Univ

15 new commers

(13 are international)

Many returns

Geographical Distribution of Participants



Geographical Distribution of Active Participants



What were New to NTCIR-4

- Open Submission Session
- ACM-TALIP Special Issue Recommendation
- Open Attendance
- Submission Raw Data
- Online Working Notes and Slides

What's New to NTCIR-5

- Open Submission >>>> continued
- Special Issue on Patent at IP&M
- Open Attendance >>>>continued
- Submission Raw Data >>>>continued
- Online Proceedings and Slides >>>>

Proceedings Only (No working notes)

- Pilot: MuST

What's New to NTCIR-6

- Open Submission >>>> enhanced to **EVIA**
- Special Issue on Patent (IP&M) published
- Open Attendance >>>>continued
- Submission Raw Data >>>> part of participants' dataset
- Online Proceedings and Slides >>>>continued
 - + Proceedings Only (No working notes)>>continued
 - + Publisher's version (page # and running title)
 - + CD contains draft papers.
- **Pilot: MuST, Opinion**
- **Multiple Collections (CLIR, PATENT)**

Multiple TCs

- For more stable/robust evaluation
- Improvements from previous years
 - CLIR, Patent IR (using NTCIR-3,-4,-5)
- For larger test sets with reasonable/manageable work
 - Patent IR (Using NTCIR-3,-4,-5,-6 collections)
 - Need Large # of topics, but limited resources
 - 34 topics: Rel Judgments by Human Experts
 - x K topics of judgments by external searchers
 - x 10K topics of judgments by patent examiners (a few relevant doc per topic)
 - Similar to Click Thro on Web.

NTCIR Workshop 6 (2006-2007) Organizers



Program chair: Noriko Kando, NII

+CLIR

Hsin-Hsi Chen, NTU
Kuang-hua Chen, NTU
Kazuaki Kishida, Surugadai U
Kazuko Kuriyama, Shirayuri U
Sukhoon Lee, NCU

+CLQA

Kuang-hua Chen, NTU
Chuan-Jie Lin, Nat Taiwan Ocean U
Yutaka Sakaki, ATR

+OPINION

Hsin-Hsi Chen, NTU
David K Evans, NII
LunWei Ku, NTU
Chin-Yew Lin, Microsoft Research Asia
Yohei Seki, Toyohashi U Tech,

+PATENT

Atsushi Fujii, Tsukuba U
Makoto Iwayama, Hitachi/TITEC

+QA

Junichi Fukumoto, Ritsumeikan U
Tsuneaki Kato, U Tokyo
Fumito Masui, Mie U
Tatsunori Mori, Yokohama nat U.

+MuST [piloy eotkdhop]

Tsuneaki Kato, Tokyo Univ
Mitsuteru Matsushita, NTT

Acknowledgment

- Central Daily News
 - China Daily News
 - China Times Inc.
 - Chosunilbo
 - Hankooki.com
 - Industrial Property Cooperation Center
 - Japan Patent Office
 - Japan Patent Information Organization
- Korea Economic Daily
 - Linguistic Data Consortium
 - Mainichi Newspaper
 - Nippon Database Kaihatsu, Co. Ltd.
 - NTT
 - NRI Cyber Patent
 - PATOLIS
 - the Sing Tao Group
 - Taiwan News
 - Tokyo Univ
 - 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, Noriko Kando,
Kazuko Kuriyama,

Design of NTCIR-6 CLIR Task

- **STAGE1:** ad hoc retrieval 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.
>>>A New Challenge
 - Purpose: To obtain the more reliable
 - Run the same system across 3 TCs

Evaluation

- **Measures**
 - Official: trec_eval
 - Mean average precision (MAP), R-precision, Recall-Precision graph, B-pref etc.
 - Add: multi-grade relevance based metrics
 - nDCG, Q-measure, Generalized average precision (GAP)

STAGE2 (cont.)

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

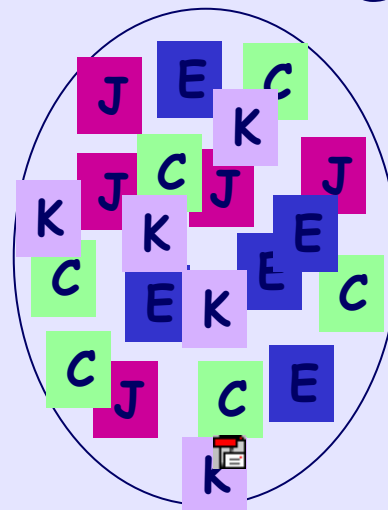
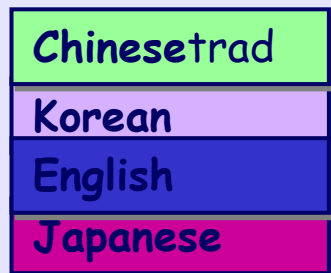


This result shows more clearly dominance of System A than that obtained from an experiment using just a single test collection.

NTCIR CLIR on News

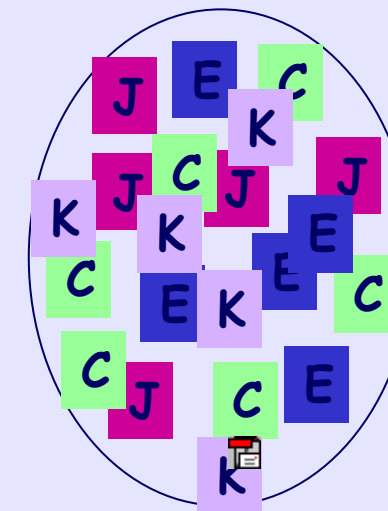
Documents

NTCIR-3
50 topics



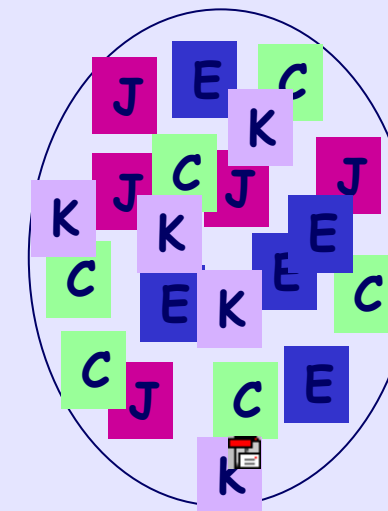
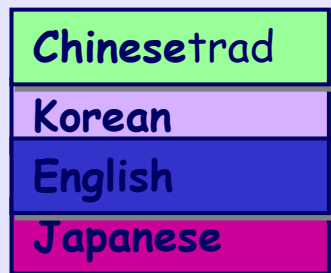
Published in
1998-1999

NTCIR-4
60 topics

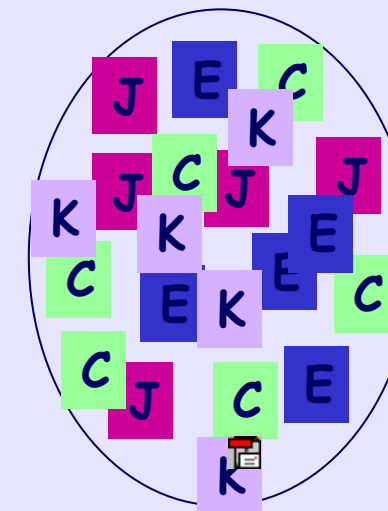


Published in
2000-2001

NTCIR-5
50 topics



NTCIR-6
50 topics



NTCIR CLIR on News

Subtasks

- Multilingual CLIR (MLIR) : e.g., C - CJKE
- Bilingual CLIR (BLIR): e.g., C - J
- Single Language IR (SLIR): e.g., C - C

Languages

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

Relevance Judgments - 4 grades

Highly Relevant (S), Relevant (A), Partial Relevant (B), Non-Relevant (C)

- **Short Q:** D-only and T-only are mandatory
- **Background info** of search requests
- **Balance btw topic-types:**
 - Forcus: NE, OOV
 - proper nouns vs without PN
 - domestic/regional/international

NTCIR
50

NTCIR
60

NTCIR
50

NTCIR
50

ntcir

S

in
9

in
1

NTCIR-6 CLIR

NTCIR-3
30 topics
for 1994



50 topics
for 1998-99

NTCIR-4
60 topics



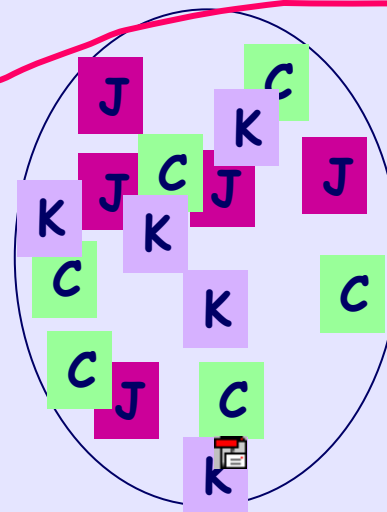
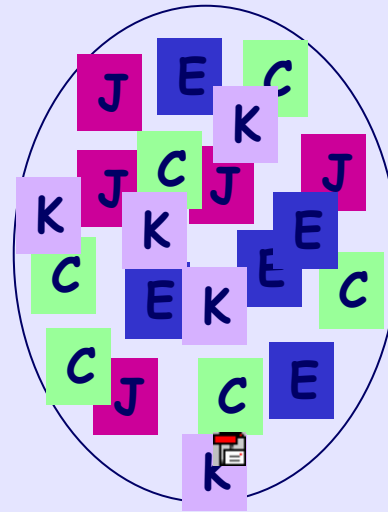
NTCIR-5
50 topics

Selected from
NTCIR-3 & 4
140 topics and
reuse

NTCIR-6
50 topics



Documents



Stage 1

NTCIR-6 CLIR

NTCIR-3
30 topics
for 1994

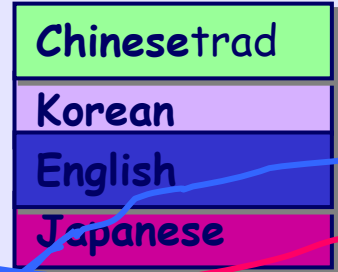


50 topics
for 1998-99

NTCIR-4
60 topics



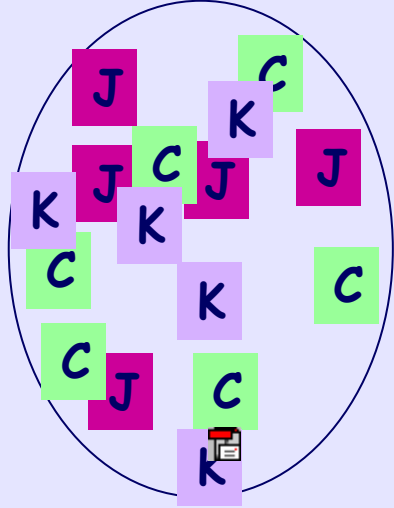
NTCIR-5
50 topics



NTCIR-6
50 topics



Documents



Stage 2

Published in
1998-1999

Stage 1

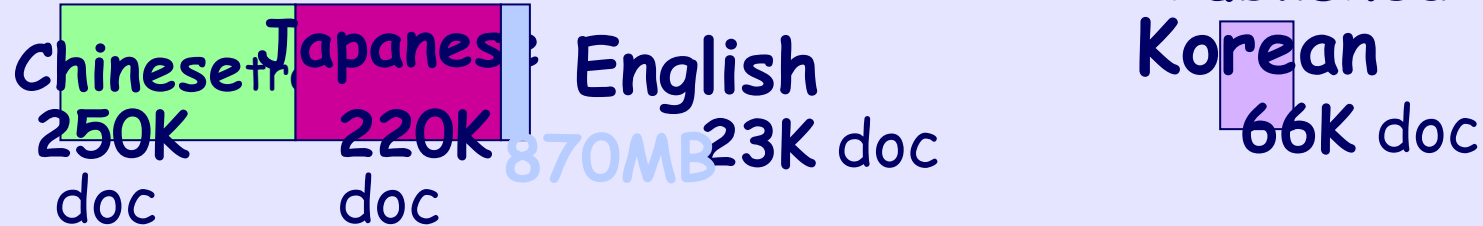
Published in
2000-2001

Documents for CLIR at NTCIR

NTCIR-3

Published in 1998-1999

Published in 1994



NTCIR-4

Published in 1998-1999

3.3GB



NTCIR-5

Published in 2000-2001



Every language is multi-sources.

Documents for CLIR at NTCIR-6

NTCIR-3

Published in 1998-1999

Published in 1994



NTCIR-4

Published in 1998-1999

3.3GB



NTCIR-6

Published in 2000-2001



Every language is multi-sources.

Techniques Used (NTCIR-6)

- IR Models: Logistic Reg, pircs, vsm, okapi, LM, BM25+GAetc.
- Indexing: bigram vs word vs others, hybrid
- Translation disambiguation w/ WEB w/target doc
- Out-of-vocabulary (OOV) problem
 - Use of Web, Wikipedia
 - NE identification
 - Transliteration
 - Cognate
- Query expansion techniques
 - Selective application PRF
- Document re-ranking

BLIR - Comparison of MAP

(D-run, Rigid)

	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%)

Results of STAGE2

- MAP across 3 different test collections: correlation coefficients by types of 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

Cross-Language Question Answering

Task Organizers

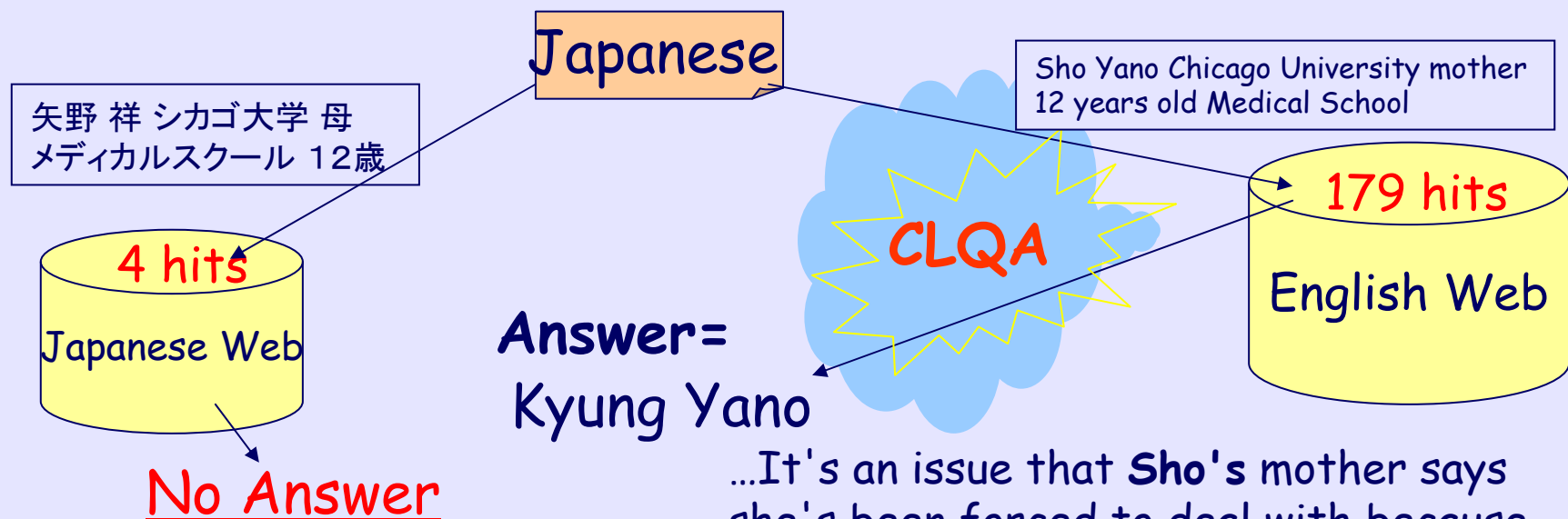
Kuang-hua Chen, NTU

Chuan-Jie Lin , Nat Taiwan Ocean U

Yutaka Sakaki, ATR

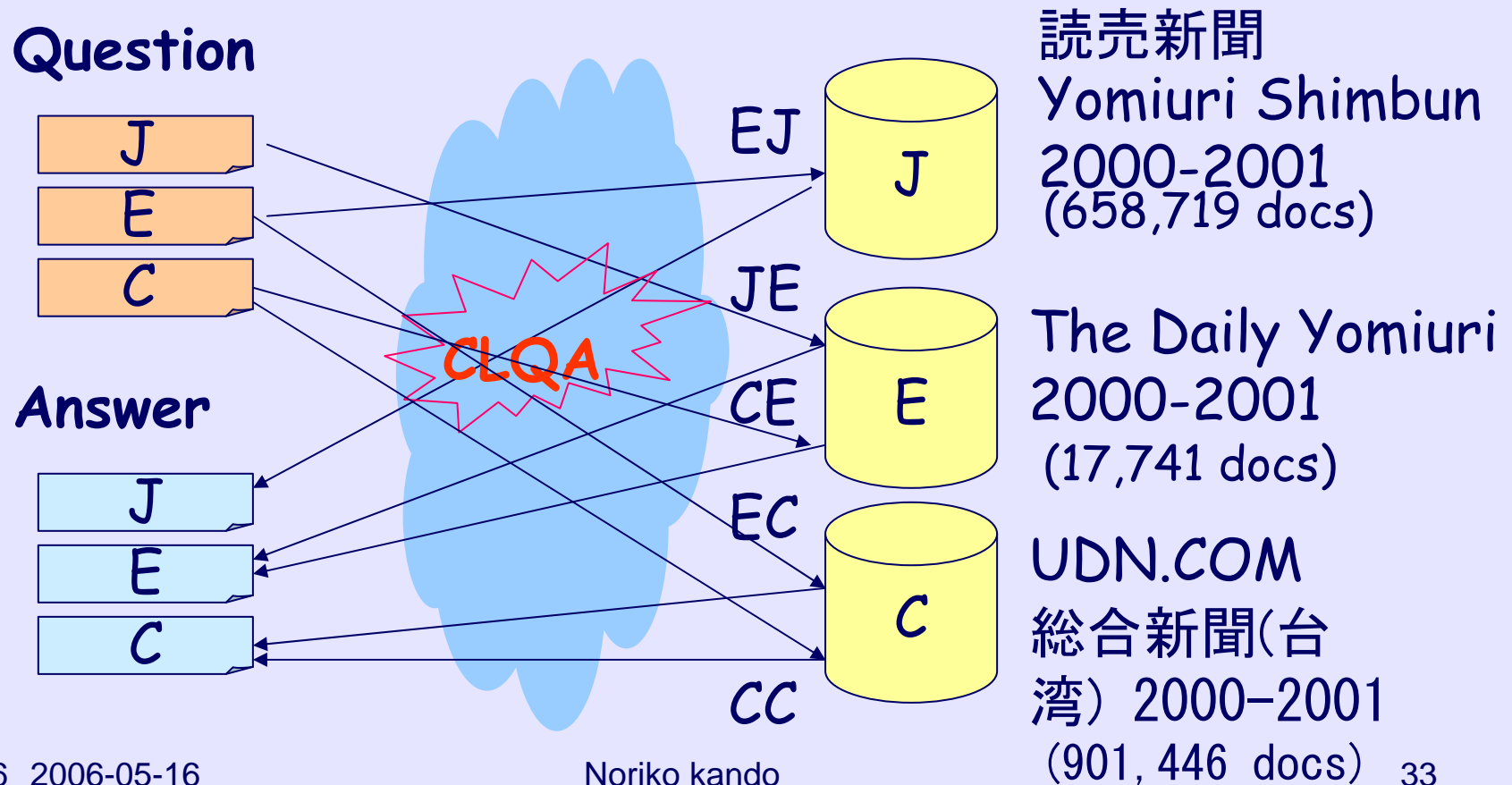
Necessity for Cross-Lingual QA

Question 「12歳でシカゴ大学メディカルスクールに入学した矢野祥君のお母さんの名前は？」
"What is mother's name of the student who goes to the University of Chicago Medical School at 12 years old."



NTCIR-5 CLQA

JE/EJ/CE/CC/EC subtasks



NTCIR-6 CLQA

EC/CC, EE/CE, EJ/JJ, and JE/EE subtasks

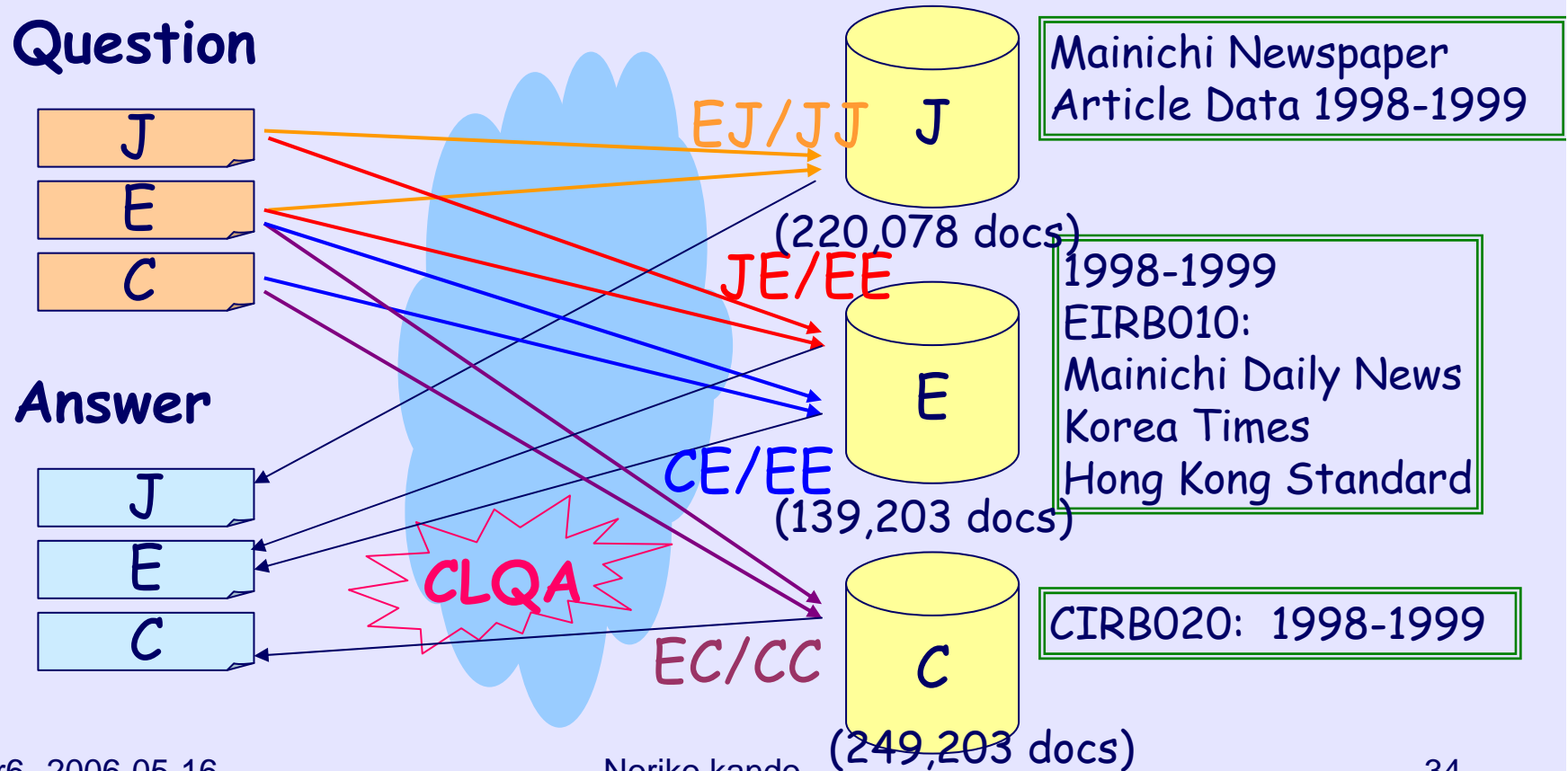


Table 1. Question type distribution of formal run questions

	E-J/J-J/J-E	E-C/C-C/C-E/E-E
ARTIFACT	20	7
DATE	31	39
LOCATION	31	16
MONEY	13	8
NUMEX	20	11
ORGANIZATION	20	16
PERCENT	15	4
PERSON	35	47
TIME	15	2
Total	200	150

Evaluation Metrics

- **Official run**
 - **Accuracy**: the rate at which the top 1 answers are correct.
- **Unofficial run**
 - **MRR (Mean Reciprocal Rank)**: the average reciprocal rank ($1/n$) of the highest rank n of a correct answer for each question.
 - **Top5**: the rate at which at least one correct answer is included in the top 5 answers.

Table 4. Japanese-related CLQA accuracy

Run	NTCIR-6 Right	Right+ Unsupported	NTCIR-5 Right	Right+ Unsupported
Forst-E-J	0.175	0.195	0.125	0.155
Forst-J-J	0.310	0.335	0.170	0.265
HARAD-J-J	0.085	0.110	-	-
LTI-E-J	0.095	0.115	0.100	0.125
LTI-J-J-u	0.335	0.360	0.080	0.200
TITFL-E-J	0.030	0.065	-	-
TITFL-J-J	0.155	0.190	-	-
TTH-E-J	0.130	0.165	-	-
TTH-J-J	0.270	0.295	-	-

Table 5. Chinese-related CLQA accuracy

Run	NTCIR-6 Right	Right+ Unsupported	NTCIR-5 Right	Right+ Unsupported
IASL-EC	0.253	0.340	-	-
IASL-CC	0.520	0.547	0.375	0.445
ICDCU-CC	0.287	0.340	-	-
ILS-EC	0.093	0.107	-	-
LTI-EC	0.147	0.200	0.075	0.095
LTI-CC	0.253	0.260	-	-
MHC-EC	0.040	0.073	-	-
MHC-CC	0.187	0.213	-	-
MHC-EE	0.187	0.207	-	-
NCUTW-EC	0.000	0.040	-	-
NCUTW-CC	0.087	0.113	-	-
pircs-EC	0.253	0.280	0.125	0.165
pircs-CC	0.420	0.447	-	-
WMMKS-EC	0.053	0.067	0.040	0.045
WMMKS-CC	0.133	0.153	0.320	0.350

Findings

- CL vs Mono
 - E-J vs J-J: about 50% of Accuracy
 - E-C vs C-C: "Veterans" worked better
 - LTI, PIRCS 60%
 - IASL, WMMKS 40%, 47.2
 - Other newcomers less than 20%
- Synonyms
 - QID T0054: What is Japan's unemployment rate for May of 1997? → no answers reported
 - QID T0123: What was the Japan's jobless in May 1986
- IR for QA
 - IR module showed largest performance drop in module by module analysis.
 - Extrinsic Evaluation of IR?

NTCIR-6 Opinion Task

- Hsin-Hsi Chen, David Kirk Evans, Lun We Ku, Chin-Yew Lin, Yohei Seki

Opinion Analysis - Roadmap

Genre	Subjectivity	Holder	Polarity	Strength
News	NTCIR-6	NTCIR-6	NTCIR-6	
Review	NTCIR-7	NTCIR-7	NTCIR-7	NTCIR-7
Blog	NTCIR-8	NTCIR-8	NTCIR-8	NTCIR-8

Stakeholder	Temporal	Language	Granularity	Application
		Chinese	single-sent	Summarization
NTCIR-7		English	clause	QA
NTCIR-8	NTCIR-8	Japanese	multi-sent	Opinion tracking
		CJE	document	Consistency checkir
				Trend

Chinese, Japanese,
English

Corpus Annotation

- Three annotators per document
- ~ 20 docs per topic (EN, JA), 40 CH
- 1998~2001 data
- CH annotators students, JA news-related, EN translators & teachers

Feature	Value	Req'd?
Opinionated	YES, NO	Yes
Opinion Holder	String, multiple per sentence possible	Yes
Relevant	YES, NO	No
Polarity	Positive, Neutral, Negative	No ₄₂

Corpus Sources

- Document

- Japanese: 1998-2001 Yomiuri, Mainichi newspapers
- Chinese: 1998-2001 CIRB020, CIRB040
- English: 1998-2001 Mainichi Daily News, Korea Times, Xinghua

- Topics

- 32(C), 30 (J), 28 (E) topics and associated document sets were selected from 160 NTCIR3-5 CLIR Topics (translated into CKJE) for 1998-2001

Annotator Agreement

Lang	Min	Max	Avg.
CH	.0537	.4065	.2328
EN	.1704	.4806	.2947
JA	.5997	.7681	.6740

Annotator Agreement

- EN, JA have consistent annotators
- CH uses 3 annotators from pool of 7 (per-topic agreement)
- JA high agreement
- EN #3 difficult!

La	r	Task	Kappa
E	1-2	Opinionated	0.4806
E	1-3	Opinionated	0.1704
E	2-3	Opinionated	0.2332
E	1-2	Relevant	0.5240
E	1-3	Relevant	0.0618
E	2-3	Relevant	0.5298
E	1-2	Polarity	0.5457
E	1-3	Polarity	0.2039
E	2-3	Polarity	0.2645
J	1-2	Opinionated	0.6541
J	1-3	Opinionated	0.5997
J	2-3	Opinionated	0.7681
J	1-2	Relevant	0.7176
J	1-3	Relevant	0.6966
J	2-3	Relevant	0.8394
J	1-2	Polarity	0.6919
J	1-3	Polarity	0.6367
J	2-3	Polarity	0.7875

Corpus

Lang	Topics	Docs	Sents	Opin.	Rel.
CH	32	843	8,546	62% / 25%	39% / 16%
EN	28	439	12,525	30% / 7%	69% / 37%
JA	30	490	8,523	29% / 22%	64% / 49%

Evaluation Metrics

- Precision, Recall, F-Measure over opinionated, relevant, polarity
- Semi-automatic evaluation of opinion holders (precision, recall, f-measure)
- Multiple approaches developed

Polarity Differences (Strict System POS)

Annotation				Behavior	Annotation				Behavior
PO S	NE U	NE G	NO T		PO S	NE U	NE G	NO T	
3	0	0	0	LWK+, DKE+, YS+	3	0	0	0	LWK+, DKE+, YS+
2	0	1	0	LWK skip, DKE-, YS-	2	0	1	0	LWK +, DKE ^{+2/3} , YS+
0	0	0	3	LWK-, DKE-, YS-	0	0	0	3	LWK-, DKE-, YS-
0	0	1	2	LWK skip, DKE-, YS-	0	0	1	2	LWK-, DKE-, YS-
					1	0	2	0	LWK-, DKE ^{+1/3} , YS-
					1	1	0	1	LWK+, DKE ^{+1/3} , YS prec. down, recall no change

Holder evaluation

- Semi-automatic evaluation
- Match system extracted holders to annotator holder list, automate the process in some way
- Time consuming, only first priority run evaluated

Discussion

- Easiest? Relevance < Opinionated < Polarity
- CH, EN, JA corpora have different annotator agreement: training issue or data issue?
- How to evaluate with different annotator tags? 3 approaches show different results

Future Work

- Increase group participation in multiple languages (only TUT, GATE this year)
- What is upper bound on annotator performance
- How good is "good enough"
- Towards consistency across languages

Patent Retrieval Task at NTCIR-6

Atsushi Fujii (Univ of Tsukuba)

Makoto Iwayama (Hitachi,
Ltd./TITECH)

Noriko Kando (NII)

Outline: Subtasks

- Japanese Retrieval
 - Invalidity search for Japanese patent applications
- English Retrieval
 - Invalidity search for USPTO patents
- Japanese Classification
 - F-term classification for Japanese patent applications

Japanese Retrieval Subtask

- Find the patents that can invalidate the demand in a patent claim
- patent-to-patent retrieval
 - Both queries and documents are patents
- This task is usually performed by
 - examiners in a government patent office
 - searchers of IP division in private companies
- Document collection
 - 10 years of Unexamined patent applications published in 1993-2002
 - 3.5 M documents

Japanese Retrieval Subtask (cont.)

- Document collection
 - 10 years of Unexamined patent applications published in 1993-2002
 - 3.5 M documents
 - Search topic sets
 - NTCIR-4: 34
 - NTCIR-5: 1189
 - Search report (SR): 349
 - NTCIR-6: 1685
- with expert judgments
- only citations were used as relevant docs

Search topic sets

Recall-oriented

34 Qs with human judgment from NTCIR-4

36 Qs with human judgment from NTCIR-3 for technological survey task

349 Q from search reports

Max. ~100,000/year

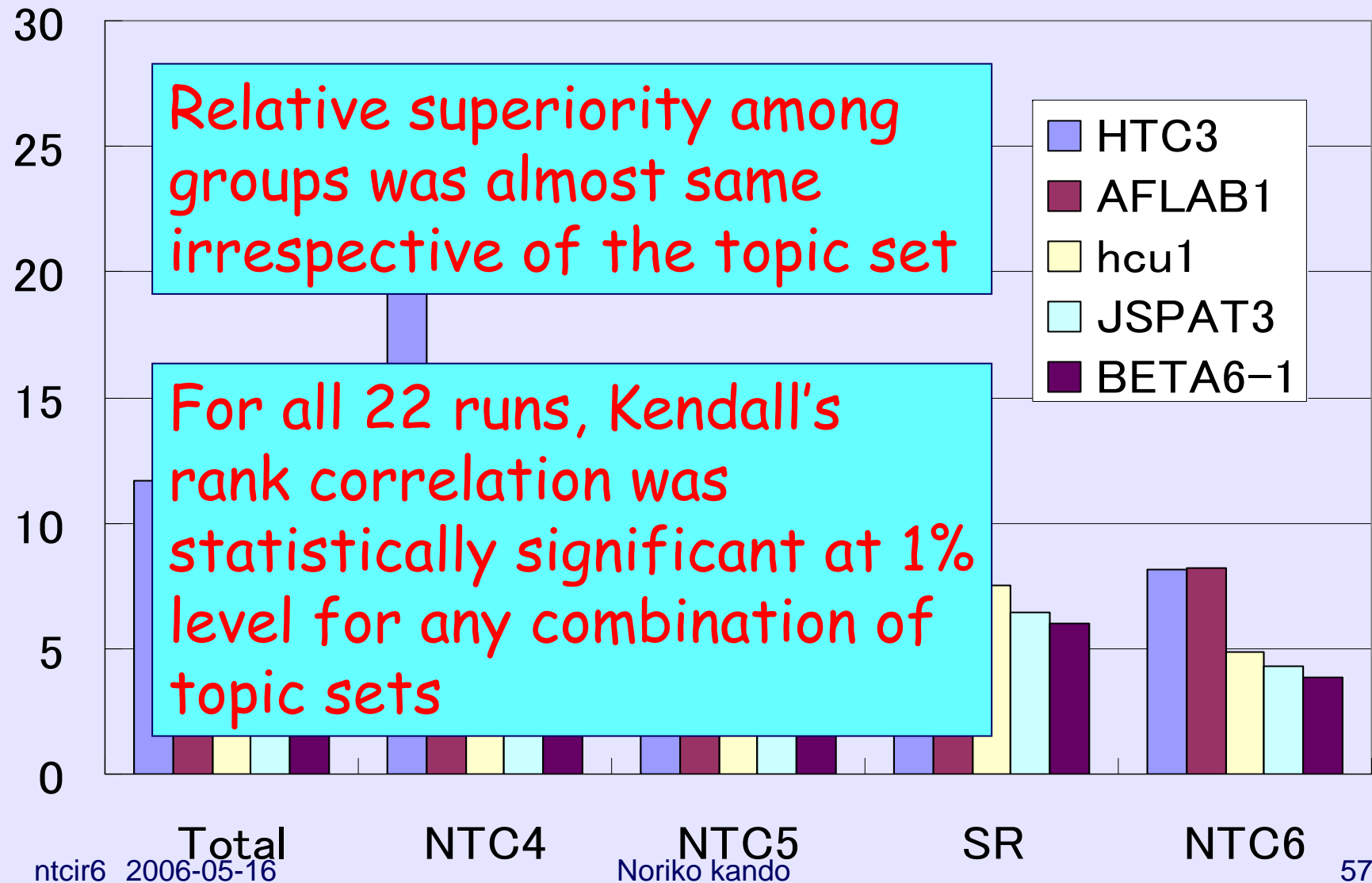
1189 Q from Patent Office Examiners' Citation (ntcir5)

1685 Q from Patent Office Examiners' Citation (NTCIR6)

Precision oriented

Max. ~300,000/year

MAP of Japanese Retrieval Relaxed



English Retrieval Subtask

- Document collection
 - Patents granted by USPTO in 1993-2000
 - 980 K documents
- Search topics
 - Patents granted by USPTO in 2001-2002
 - 1000 topics for training
 - 2221 topics for formal run
- Relevant documents
 - Citations listed in the topic patent (provided by applicants and examiners)

Evaluation Result: MAP

Run ID	Rigid	Relaxed
AFLAB1	3.65	7.12
hcu1	3.37	6.49
hcu2	3.37	6.49
KLE1	2.82	5.72
NTNU	2.30	4.50
JSPATO	1.27	2.12
JSPAT1	1.26	2.10

Discussion

- MAP was low irrespective of the run and the relevance degree
- Selection of search topics should be revised
- AFLAB improved the result by integrating text retrieval and citation analysis (as optional runs)

Classification Subtask

- Purpose is to identify F-term for input patent applications
- Training data
 - Patent applications in 1993-1997
- Test data
 - Patent applications in 1998-1999
 - Topics: 21,606 applications

Goal: Patent Map Creation

Example map for optical disk

← problems →

	high density	erasing	rewriting
managing the number of rewriting			1993-000003
shifting the writing position			1994-000002
laser power pulse waveform		1996-000005 1994-000008	

← solutions →

patent map creation =
multi faceted patent clustering

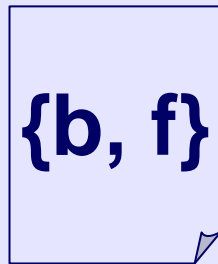
Category Matching based on Text Retrieval (exact match)

queries that can retrieve the test doc.

$\{B, F\}$ ← matching → $\{E, F\}$ $R=1/2$
 $P=1/2$

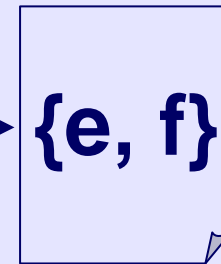
query "B" =
retrieve documents
with "b"

test
doc.



**correct
categories**

conventional
matching

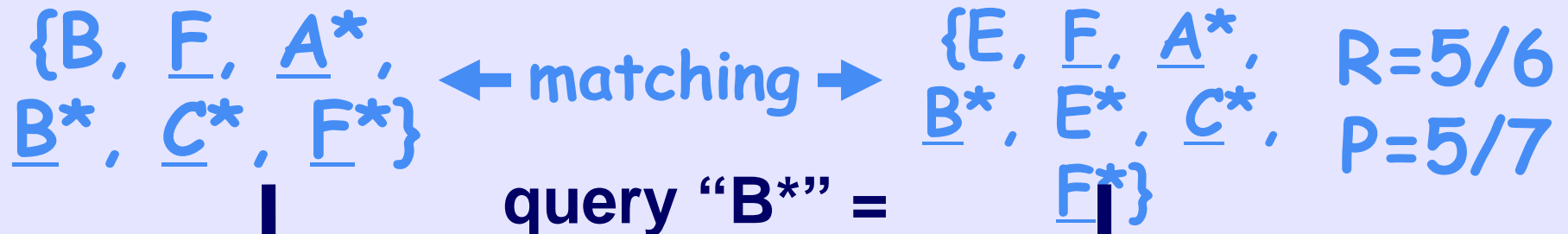


**submitted
categories**

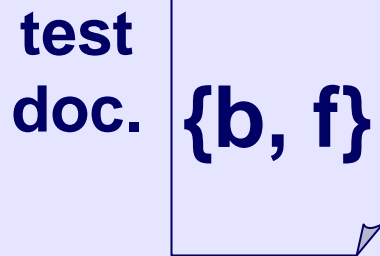
$R=1/2$
 $P=1/2$

Category Matching based on Text Retrieval (relaxed match)

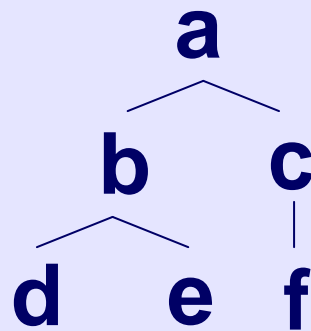
queries that can retrieve the test doc.



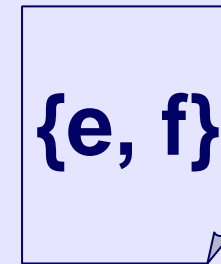
query "B*" =
 retrieve documents with
 "b" or any subcategory
 under "b"



correct categories



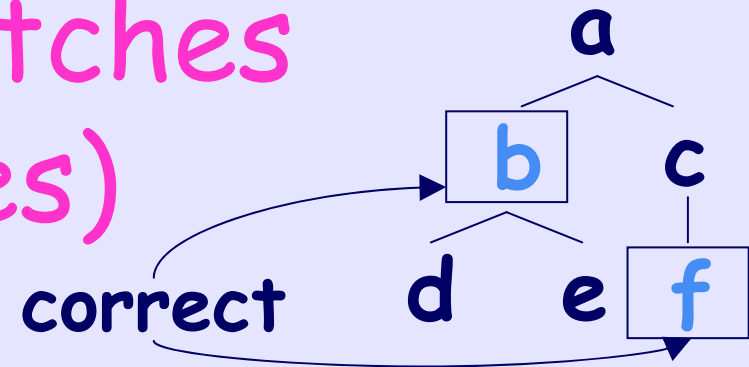
Noriko kando



submitted categories

$R=1/2$
 $P=1/2$

Relaxed Matches (examples)

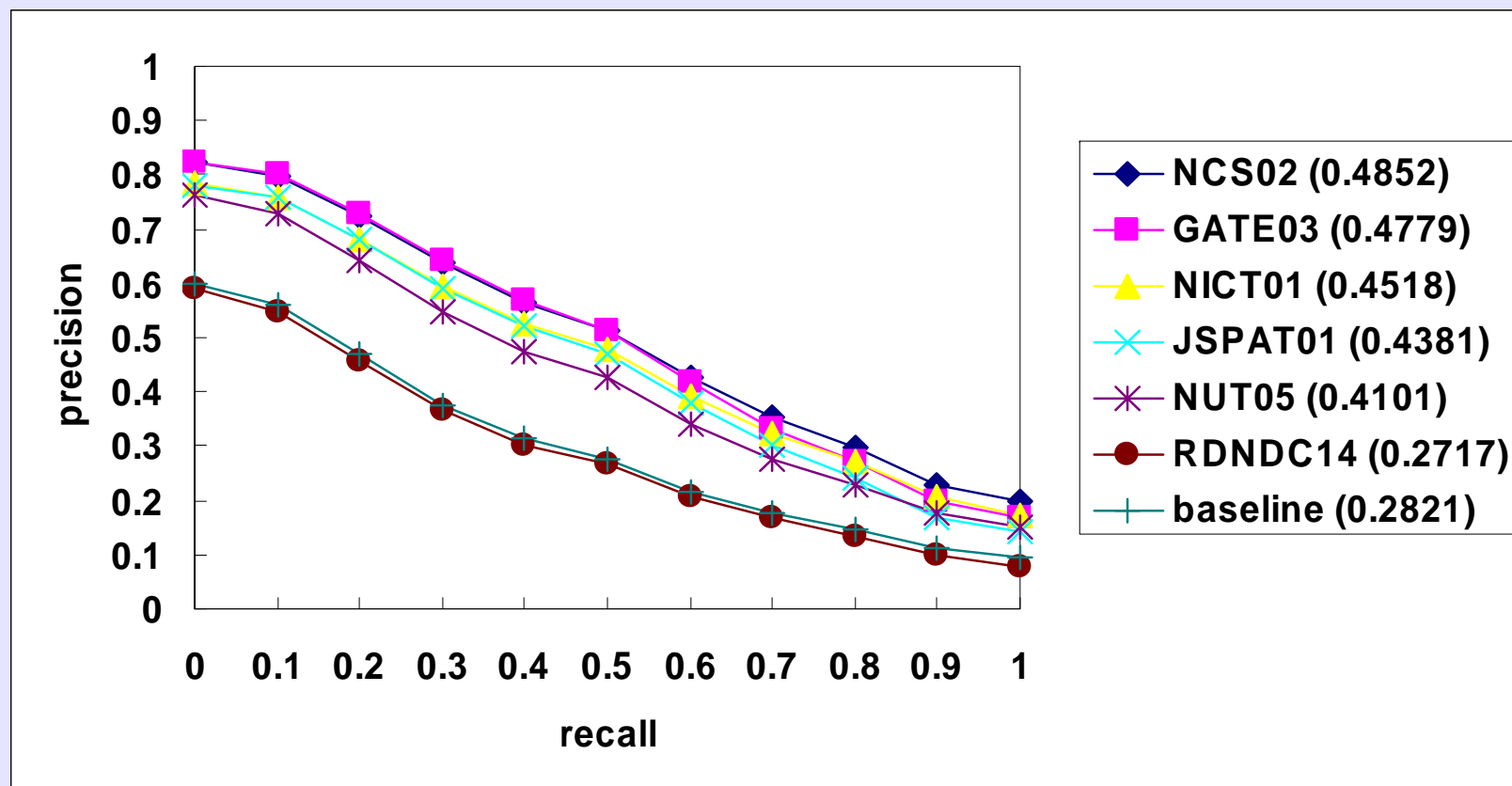


categories	queries	recall	precision
{b,f}	{B,F,A*,B*,C*,F*}	1	1
{f}	{F,A*,C*,F*}	4/6	1
{b}	{B,A*,B*}	3/6	1
{c}	{C,A*,C*}	2/6	2/3
{a}	{A,A*}	1/6	1/2
{b,f,c}	{B,F,C,A*,B*,C*,F*}	1	6/7

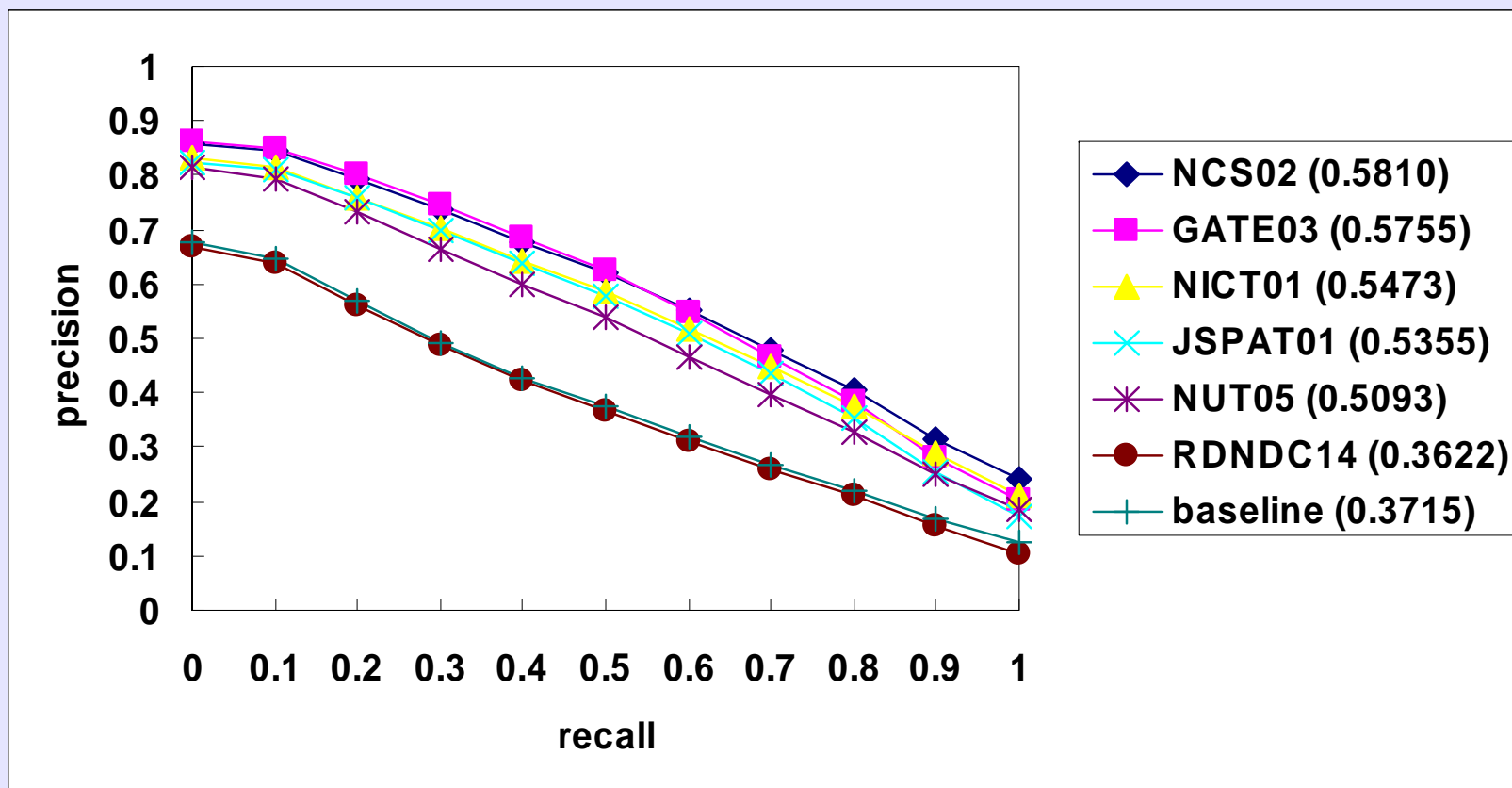
Results (MAP, F-measure)

system	exact match		relaxed match	
	MAP	F-measure	MAP	F-measure
NCS02	0.4852	0.4037	0.5810	0.4970
GATE03	0.4779	0.4125	0.5755	0.5109
NICT01	0.4518	0.3840	0.5473	0.4767
JSPAT01	0.4381	0.3038	0.5355	0.3680
NUT05	0.4101	0.2432	0.5093	0.3838
RDNDC14	0.2717	0.2414	0.3622	0.3431
baseline	0.2821		0.3715	

Results (exact match)



Results (relaxed match)



Plan for NTCIR-7 (tentative)

- Two subtasks
 - Patent mining
 - Related to retrieval and classification
 - Patent translation
 - Patent families as parallel corpus
 - Statistical MT engines are available to the public

Previous QACs

- Evaluation of open domain question answering
 - Main task (5 ranked answers)
 - List task (all answers)
 - Information Access Dialogue (IAD) task
- Factoid question in QAC-1,2,3

Task Description

- Question Answering Track
 - Question answering evaluation using non-factoid questions 100 Q.
- Evaluation Track
 - Open evaluation using QAC-4 evaluation results

Evaluation criterion

- Human evaluation measure
 - Level A: System answer has almost the same contents as one of the correct answers.
 - Level B: System answer includes the contents of one of the correct answers.
 - Level C: System answer includes some part (not all one) of the contents of the correct answers.
 - Level D: System answer includes no information of any of the contents of the correct answers.

Evaluation results of system answers

System ID	All answers	A	B	C	D	No answer
Forest1	591	45	104	34	408	0
Forest2	317	30	52	21	214	2
HOMIO1	100	5	4	7	84	0
HOMIO2	100	3	7	4	86	0
LTI-J	377	24	30	13	310	1
NCQAW1	330	37	15	6	272	32
NCQAW2	323	31	11	4	277	32
NICT1	345	25	65	14	241	0
NICT2	363	6	119	24	214	0
HARAD	204	21	7	7	169	38
RitsQ	286	31	6	14	235	15
TTH1	353	34	36	24	259	0
TTH2	394	22	42	24	306	0
TTH3	354	30	43	26	255	0
Sum	4236	344	541	222	3330	120
average	302.6	24.6	38.6	15.9	237.9	8.6

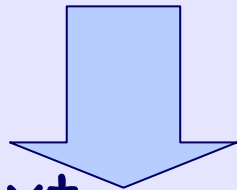
Multimodal summarization for Trend Information

Queries on trends

"How the price of gasoline shifted during the year?"

"What the situation has been in the PC market?"

"How terrible the typhoons were last autumn?"



- Concise, plain text
- Information graphics
- Multimedia presentation
 - text including references to graphics
 - graphics annotated with text



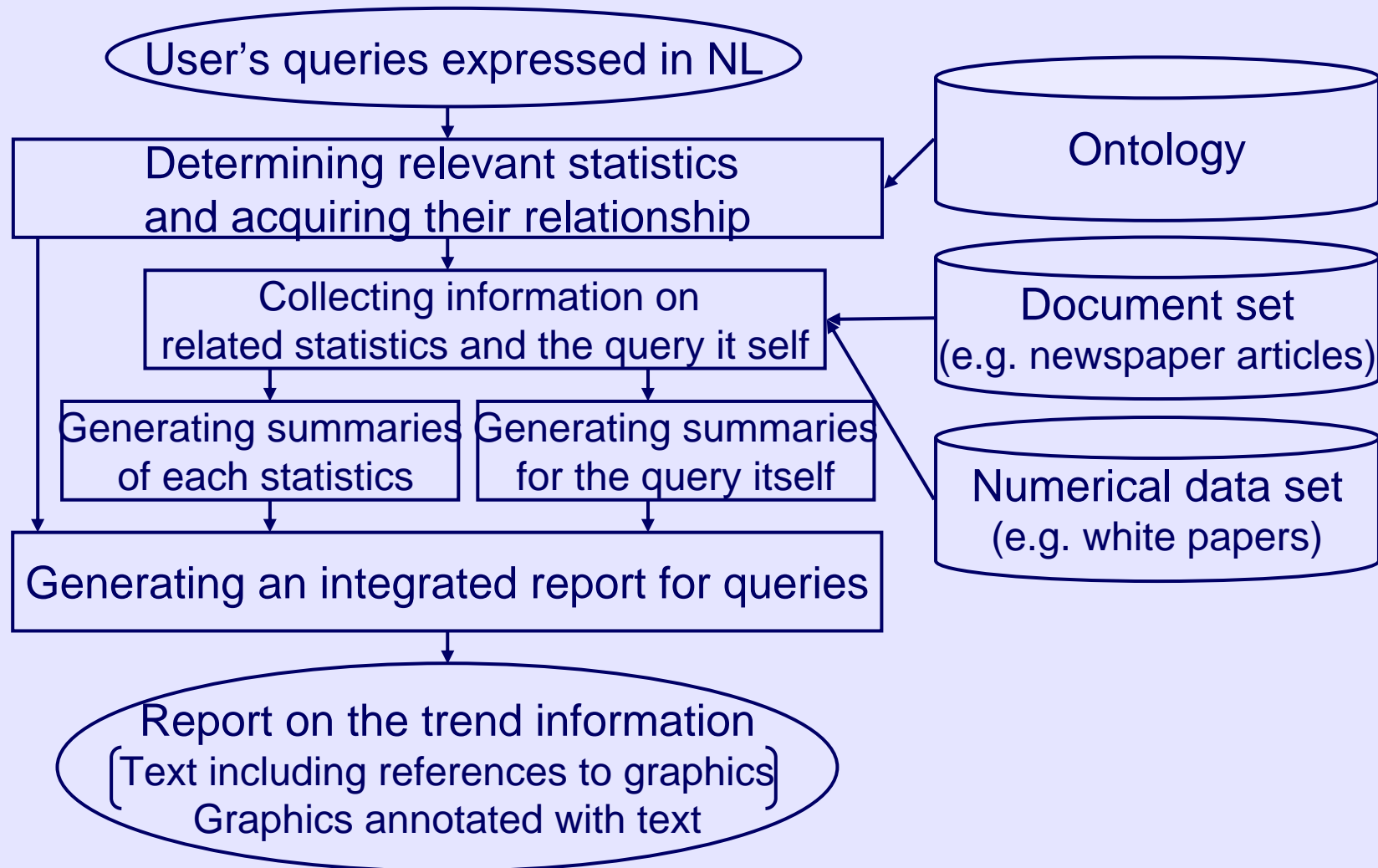
Trend Information

- Summarization of temporal statistical data, obtained through synthesis rather than enumeration
 - changes in product price and sales
 - public approval rating of political parties
- Not always single-dimensional temporal information; can be multi-dimensional
 - Market share of a given product
 - Land price

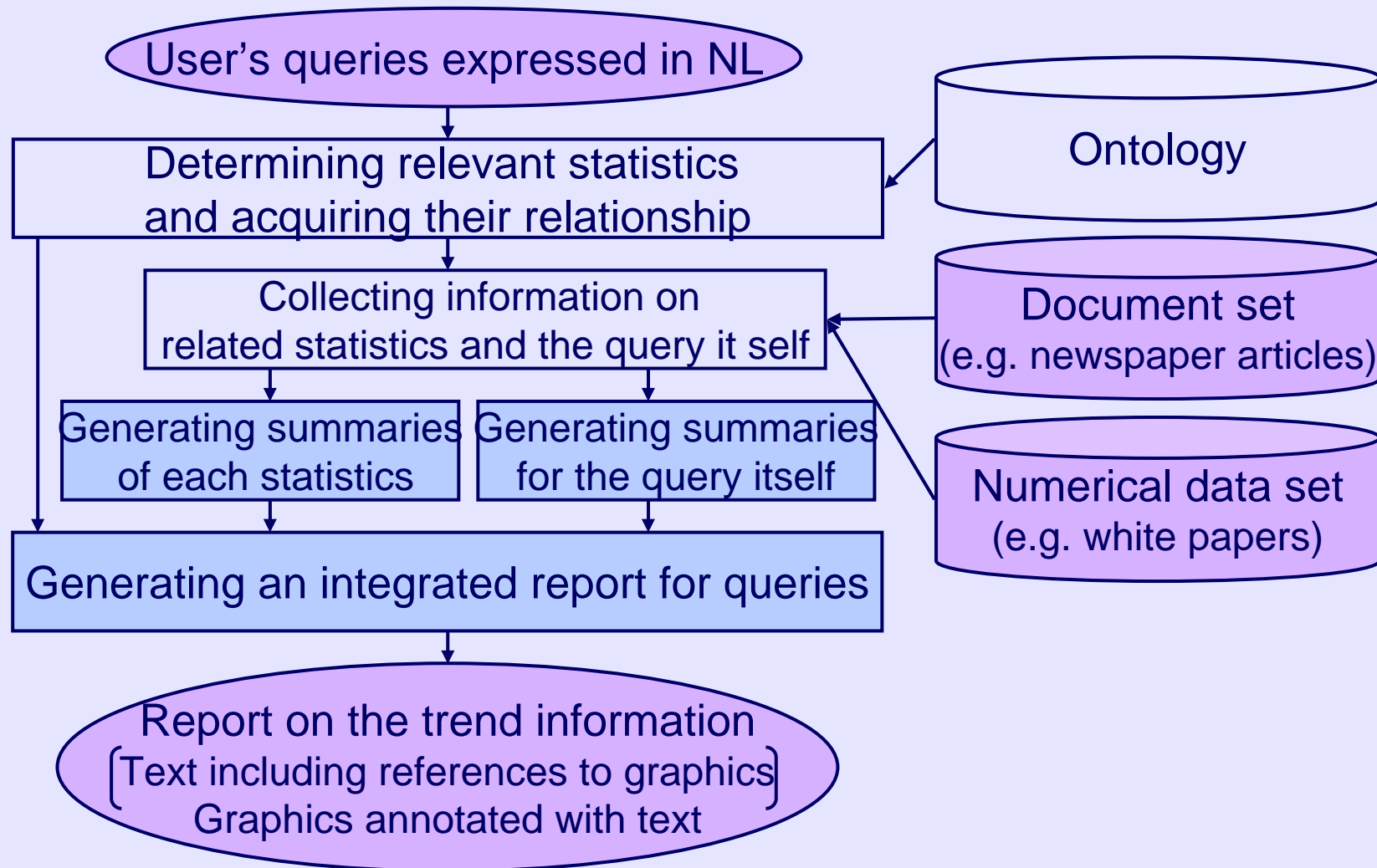
Characteristics

- To encouraging cooperative studies
 - Promoting discussion
 - Conforming communities
 - Constructing and accumulating resources
- Shared research resource
- Building a Community
 - (Loosely) shared theme of research

Framework



Framework



The Roles of Data Set

Information Collected

Articles, Tables and Charts

Multimodal
Summarization



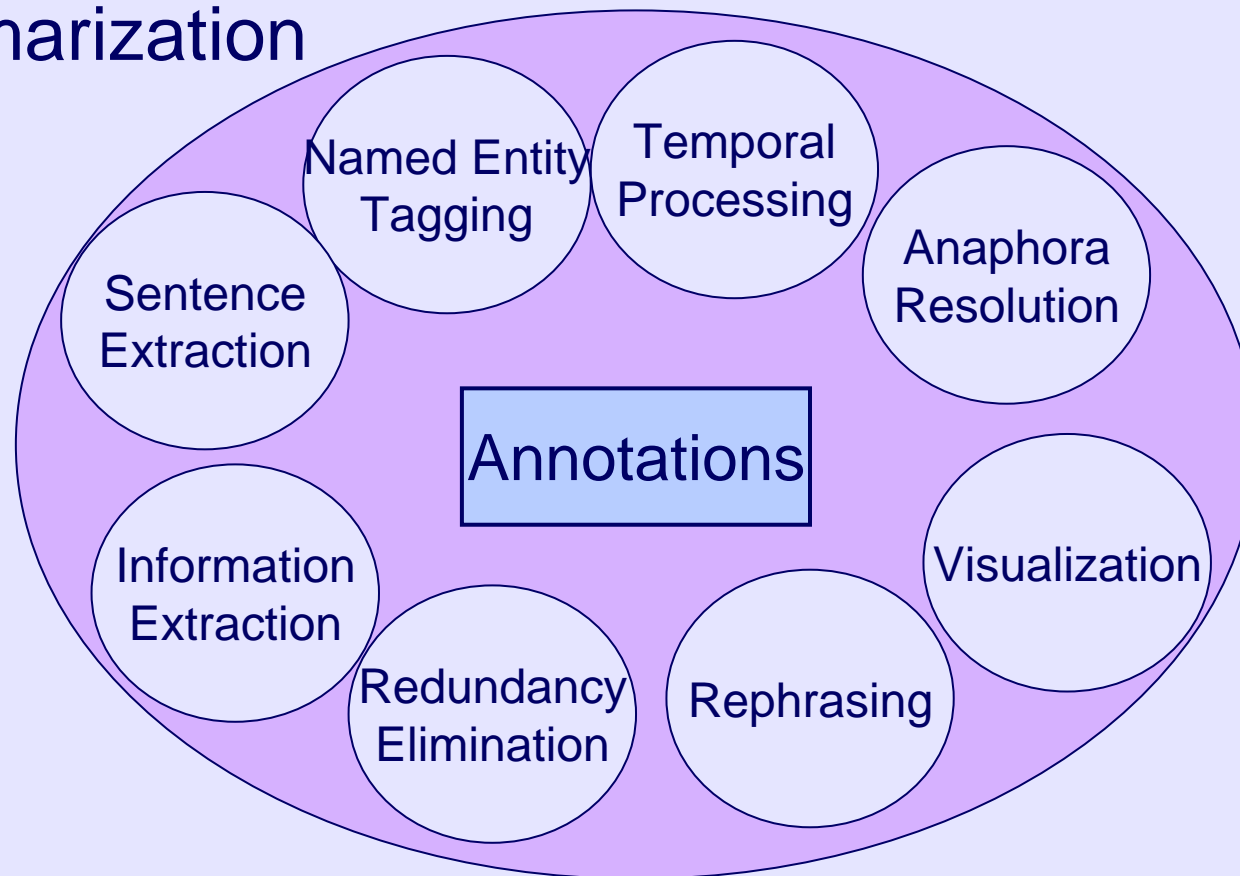
Annotations

Summaries, Reports

Textual summaries, Charts and Tables

The Roles of Annotation

Multimodal Summarization



NTCIR-7 Proposals

- Complex CLQA (CCLQA)
- CLIR For Blog (CLIRB)
- Multilingual Opinion Analysis Task (or Multilingual Evaluation of Opinions on the Web) (MOAT)
- Multi-modal Summarization for Trend Information (MuST)
- Patent Processing Task (translation, mining) (PAT)
- Question Answering Challenge (QAC5)
- Simplified Chinese Information Retrieval, as part of CLIR (CLIR-SC)
- User Satisfaction Task (USAT)

**Under review by NTCIR-7 PC With consideration
of the discussion during this Meeting, selection will
be made in June and announce through ML and
WEB**

Some thought on Future

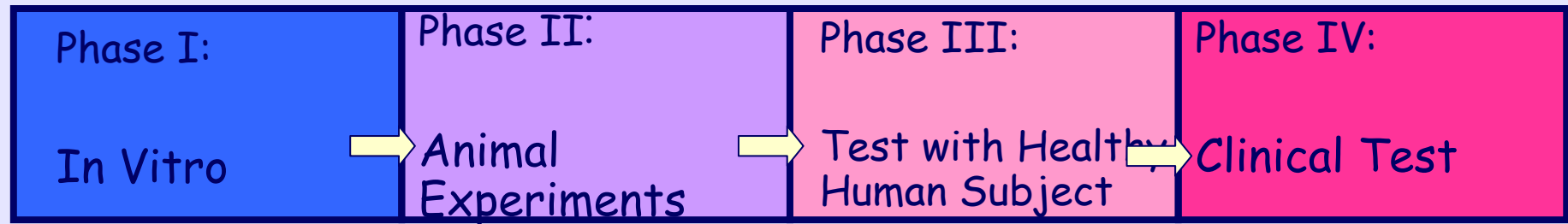
- Evaluation Methodology must keep improving according to the improvement of technologies and social environment.
 - WEB and various document genres including traditionally available
 - Users: User's Task, purpose, situation, adaptive information access
 - Interactive & Exploratory: estimate the users situation and query characteristics
 - Intrinsic vs Extrinsic Evaluation ex.CLIR for QA
 - Synergy
 - Retrieval -> Utilize Info in Doc -> "To know"

IR Systems Evaluation

- Engineering Level: Efficiency
- Input Level: ex. Exhaustivity, quality, novelty of DB
- ➔ **Process Level: Effectiveness ex. recall, precision**
- Output Level: Display of output
- User Level: ex. Effort that users need
- Social Level: ex. Importance (Cleverdon & Keen 1966)

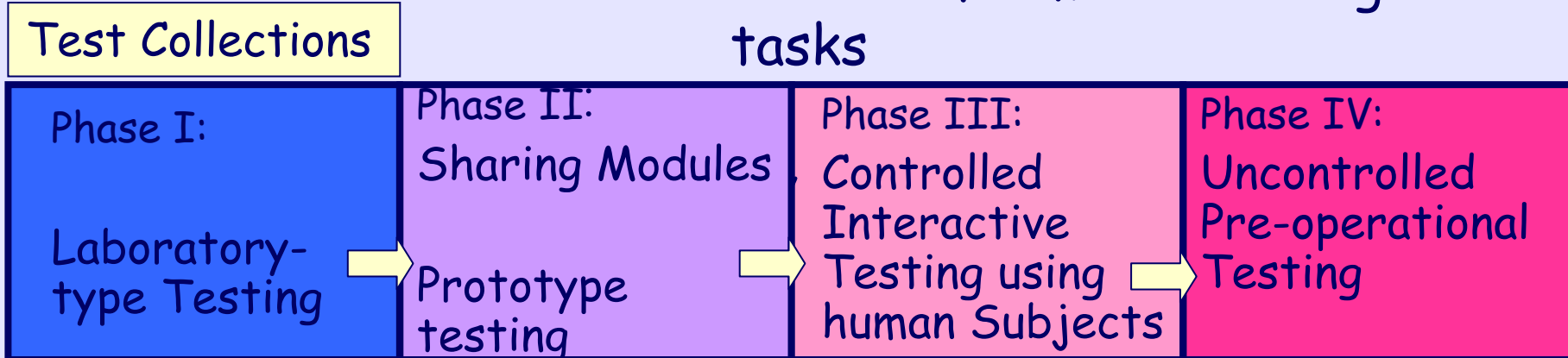
TC usable to evaluate?

Pharmaceutical R & D

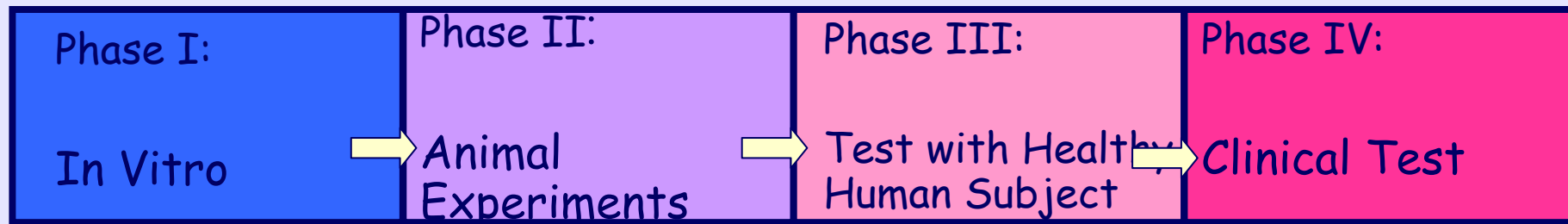


TC usable to evaluate what?

Users' information seeing tasks



Pharmaceutical R & D



Levels of Evaluation



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/>

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/>