

From Shared Tasks to Communities: Lessons from NTCIR's 26-Year Journey

pt.1 NTCIR History and Test Collection Development

LREC 2026 Tutorial
Noriko Kando
NII, Japan

In this tutorial,

- 1) Why Evaluation? Why shared tasks?
- 2) How to design and run a shared task?
- 3) How to enjoy running a shared task and community building
- 4) Why and how communities work?

2) How to design and run a shared task?

- Task Description
 - Any Ground challenge?
- Dataset
 - Documents
 - Topics
 - Annotations
 - Criteria
 - Relevance Judgments, Right Answers, etc
 - Who ? Human or LLM
 - Quality assessment
- Evaluation Method and Metrics
 - Intrinsic, Extrinsic, evaluation by outcomes

Pt.1 NTCIR History and Test Collection Development

Pre-NTCIR, NTCIR-1 (1998/1999) to NTCIR-9 (2010/2011)

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

■ Data sets, evaluation methodologies, and forum

Project started in late 1997

■ Once every 18 months

Data sets (Test collections or TCs)

■ Scientific, news, patents, web, CQA, Wikipedia, Entrance Exams

■ Chinese, Korean, Japanese, and English

Tasks (Research Areas)

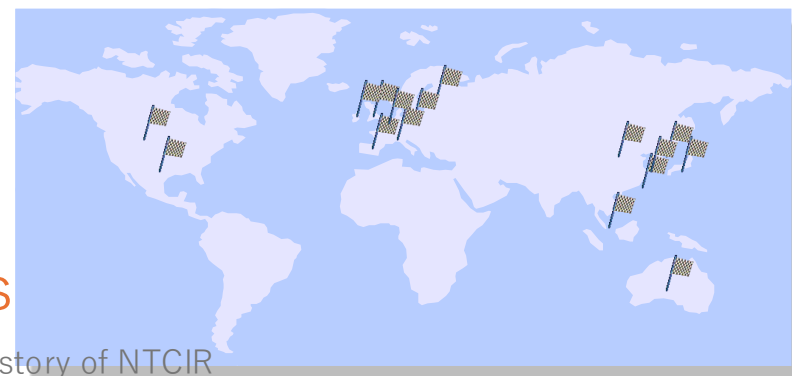
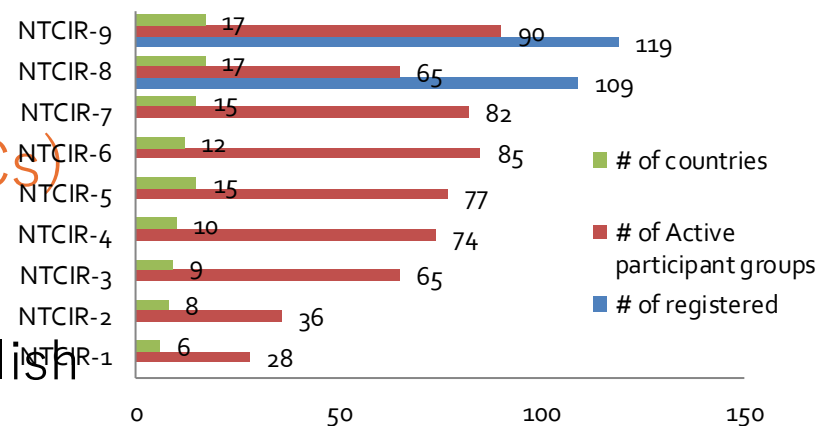
■ IR: Cross-lingual tasks, patents, web, Geo

■ QA : Monolingual tasks, cross-lingual tasks

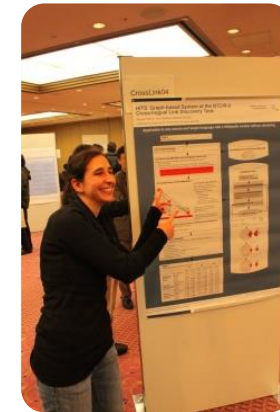
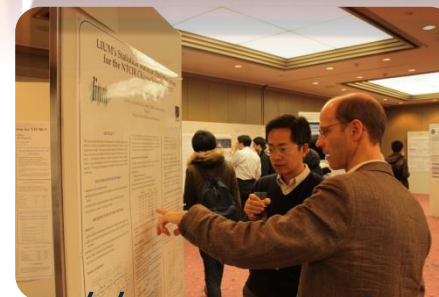
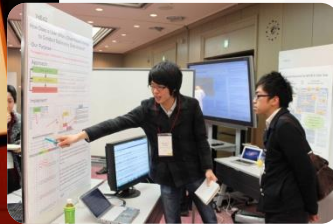
■ Summarization, trend info., patent maps

■ Opinion analysis, text mining

Community-based Research Activities



NTCIR:



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

NTCIR: Workshop Meeting



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

Test Collections = Docs + Topics/Questions + Answers

	Ad Hoc/ CLIR [Scientific Abstracts] (日本語・ 英語検索)	Chinese IR (中国語検索)	CLIR [News] (言語検索 検索)	CLQA (言語検索・ 質問応答)	MuST (「動向情報 要約と可視 化」)	OPINION (意見分析)	PATENT (特許検索・ 分類)	QAC (質問応答)	TMREC (用語抽出)	TSC (要約)	WEB
			ACLIA (高度言語検索 情報アクセス)			MOAT (多言語 意見分析)	PATMH (特許マイ ニング)	PATMT (特許類 聚)			
NTCIR-1	NTCIR-1 Ad Hoc/ CLIR	-	-	-	-	-	-	-	NTCIR-1 TMREC	-	-
NTCIR-2	NTCIR-2 Ad Hoc/ CLIR	CIRBO10	-	-	-	-	-	-	-	NTCIR-2 SUMM	-
NTCIR-3	-	-	NTCIR-3 CLIR	-	-	-	NTCIR-3 PATENT 特許検索	NTCIR-3 QA	-	NTCIR-3 SUMM	NTCIR-3 WEB
NTCIR-4	-	-	NTCIR-4 CLIR	-	-	-	NTCIR-4 PATENT 特許検索	NTCIR-4 QA	-	NTCIR-4 SUMM	NTCIR-4 WEB
NTCIR-5	-	-	NTCIR-5 CLIR	NTCIR-5 CLQA	-	-	NTCIR-5 PATENT 特許検索 分類	NTCIR-5 QA	-	-	NTCIR-5 WEB
NTCIR-6	-	-	NTCIR-6 CLIR	NTCIR-6 CLQA	NTCIR-6 MuST	NTCIR-6 OPINION	NTCIR-6 PATENT 特許検索 分類	NTCIR-6 QA	-	-	-
NTCIR-7	-	-	NTCIR-7 ACLIA		NTCIR-7 MuST	NTCIR-7 MOAT	NTCIR-7 PATMH	NTCIR-7 PATMT	-	-	-

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
Interactive/Exploratory search
QA types at topic crea

Forum for Researchers and Other Experts/users

Idea Exchange
Discussion/Investigation on Evaluation
methods/metrics

Focus of NTCIR at NTCIR-1

**Lab-type IR
Test, Dataset**

New Challenges

Forum for Researchers

Focus of NTCIR at NTCIR-1

**Lab-type IR
Test, Dataset**

New Challenges

Forum for Researchers

NTCIR-1 (1997-2000): NACSIS Test Collections
for Information Retrieval Systems

Focus of NTCIR , gradually

Lab-type Test,
Dataset

NTCIR-2 to -8 (2009) : NII Test Collections
for Information Retrieval Systems

**New
Challenges**

Forum for Researchers

Focus of NTCIR

Lab-type Test,
Dataset

Forum for Researchers

New Challenges

**NTCIR-9 to present:
NII Testbeds and Community for
Information Access
Technologies**

Information retrieval (IR)

- Retrieve RELEVANT information from vast collection to meet users' information needs
- Using computers since the 1950s
- First CS uses human assessments as success criteria
 - Judgments vary
 - Comparative evaluations on the same infrastructure

Information access (IA)

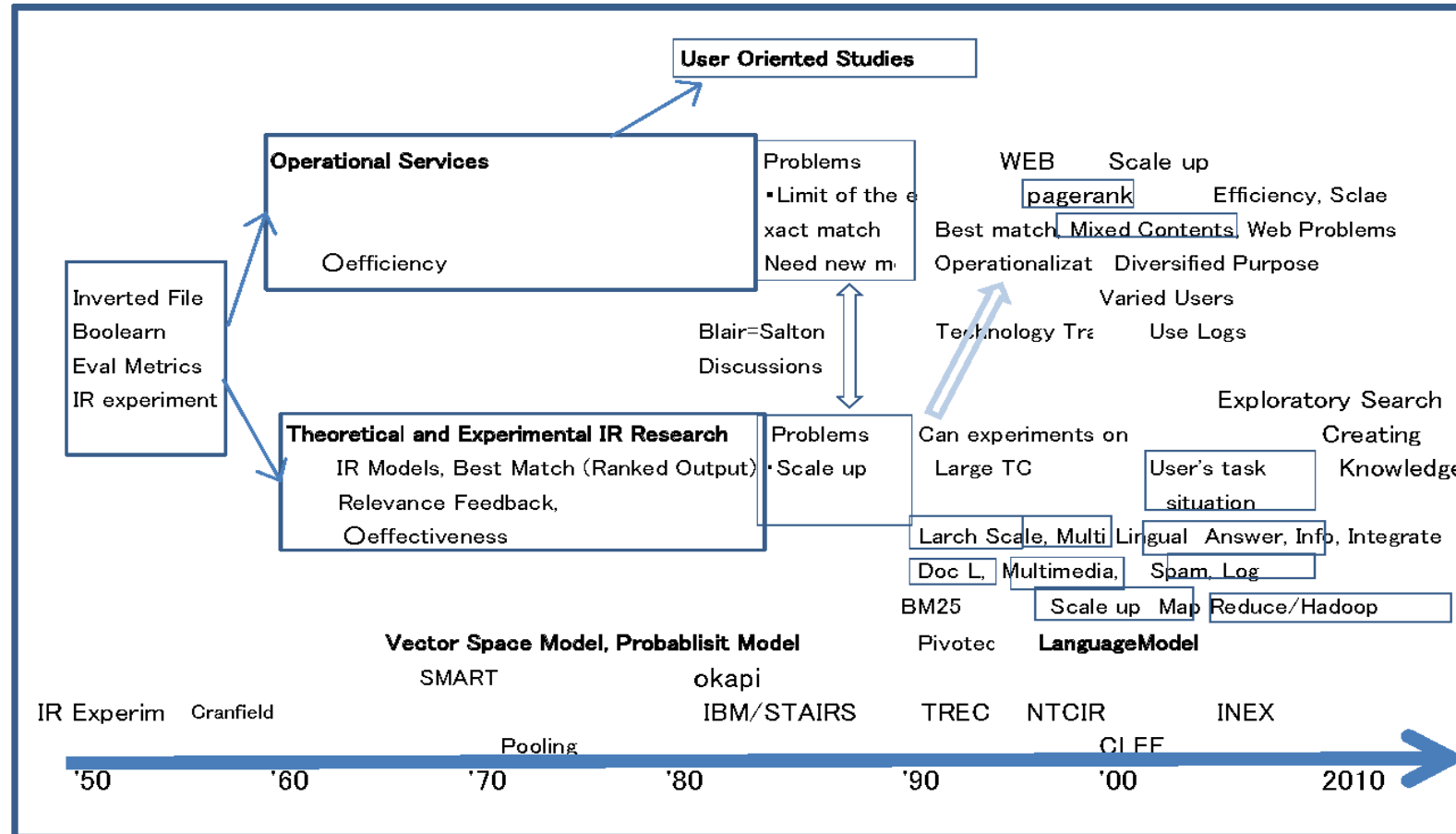
Whole process to make information usable by users.

ex.: IR, text summarization, QA, text mining, and clustering

NTCIR-1: Cross Lingual IR and Terminology on Scientific (J-E Paired) Documents

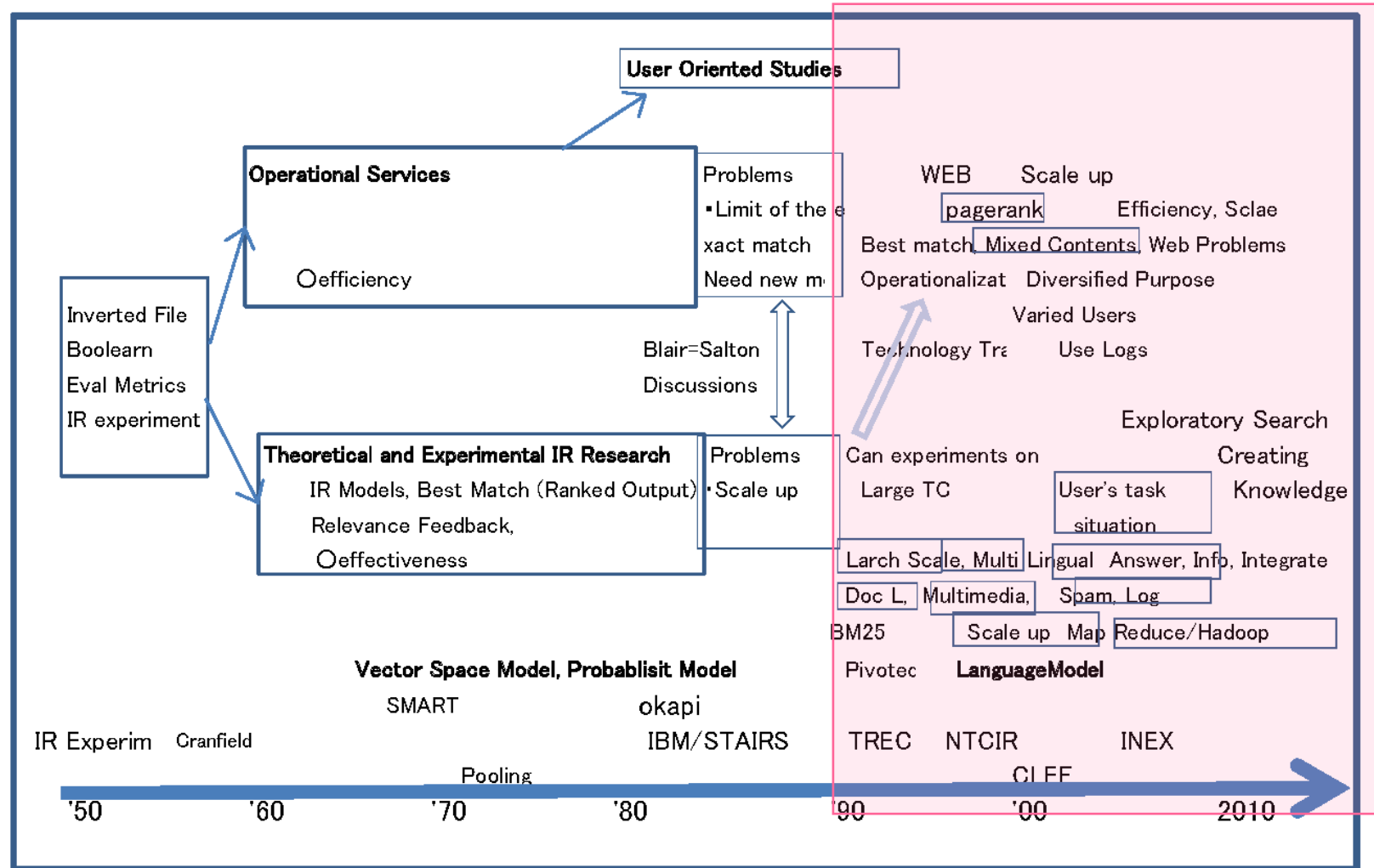
- Social Recognition
- Licences for Documents
- Test Collection
 - Documents, Topics, Relevance Judgments
 - “Topic engineering” and Pooling
 - Relevance Judgements
 - Platform (tool)
 - Multi-Grade Relevance Judgments
 - Assessors, Inter-assessor agreements

IR using Computer Started in 1950s

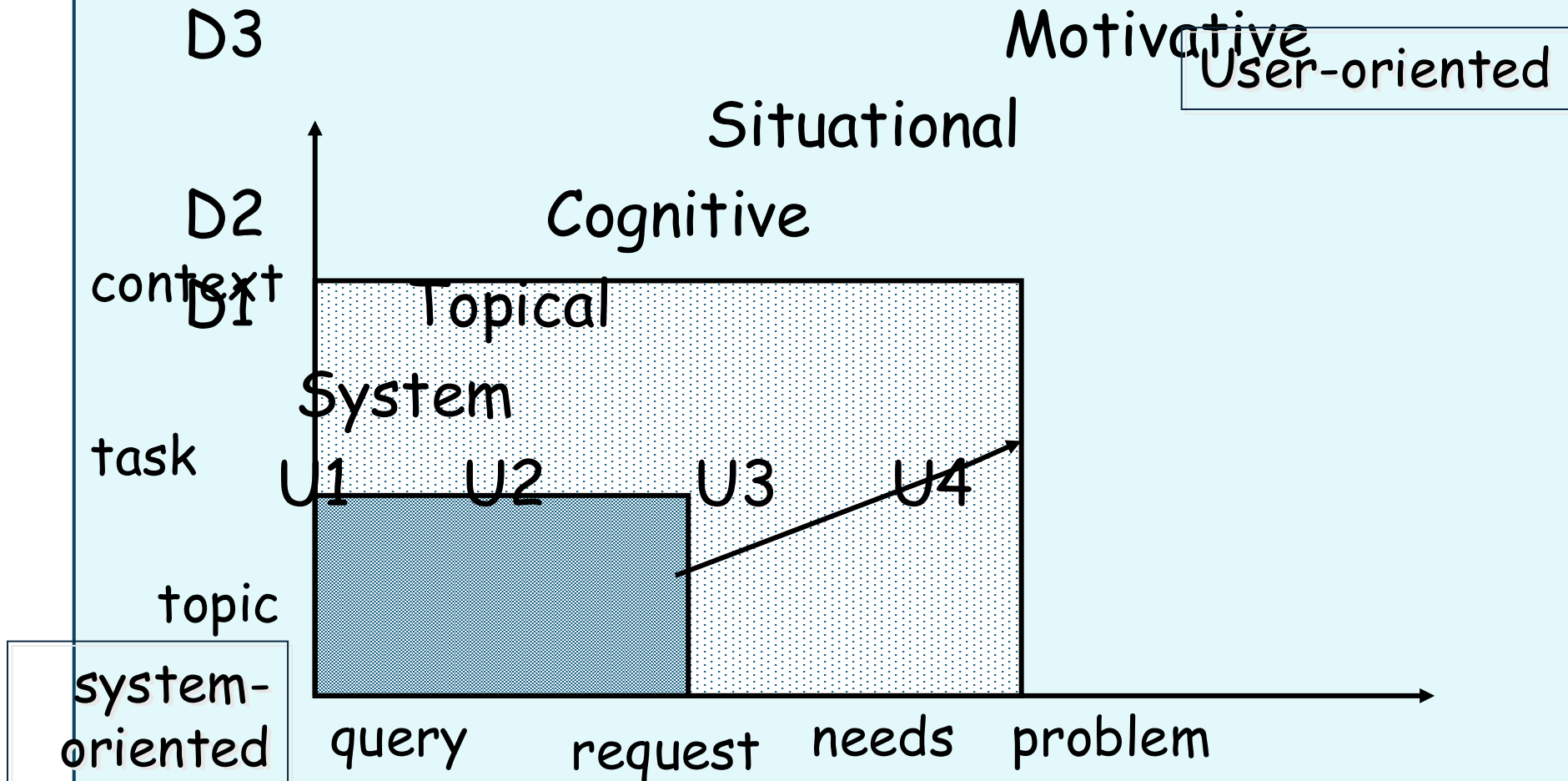


Kando, N. "Information Retrieval – 50 years history" In 50 years of Information Processing Society Japan, translated

After 1992, major research trends have been proposed and led by evaluation forums



Success Criteria - Relevance?

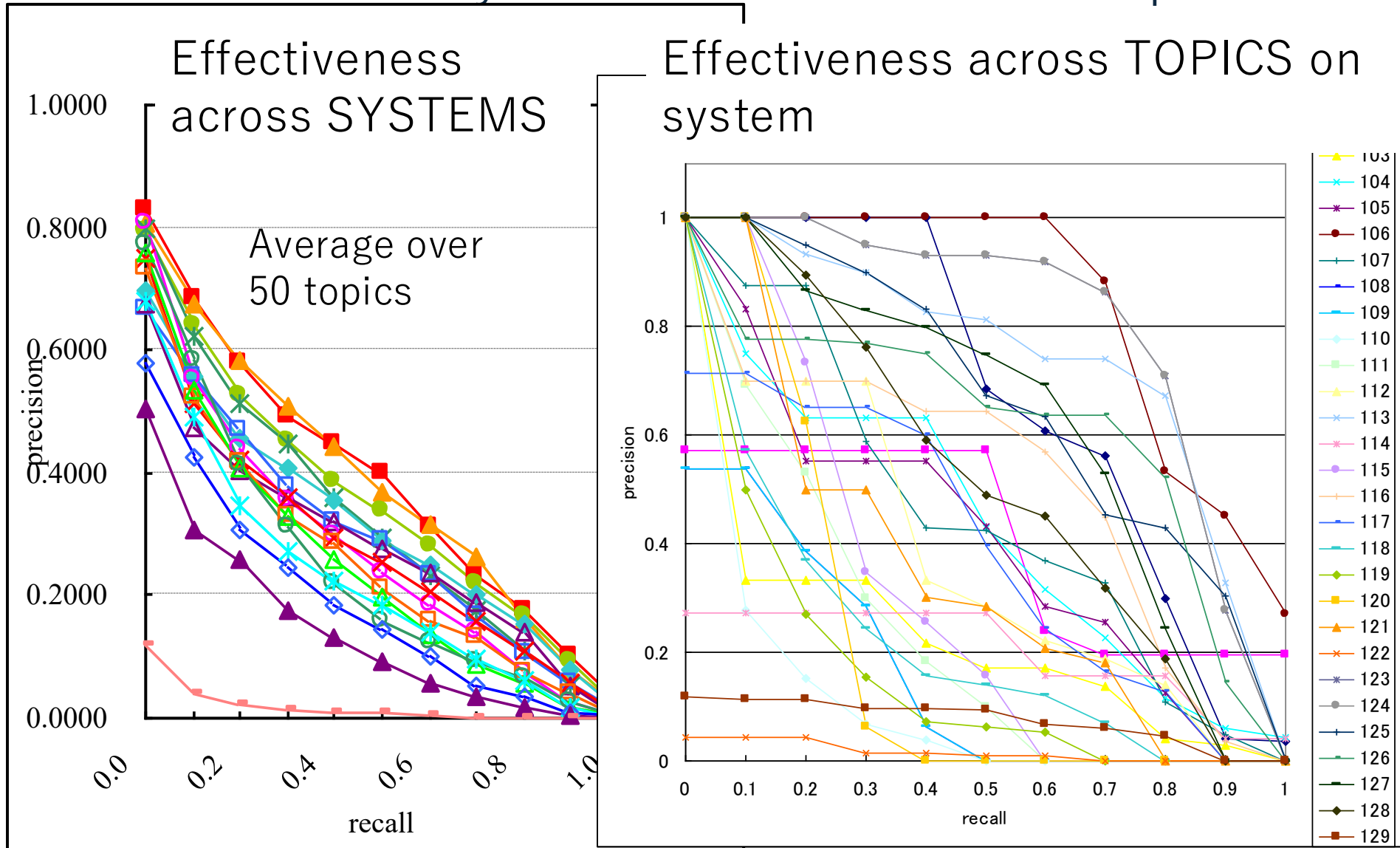


-Saracevic, T. "Relevance reconsidered '96", In *Proc. of CoLIS-2. 1996*, p.201-218

-Mizzaro, S. "Relevance: The whole history" *JASIS Vol.48, No.9*, pp.810-32, 1996

-Kando, N. "Relevance Re-Examined: In the Context of Information Retrieval System Testing", ISLoRWI, 2002

Difficulty of retrieval varies with topics

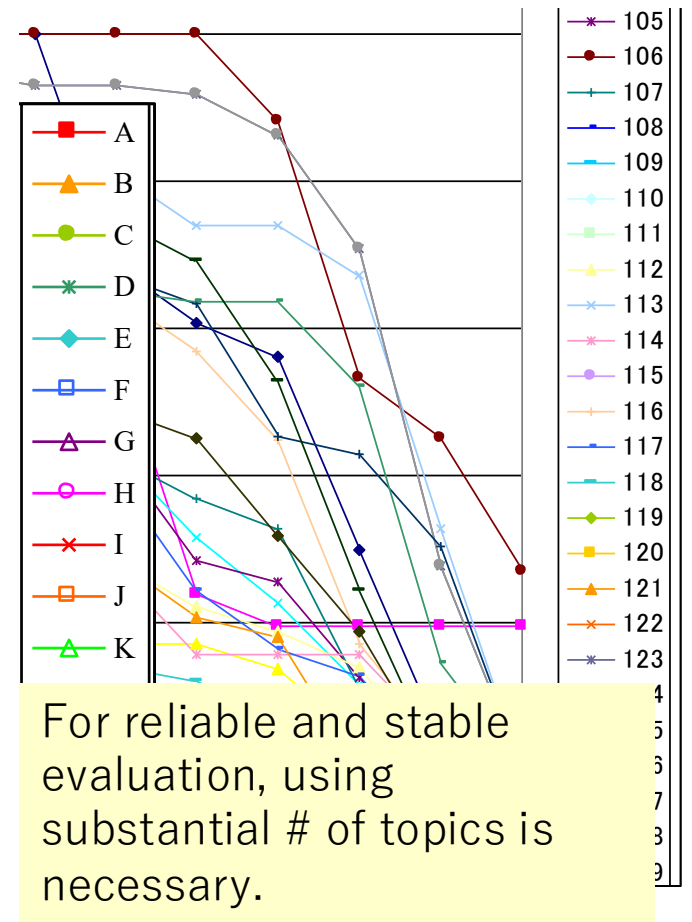
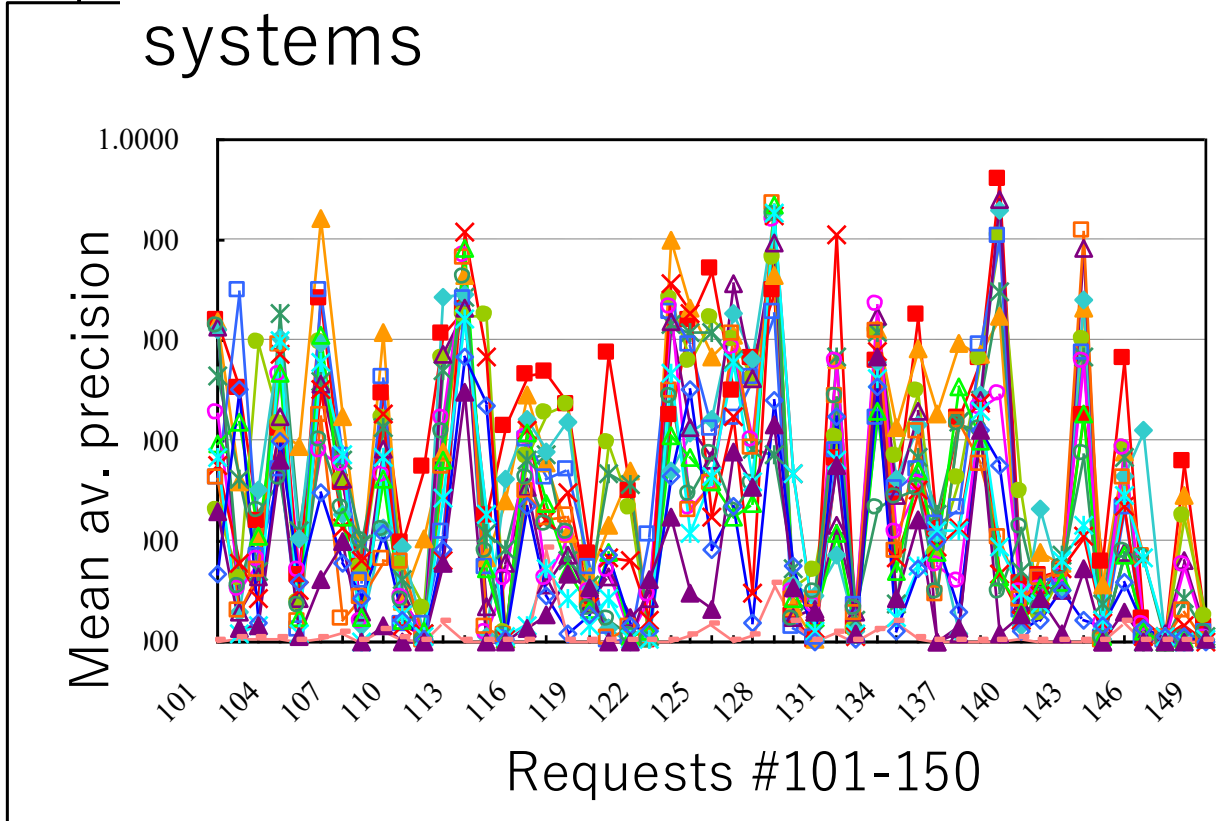


Difficulty of retrieval varies with topics

Effectiveness across SYSTEMS

Effectiveness across TOPICS on system

“Difficult topics” vary with systems

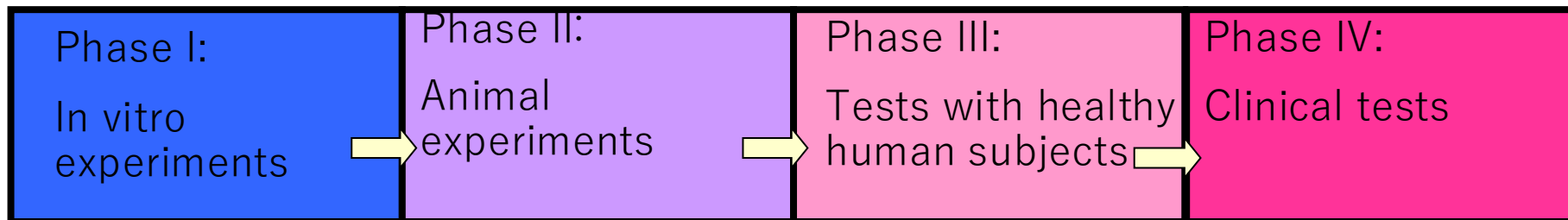


IR Systems Evaluation

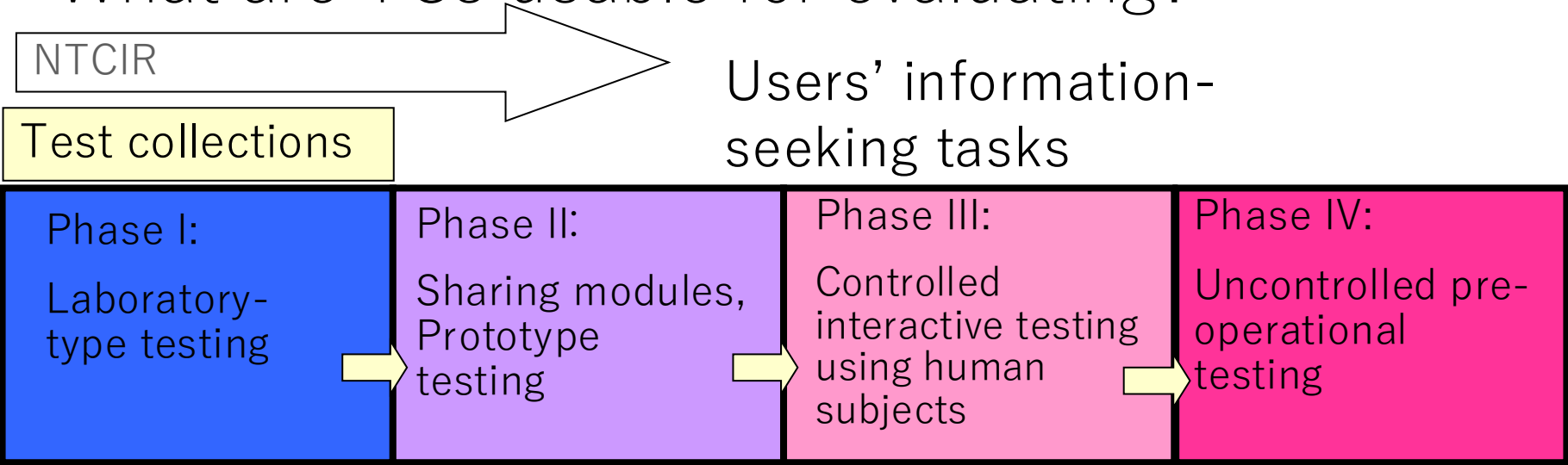
- Engineering Level: Efficiency
- Input Level: ex. Exhaustivity, quality, novelty of DB
- ➔ **Process Level: Effectiveness ex. recall, precision**
- Output Level: Display of output
- User Level: ex. Effort that users need
- Social Level: ex. Importance (Cleverdon & Keen 1966)

What are TCs usable for evaluating?

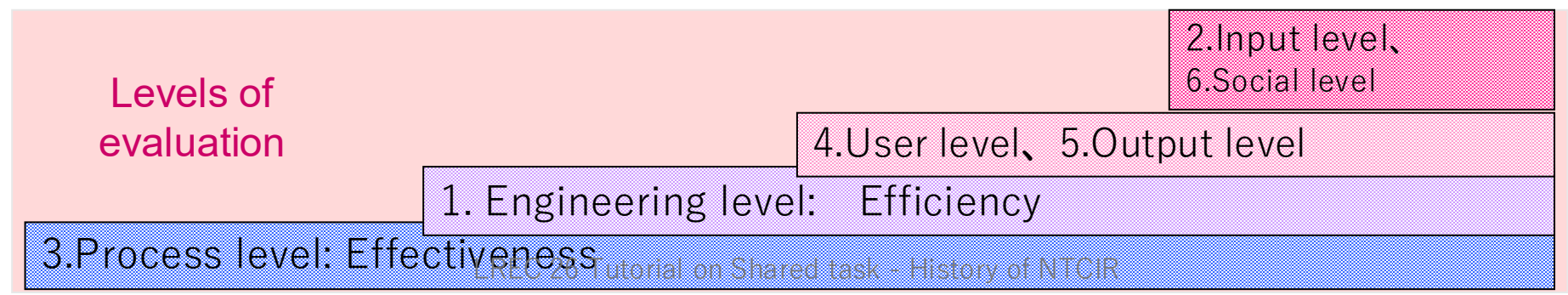
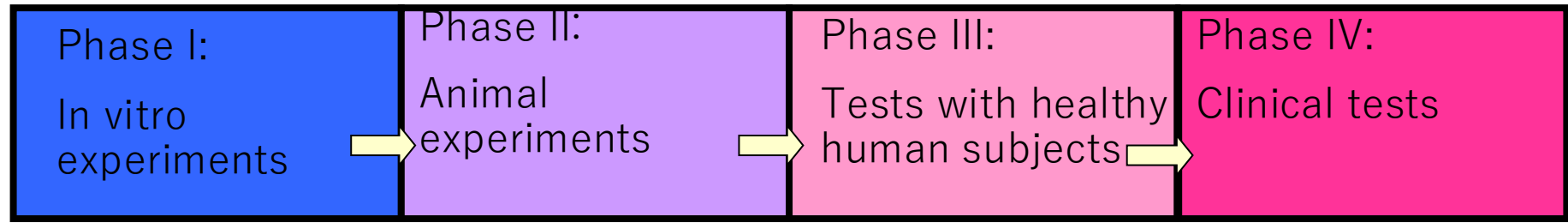
Pharmaceutical R & D



What are TCs usable for evaluating?



Pharmeceutical R & D



NTCIR-2

- J-E retrieval
- Chinese Retrieval (Hsin-His Chen, National Taiwan University)
- Text Summarization (Manabu Okumura, Institute of Science Tokyo)

Evaluation of Summarization before Rouge

- Intrinsic evaluation
 - Quality Questions
 - Edit Distance
- Extrinsic Evaluation
 - .Task-Based
 - IR task
 - QA task, etc

NTCIR-3 to -6

- CLIR - CJKE News, Comparative Corpora
- Summarisation
- QA → IAD
- Patent Retrieval → Classification, Mapping, MT
 - User's task, Briding Communities
- Web
 - Crowling documents, Distiration

- Opinion
- CLQA
- MuSt

NTCIR-7, and -8

- ACLIA: Advanced Cross-Lingual Information Access
 - Module-based Evaluation of IR and QA
- User Generated Contents
 - CQA
 - Opinion
- Knowledge-Intensive
 - Patent Mining
 - Patent MT

We have always faced various challenges

- Technologies and Society have been evolving
- Always new technologies are emerged
- We have enjoyed such challenged and overcome with the power of Community