Overview of Patent Retrieval Task at NTCIR-5

Atsushi Fujii (Univ of Tsukuba)
Makoto Iwayama (Hitachi, Ltd./TITECH)
Noriko Kando (NII)

Background

- Large test collections for Human Language Technology (HLT) have been produced in TREC, CLEF, and NTCIR
  - newspapers, Web, and spoken documents

- Commercial patent retrieval systems have operated for a long time
- Less attention has been paid in the HLT research community

History of patent IR in NTCIR

- NTCIR-3 (2001-2002)
  - technology survey

    First effort was made to produce a large test collection for patent retrieval
- NTCIR-4 (2003-2004)
  - invalidity search
  - automatic patent map generation

Outstanding issues in NTCIR-4

- Invalidity search
  - Document retrieval
    - The number of search topics was small
  - Passage retrieval
    - Evaluation was not performed
- Patent map generation
  - Quantitative evaluation was not performed

NTCIR-5 Patent Retrieval Task

focus of my talk

- Document Retrieval Subtask
  - invalidity search for larger number of topics
- Passage Retrieval Subtask
  - sorting passages in documents retrieved for invalidity search
- Classification Subtask

focus of next talk

  - Classifying patent documents has promise to improve the quality of patent map generation
  - Quantitative evaluation is relatively easy

Document Retrieval Subtask — Invalidity search —

- Find the patents that can invalidate the demand in a patent claim
- This is a patent-to-patent associative 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

Process of the subtask

1. Organizers provided
   - document sets
   - search topics
2. Participating group submitted
   - one or more retrieval results (runs)
   - system description
3. Organizers evaluated each submission

Document sets

- Unexamined patent application
  - Japanese full text published in 1993-2002
  - approx. 3.5 M documents

- Patent Abstracts of Japan (PAJ)
  - English translations of Japanese Abstracts published in 1993-2002
  - training purposes for CLIR
  - optional document set

Search topics

- Japanese patent application rejected by the Japanese Patent Office (JPO)
  - One or more citations (i.e., prior art) exist
  - The first claim was used as a target of invalidation
- To increase the number of topics with minimal cost, only JPO citations were used as relevant documents
  - Additional relevance judgment was not performed

Selection of search topics

From recently filed applications, we selected 1200 applications which satisfy the following criteria as the search topics

- Application was not used as a topic for NTCIR-4
- Texts of both application and citations are in our document collection

Selection of search topics (cont.)

- Citation had been published before application was filed
  - Relevant documents must be prior art
- Application does not claim the priority
  - Otherwise it is difficult to identify the filing date of the application automatically

During the formal run, we identified 11 inappropriate applications

Example search topic

Date of filing (June 29, 1995)

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

<TOPIC>
<NUM>1048</NUM>
<FDATE>19950629</FDATE>
<CLAIM>A milk-derived calcium-containing composition comprising an inorganic salt mainly composed of calcium obtained by baking a milk-derived prepared matter containing milk casein-bonding calcium and/or colloidal calcium.</CLAIM>
...
</TOPIC>

Target of invalidation

Submissions from participants

- 1223 topics were used
  - 34 topics in NTCIR-4
  - 1189 new topics
- For each topic, up to 1000 documents were sorted according to score
  - Mandatory runs must be obtained by only <CLAIM> and <FDATE>
- 84 runs from 10 groups
  - Japan (8), Korea (2)
  - 7 groups submitted mandatory runs

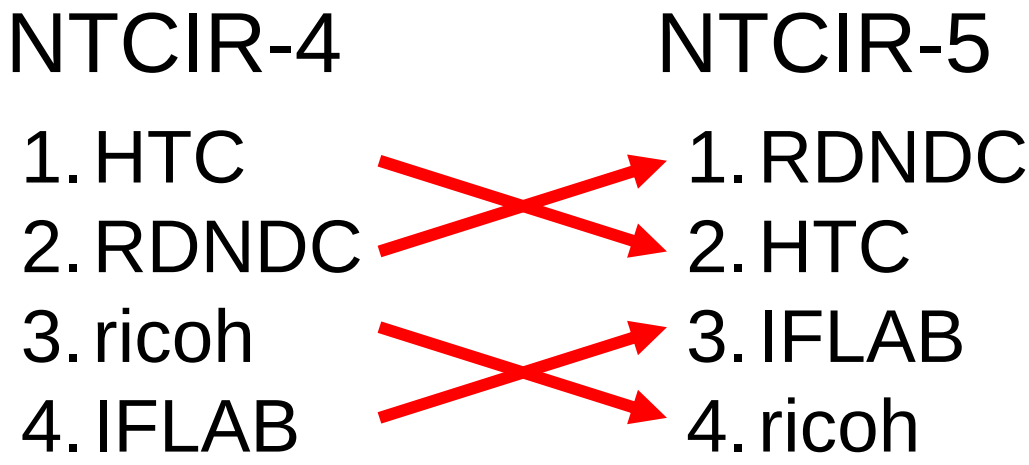
Relevance degree of citations

- The citation used to reject an application was relevant (A)
  - The decision was made confidently
- Multiple citations used together to reject an application were partially relevant (B)
  - Each citation was partially related to the claim in the application

Mean Average Precision (MAP) was used as the evaluation measure

Evaluation results

- No difference in relative superiority between runs depending on the relevance degree (rigid and relaxed)
- Comparing MAP for NTCIR-4 and NTCIR-5 topics, the ranks of top groups changed



Analyzing difficulty of topics

- It can be predicted that invalidity search is easy for the following cases
  - The target application and citation were filed by the same applicant
  - The target application and citation were assigned with the same IPC code

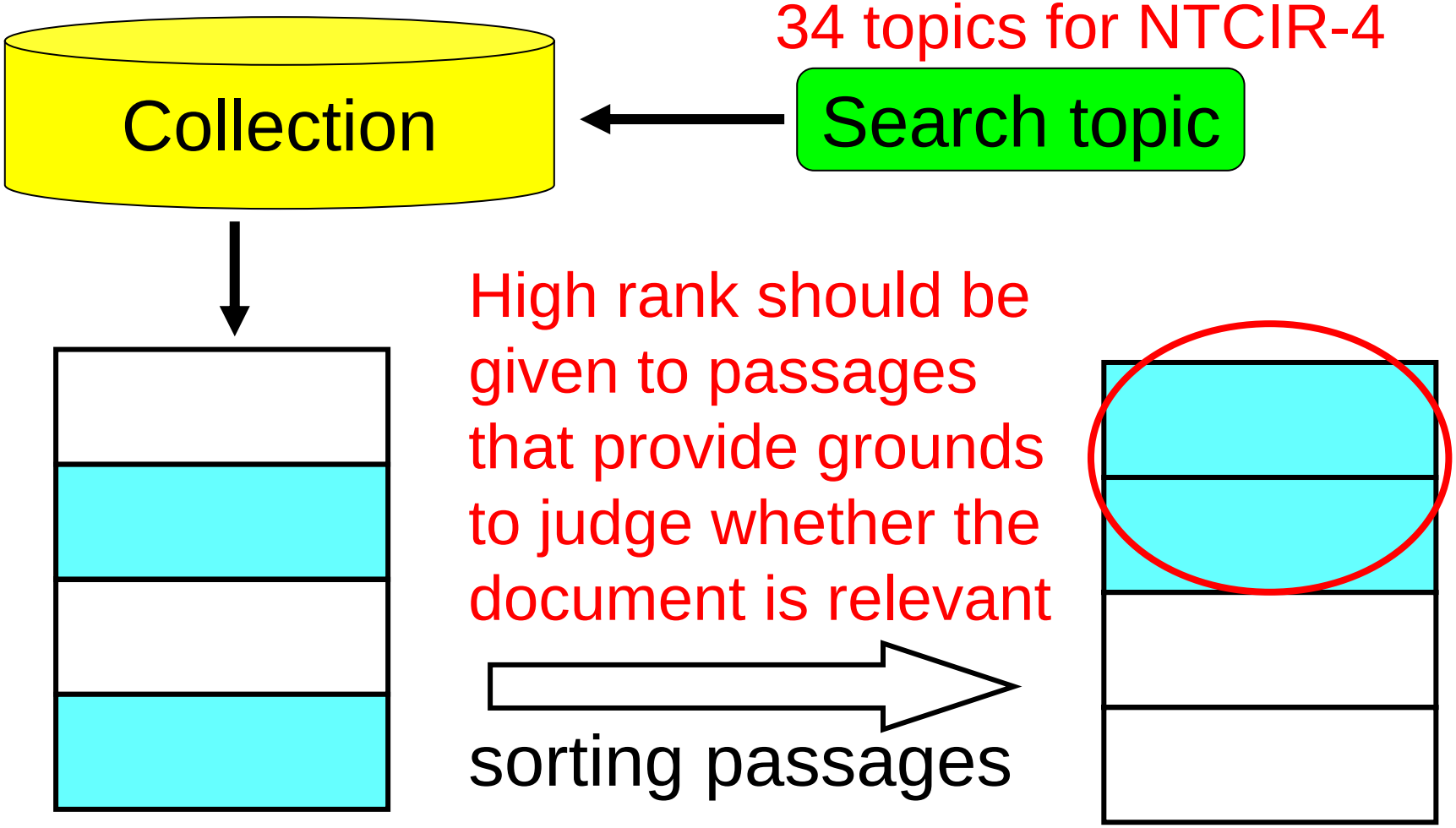
Difficulty of topics (rigid)

Group	ALL	Applicant		IPC	
		same	diff	same	diff
RDNDC	19.49	50.38	13.18	20.30	9.16
HTC	19.44	43.86	14.45	20.33	8.16
IFLAB	19.16	48.08	13.26	20.20	5.96
ricoh	17.66	40.77	12.90	18.98	0
kle	7.86	13.29	6.75	7.85	8.00
JSPAT	6.83	20.20	4.10	7.29	0.93
TUT	3.48	9.57	2.24	3.74	0.25
#Topics	619	105	514	574	45

Summary of Document IR

- Evaluation results were marginally different depending on the search topic
  - NTCIR-4: small number of topics with additional relevance judgment
  - NTCIR-5: large number of topics for which only JPO citations were used as relevant documents
- Difficulty of topics can partially be predicted by Applicant and IPC codes

Passage Retrieval Subtask



Relevance judgment for Passage Retrieval

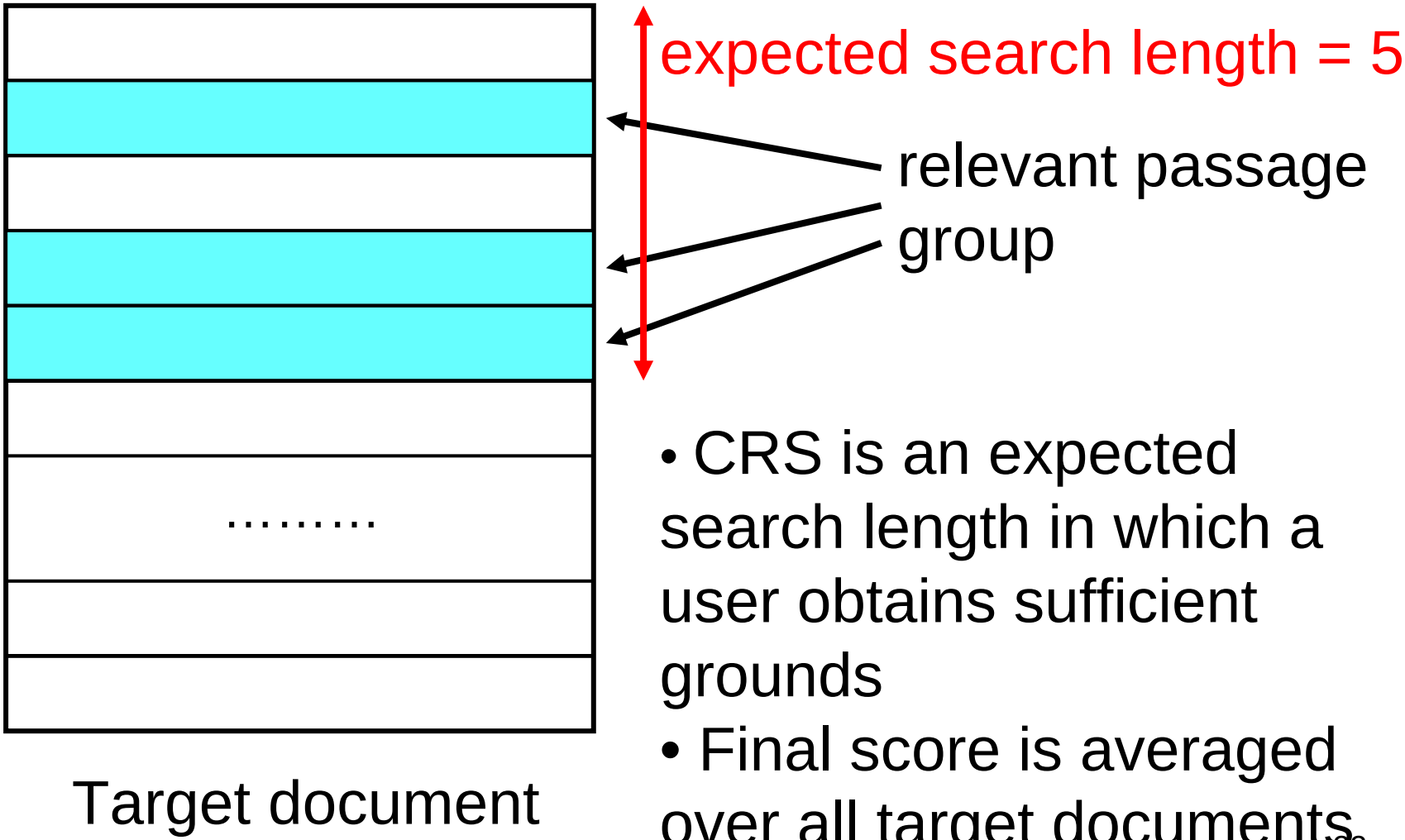
- For each target document, relevant judgment was performed as follows:
  - if a passage can be grounds to judge the document as relevant, this passage is relevant
  - if a group of passages can be grounds to judge the document as relevant, this passage group is relevant
- Assessors searched for relevant passages and groups exhaustively

Evaluation of Passage Retrieval

- A relevant passage group is equally informative as a single relevant passage
- New concept of combinational relevance is proposed

- In the conventional evaluation for IR, relevant items (documents or passages) are independent and combinations are not considered

Combinational Relevance Score (CRS)



- CRS is an expected search length in which a user obtains sufficient grounds
- Final score is averaged over all target documents

Evaluation result

Run ID	CRS	MAP
IFLAB4	10.91	46.14
IFLAB5	11.23	46.36
RDNDCP503	11.67	46.10
RDNDCP507	11.70	47.44
HTC1	12.10	42.31
JSPAT1	12.13	43.52
HTC2	12.14	47.97

Comparing CRS and MAP, relative superiority between runs was different

Summary of Passage IR

- Combinational Relevance Score (CRS) was proposed to evaluate passage retrieval results
- MAP cannot be used with combinational relevance

Conclusion and Future work

- Outstanding issues in NTCIR-4 were resolved

- Plan for NTCIR-6
  - USPTO document collection will be used
  - CLIR, MT, Translation lexicon extraction