

CLIR11

NTCIR-6 CLIR Experiments at Osaka Kyoiku University

— **Term Expansion Using Online Dictionaries and Weighting Score by Term Variety** —

Takashi SATO (Osaka Kyoiku University)

[1] Introduction

- Made **two gram-based indices**, namely indices for 'HEADLINE' and 'TEXT' tag extracted from test collection for **J-J subtask**.
- **Expanded query term** using free online dictionaries in a WEB.
- Probability model were employed for scoring.
- Modified this score multiplying by the number of **varieties of query terms**.

[2] Indexing, Term Extraction and Expansion

- Two indices (HEADLINE and TEXT index) as inverted files of *n*-grams for each of 1st and 2nd stage corpus of J-J subtask.
- Query terms are extracted from TITLE and DESC tag fields in J-J subtask topics.
- Each *compound word* are segmented in words, and all combinations of these words are also made.
- After our submission of runs, we tried to expand terms manually using *definition part of online dictionary* in a WEB (such as Wikipedia and Yahoo dictionaries) .

[3] Ranking using TVF (term variety factor)

- We ranked documents using probabilistic model.
- We prepare another run in which each document score is multiplied by **term variety factor (TVF)** i.e. the number of query term appeared in the document (t_a) divided by the number of query terms for a topic (t_t).
- For example the number of **query terms for a topic is $t_t=5$** and the number of **query terms appeared in a document for the topic is $t_a=3$** out of 5. Then score of the document for the topic is multiplied by **$0.6(=3/5)$** .

[4] Results

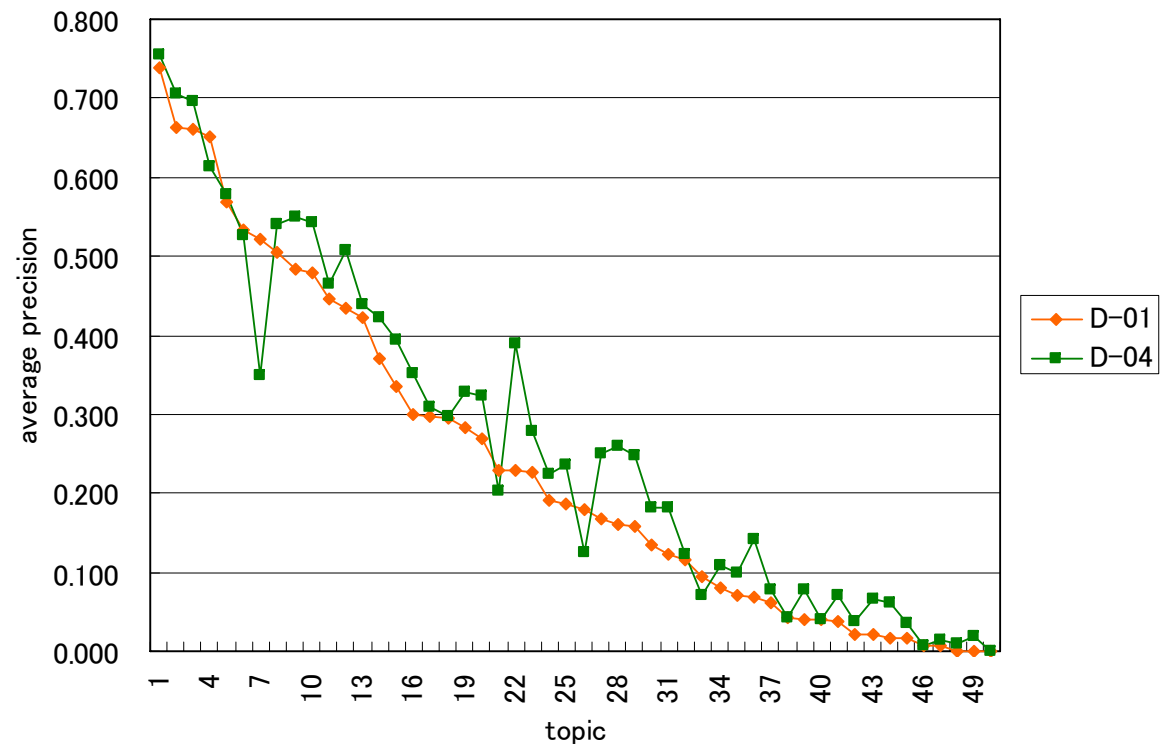
- We submitted 5 runs for 1st stage.
- Table shows the combination of query term set (TITLE and DESC) for HEADLINE and TEXT index.
- Last two runs i.e. D-04 and T-05 are scored multiplying by term variety factor (TVF in the table).

Table 1.
Submitted runs
for 1st stage
(OKSAT, J-J)

run-id	HEADLINE	TEXT	TVF
D-01	DESC	DESC	no
TD-02	TITLE	DESC	no
T-03	TITLE	TITLE	no
D-04	DESC	DESC	<u>yes</u>
T-05	TITLE	TITLE	<u>yes</u>

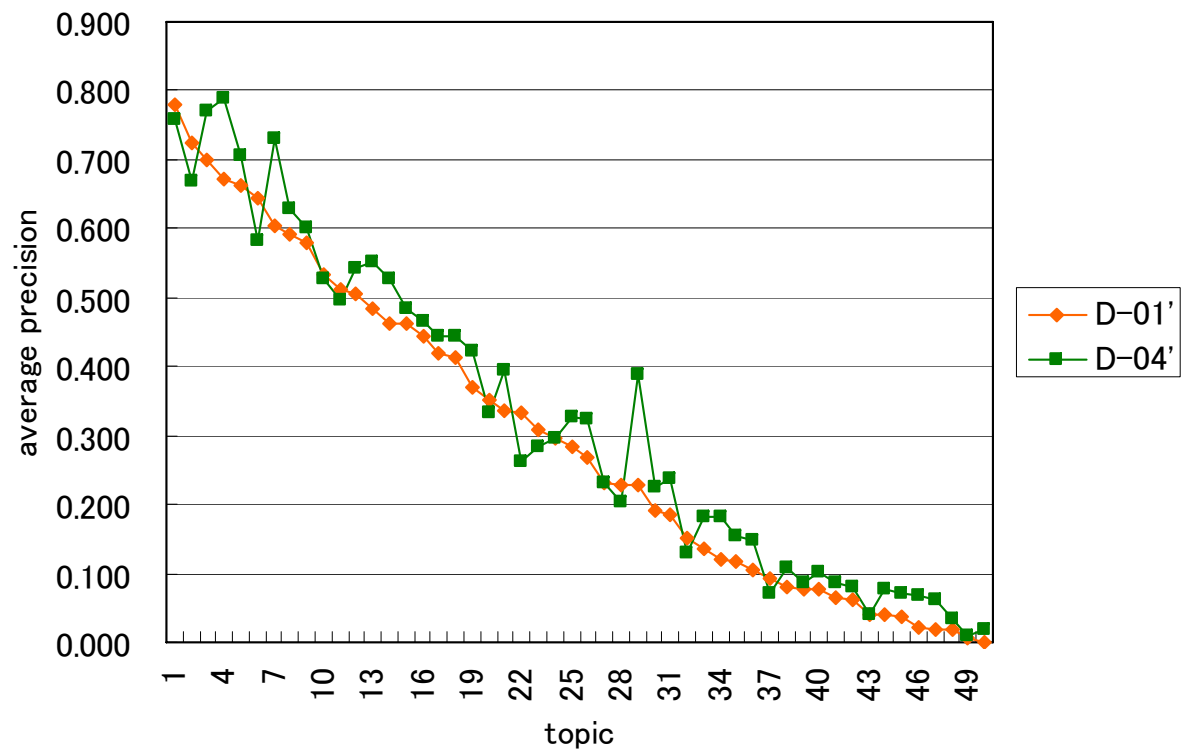
- Figure 1 shows relationship between topic and average precision of relax evaluation of D-01 and D-04.
- In this figure topics are re-ordered by their average precision in descendent order.
- Mean average precision over evaluated topics (MAP) of **D-01 run is 0.240**, and that of **D-04 run is 0.268**.
- In most topics, D-04 (TVF runs) are better then D-01 (normal one).

Figure 1.
Average precision
of submitted run
D-01 and D-04



- We made two post-submission runs **D-01'** and **D-04'**.
- These are term-arranged version of D-01 and D-04. We expanded terms using definition part of online dictionary.
- On the other hand we reduce the number of terms for topics whose average precision of D-04 (TVF version) is lower than that of D-01 (normal version).
- MAP of **D-01'** is **0.302**, and that of **D-04** is **0.327**.

Figure 2.
Average precision
of post-submitted
run D-01' and D-04'



[7] Query Term Reduction

- We reduce the number of terms for topics whose average precision of TVF version (D-04) is lower than that of normal probabilistic version (D-01).
- More concretely, we delete top popular terms in corpus for these topics.

Table 2.
Reduction of
query terms
(topic# 24 and
50)

topic#	query terms	run ID	average precision
24	5	D-01	0.179
		D-04	0.126
	2	D-01'	0.307
		D-04'	0.282
50	16	D-01	0.521
		D-04	<u>0.349</u>
	6	D-01'	0.605
		D-04'	<u>0.732</u>

[8] Conclusions

- We experimented **term expansion using online dictionaries**. It was effective for some topics of which average precision was low.
- We also tried to **weight score by query term variety factor** (TVF). In most cases this worked well.
- **Query term reduction** should be considered if TVF scoring fails.