Selecting Effective Expansion Terms for Diversity

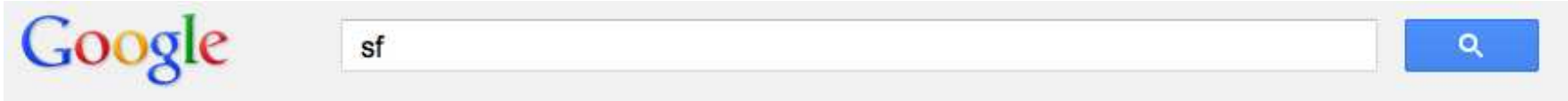
S. Vargas, R.L.T. Santos,
C. Macdonald and I. Ounis

IRGUAM

Terrier Team

University of Glasgow

Query Ambiguity and Underspecification



SF SCHWEIZER FERNSEHEN

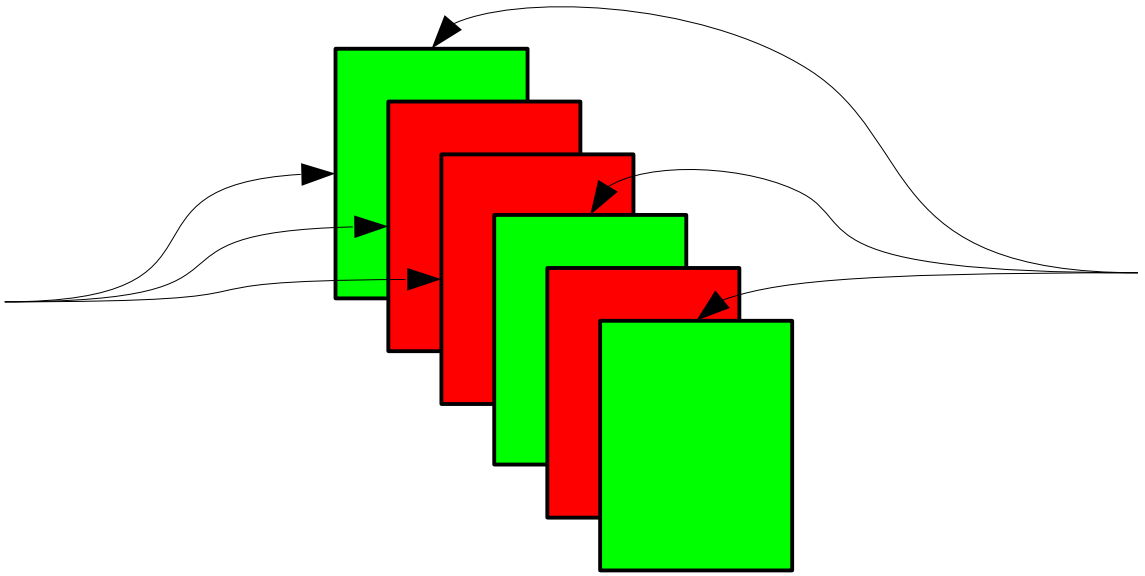
SOURCEFORGE®

Sindicato S.F. Ferroviario
CONFEDERACIÓN INTERSINDICAL

Query Expansion

q = “spiders”

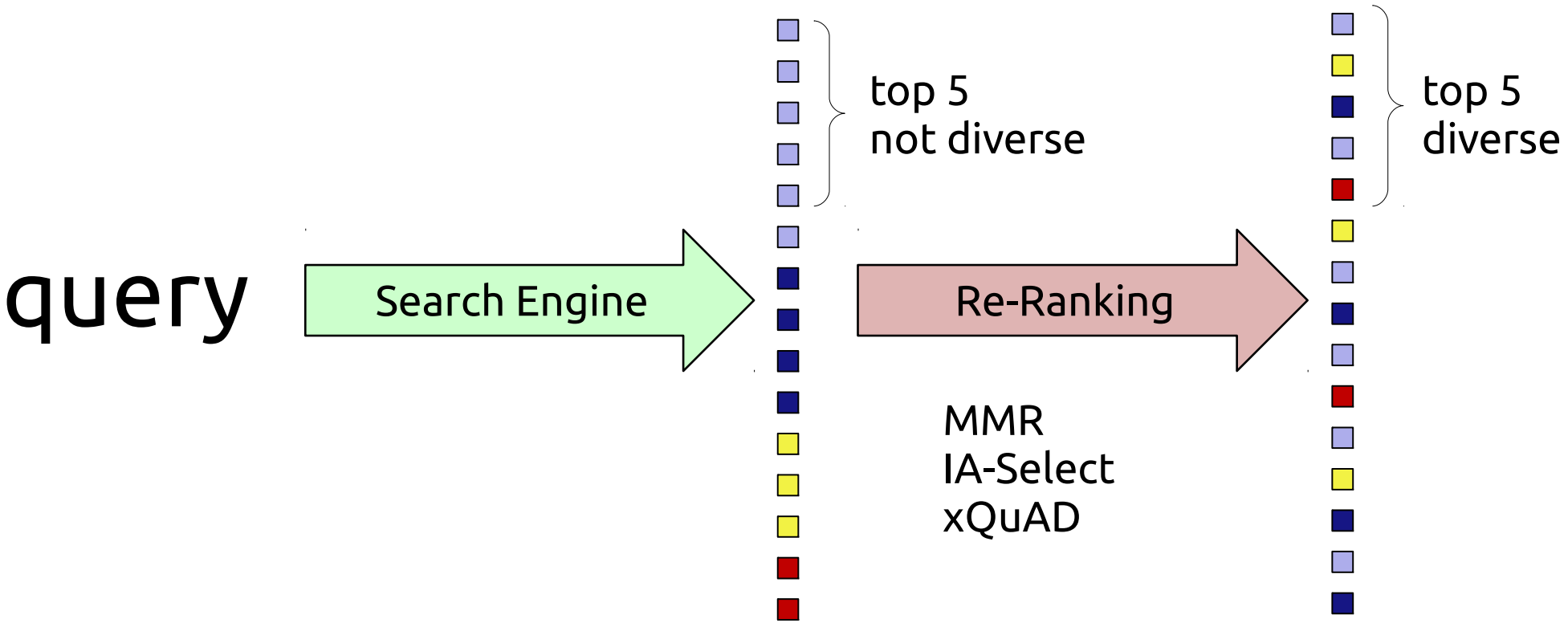
Relevance Feedback



Pseudo Relevance Feedback

q* = “types of spiders in Europe”

Search Result Diversification



Query Expansion and Search Result Diversification?

- Does Query Expansion help retrieving diverse search results?
- If not, can it be adapted to do so?
- Query Expansion can fail for difficult queries.
- Ambiguous queries are difficult!
- In this particular scenario, we identify two problems:
  - Incoherence
  - Bias

Incoherence

- Ambiguous queries result in incoherent feedback sets.
- Query Expansion techniques tend to select terms that are meaningful to the feedback set as a whole.
- This may end up selecting excessively general terms for the expanded query.

Bias (I)

- The feedback set may be biased towards documents covering a single, dominant subtopic.
- Terms important to marginal subtopics may never be selected.
- The retrieval performance may be improved, but the subtopic coverage may be degraded.

Bias (II)

query 79 in TREC 2010 Web Track = “voyager”

S_2

70 relevant documents in ClueWeb09b

S_3

9 relevant documents in ClueWeb09b

Example (III)

- If we use the relevant documents in ClueWeb09 to expand the query:

q^*_{all} = “voyager spacecraft saturn jupiter solar interstellar”

q^*_2 = “voyager spacecraft saturn jupiter solar interstellar”

q^*_3 = “voyager trek maqui borg janeway star uss quadrant”

- Result:

	nrel@20(s_2)	nrel@20(s_3)
q	2	0
q^*_{all}, q^*_2	17	0
q^*_3	1	7

Selection of Expansion Terms (I)

- We propose to identify and select “good” terms.
- The procedure is the following:
  1. Identify groups of documents covering the same subtopic.
  2. Generate a local expanded query for each feedback group.
  3. Select terms from those local expanded queries so that subtopic coverage is maximized with minimum redundancy.

Selection of Expansion Terms (II)

- We adapt the xQuAD algorithm (document selection) to the term selection problem.
- We call it ts_{xQuAD}

$Q = \emptyset$
$T = \bigcup_i \{t \in q_i^*\}$
while $|Q| < \tau$ **do**
  $t^* = \arg\max_{t \in T \setminus Q} \; (1-\lambda)\, \mathrm{P}(t|q) + \lambda\, \mathrm{P}(t, \bar{Q}|q)$
  $T = T \setminus \{t^*\}$
  $Q = Q \cup \{t^*\}$
end while
return q^*_{xQuAD} using the terms in Q

$$\mathrm{P}(t, \bar{Q}|q) = \sum_{q_i^*} \mathrm{P}(q_i^*|q)\, \mathrm{P}(t|q_i^*) \prod_{t' \in Q} \left(1 - \mathrm{P}(t'|q_i^*)\right)$$

Selection of Expansion Terms (III)

- Going back to the “voyager” example:

q^*_{xQuAD}= “voyager trek spacecraft maqui saturn nasa”

- The expanded query contains terms from both subtopics.
- The subtopic coverage is improved:

	nrel@20(s_2)	nrel@20(s_3)
q^*_{xQuAD}	6	4

Research Questions

- **RQ1**: What is the effect of state-of-art query expansion from pseudo-relevance feedback in terms of diversity metrics?
- **RQ2:** How does ts_{xQuAD} perform in terms of ad-hoc retrieval and diversity compared to existing query expansion approaches?

Experimental Setup

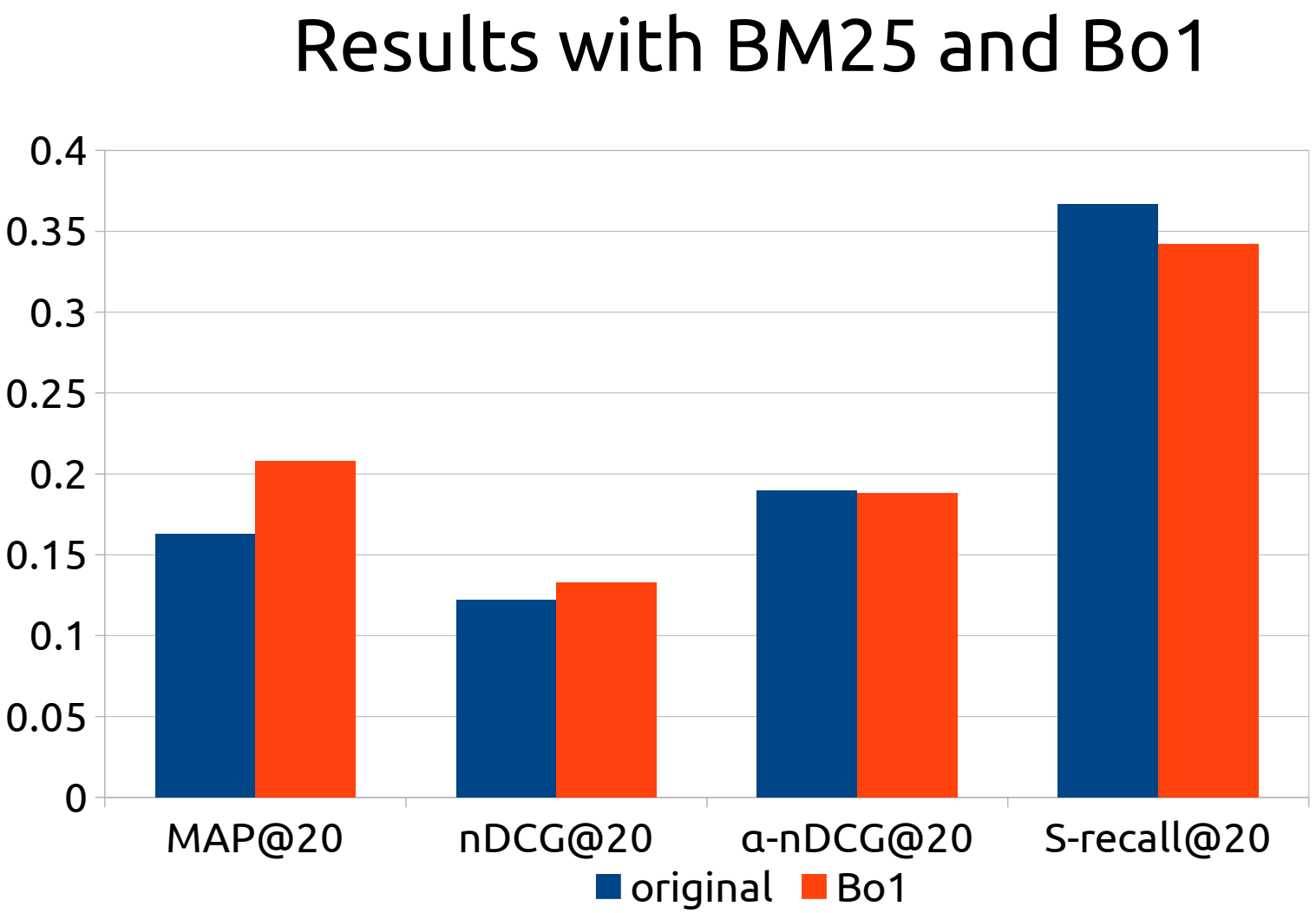
- Context: diversity task of the TREC 2009, 2010 and 2011 Web Tracks.
  - Corpus: ClueWeb09 Category B.
  - 150 queries with 3 to 8 subtopics.
- Terrier for indexing and retrieval:
  - Retrieval models: BM25, DPH, TF-IDF, PL2.
  - Query Expansion techniques: Bo1, Bo2 and KL.
  - Ad-hoc metrics: MAP, nDCG.
  - Diversity metrics: α-nDCG, ERR-IA, S-recall.

RQ1: Experiment

What is the effect of state-of-art query expansion from PRF in terms of diversity metrics?

- We evaluate query expansion techniques in a pseudo-relevance feedback setting.
- We expand queries using the first 5 and 10 retrieved documents from the original query.

RQ1: Results


Results with BM25 and Bo1
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
MAP@20
nDCG@20
α-nDCG@20
S-recall@20
original
Bo1

RQ2: Experiment (I)

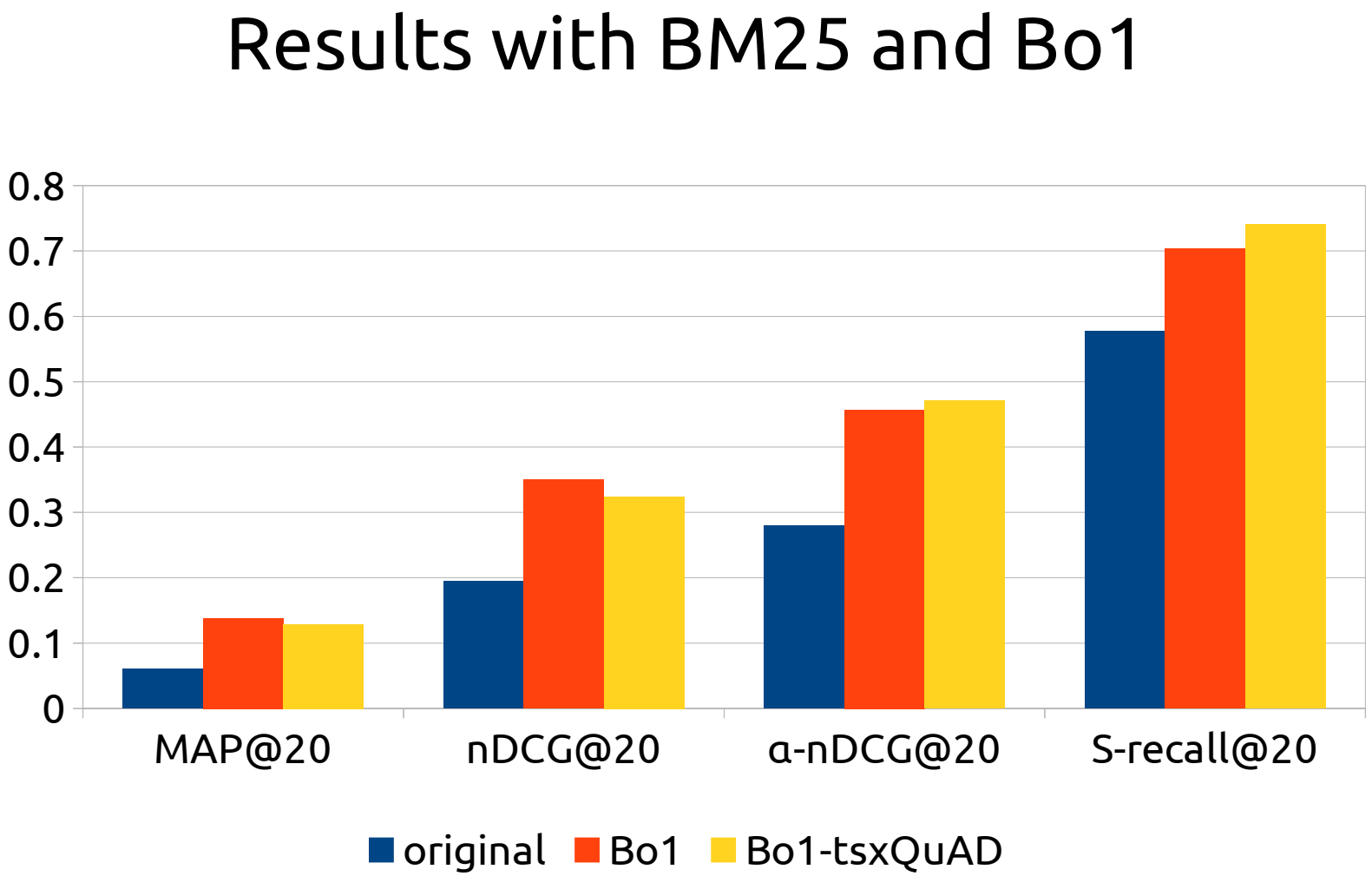
How does ts$_{xQuAD}$ perform in terms of ad-hoc retrieval and diversity compared to existing query expansion approaches?

- We consider a relevance feedback setting where feedback from the assessors for a given query is used to generate an expanded query.
- We simulate a situation where users provide feedback for their interpretation of the query:
  - The problem lies in the combination of different sources referring to possibly more than one subtopic.
  - We assume that there is complete information about the subtopics each document covers

RQ2: Experiment (II)

- We compare our proposed ts_{xQuAD} (λ=1.0) built on top of different retrieval and query expansion models with their standard variants.
- Feedback documents are extracted from the TREC relevance judgments with the following constraints:
  - Residual evaluation method.
  - Similar number of documents for each subtopic in feedback and evaluation.
  - We chose subtopics with, at least, 6 relevant documents.

RQ2: Results


Results with BM25 and Bo1
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
MAP@20
nDCG@20
α-nDCG@20
S-recall@20
original
Bo1
Bo1-tsxQuAD

Conclusions

- We have analyzed the suitability of query expansion techniques for search result diversification.
- We have proposed a term selection strategy to improve the diversity of expanded queries.
- A thorough evaluation shows that it improves the diversity of the search results at a negligible cost in terms of ad-hoc relevance.
- Future work: apply ts_{xQuAD} to the pseudo-relevance feedback scenario.

Thanks for you attention!
Questions?

RQ1: Results

		MAP×10		nDCG		α-nDCG		ERR-IA		S-recall	
		10	20	10	20	10	20	10	20	10	20
DPH		0.123	0.200	0.144	0.141	0.193	0.221	0.139	0.147	0.331	0.402
+Bo1	**5**	0.142▲	0.231	0.145▲	0.143▲	0.183▼	0.209	0.141	0.149	0.268▼	0.347▼
	10	0.145▲	0.238	0.152▲	0.149	0.197	0.219	0.154	0.160	0.290▼	0.347▼
+Bo2	**5**	0.123▲	0.201	0.129▼	0.132▼	0.169▼	0.196▼	0.123▼	0.131▼	0.275▼	0.353▼
	10	0.129▲	0.209	0.145▲	0.141	0.181▼	0.207▼	0.133▼	0.141▼	0.299▼	0.368▼
+KL	**5**	0.140▲	0.226	0.144▲	0.141	0.184▼	0.208	0.141▲	0.148	0.284▼	0.347▼
	10	0.155	0.243	0.155▲	0.151	0.192▼	0.214	0.148▲	0.154	0.294▼	0.344▼
BM25		0.092	0.163	0.119	0.122	0.159	0.190	0.115	0.123	0.271	0.367
+Bo1	**5**	0.102▲	0.180	0.115▼	0.123▲	0.159▼	0.188	0.121	0.129	0.253▼	0.340▼
	10	0.122▲	0.208	0.129▲	0.133	0.158▼	0.188	0.115	0.123	0.263▼	0.342▼
+Bo2	**5**	0.111▲	0.196	0.115▼	0.123	0.155▼	0.177▼	0.111▼	0.117▼	0.265▼	0.320▼
	10	0.115▲	0.198	0.117▼	0.122	0.144▼	0.170▼	0.101▼	0.108▼	0.244▼	0.326▼
+KL	**5**	0.110▲	0.191	0.118▼	0.127	0.157	0.186	0.114	0.122	0.261▼	0.348▼
	10	0.124▲	0.202	0.131▲	0.133	0.159▼	0.187	0.114▼	0.122	0.268▼	0.342▼

RQ2: Results

		MAP		nDCG		α-nDCG		ERR-IA		S-recall	
		10	20	10	20	10	20	10	20	10	20
DPH		0.057	0.084	0.269	0.256	0.328	0.365	0.257	0.268	0.566	0.658
+xQuAD		0.071	0.104	0.314	0.295	0.382	0.420	0.308	0.319	0.610	0.711
+Bo1	s	0.115	0.161	0.438	0.398	0.479	0.514	0.400	0.410	0.673	0.759
	d	0.117	0.160	0.432	0.387▼	0.506▲	0.538▲	0.424▲	0.433▲	0.714▲	0.792▲
+Bo2	s	0.107	0.148	0.406	0.367	0.456	0.489	0.381	0.391	0.646	0.732
	d	0.105	0.142▼	0.390▼	0.352▼	0.478▲	0.514▲	0.400▲	0.411▲	0.688▲	0.780▲
+KL	s	0.115	0.159	0.433	0.392	0.476	0.509	0.397	0.407	0.672	0.760
	d	0.111	0.152▼	0.408▼	0.371▼	0.490▲	0.525▲	0.408	0.418▲	0.714▲	0.795▲
BM25		0.040	0.061	0.195	0.195	0.233	0.279	0.177	0.190	0.427	0.577
+xQuAD		0.054	0.079	0.233	0.232	0.298	0.343	0.236	0.249	0.515	0.637
+Bo1	s	0.098	0.138	0.378	0.350	0.423	0.457	0.346	0.356	0.616	0.703
	d	0.095	0.129▼	0.355▼	0.324▼	0.439▲	0.471▲	0.363▲	0.372▲	0.657▲	0.740▲
+Bo2	s	0.098	0.135	0.371	0.338	0.421	0.453	0.351	0.360	0.598	0.689
	d	0.098	0.132	0.359	0.327▼	0.447▲	0.479▲	0.370▲	0.379▲	0.661▲	0.742▲
+KL	s	0.097	0.138	0.375	0.349	0.417	0.452	0.336	0.346	0.625	0.710
	d	0.095	0.130▼	0.353▼	0.323▼	0.434▲	0.467▲	0.354▲	0.364▲	0.659▲	0.742▲