University of Glasgow at the NTCIR-9 Intent task

Experiments with Terrier on Subtopic Mining and Document Ranking

Rodrygo L. T. Santos
rodrygo@dcs.gla.ac.uk

Craig Macdonald
craigm@dcs.gla.ac.uk

Iadh Ounis
ounis@dcs.gla.ac.uk

School of Computing Science
University of Glasgow
G12 8QQ, Glasgow, UK

ABSTRACT

We describe our participation in the subtopic mining and document ranking subtasks of the NTCIR-9 Intent task, for both Chinese and Japanese languages. In the subtopic mining subtask, we experiment with a novel data-driven approach for ranking reformulations of an ambiguous query. In the document ranking subtask, we deploy our state-of-the-art xQuAD framework for search result diversification.

Team Name

uogTr

Subtasks/Languages

Chinese and Japanese Subtopic Mining
Chinese and Japanese Document Ranking

External Resources Used

Bing related searches

1. INTRODUCTION

In NTCIR-9, we participate in both subtasks of the Intent task [11], namely, subtopic mining and document ranking. In the subtopic mining subtask, the goal is to retrieve relevant aspects (or subtopics) for an ambiguous query. For such, we experiment with a novel data-driven approach for ranking reformulations with respect to a given query.

In the document ranking subtask, the goal is to produce a ranking with maximum coverage and minimum redundancy with respect to the relevant aspects of a given query. To this end, we experiment with our state-of-the-art xQuAD framework [5] for search result diversification. In particular, we build upon our previously successful participations in the diversity task of the TREC 2009 and 2010 Web tracks [2, 8], by leveraging learned models for an ambiguous query and its multiple aspects [7, 8], as well as a selective diversification regime for queries with different levels of ambiguity [6].

In the remainder of this paper, Section 2 describes our infrastructure for processing Chinese and Japanese text. Sections 3 and 4 describe our approaches for the subtopic mining and document ranking subtasks, respectively. Finally, Section 5 presents our concluding remarks.

2. TEXT PROCESSING

Our participation in the NTCIR-9 Intent task builds upon the Terrier Information Retrieval platform [4]. In order to process Chinese and Japanese text, we have enhanced Terrier's multilingual support, with a neat separation between text parsing and tokenisation. In particular, we have introduced a new interface for different tokeniser implementations that is now part of Terrier 3.5.[1] For our participation, we have integrated a Chinese tokeniser based on hidden Markov models.[2] For the Japanese language, we have integrated a simple character-based tokenisation scheme.[3]

Using the new tokenisation interface in Terrier and the integrated tokenisers, we have indexed the SogouT Chinese corpus and the Japanese subset of the ClueWeb09 corpus. For both corpora, we have indexed multiple fields, namely, title, body, and the anchor-text of incoming hyperlinks. The statistics of the indexed corpora are shown in Table 1.

	SogouT	ClueWeb09-JA
#documents	128,677,943	67,337,717
#tokens		
title	4,032,694,914	841,036,389
body	107,657,422,738	62,782,716,375
anchor	13,750,661,867	1,842,553,273

Table 1: Collection statistics.

3. SUBTOPIC MINING

In the subtopic mining subtask, we introduce a data-driven approach to mine the most likely aspects underlying a query from a usage log. In particular, for the Chinese subtask, we use the provided SogouQ query log,[4] comprising usage data gathered from the Sogou search engine in June 2008. For the Japanese subtask, we use the anchor-text extracted from the Japanese portion of the ClueWeb09 corpus as a query log surrogate [1], as no actual query log was provided.

Given a usage log $\mathcal{L}$ (e.g., a query log or anchor log), our approach identifies candidate aspects a for a query q that co-occur frequently with q or another query q' similar to q. The latter relaxation is introduced to handle the case where no exact match for q is found in the log $\mathcal{L}$. For the Chinese query log, the likelihood of an aspect a given the query q is estimated as:

$$\mathrm{P}(a|q) = \frac{1}{|\mathcal{L}|} \sum_{q' \in \mathcal{L}} \mathrm{WS}(q, q') \, \frac{\mathrm{SS}(q', a) + \mathrm{CS}(q', a)}{2}, \qquad (1)$$

[1] http://terrier.org
[2] http://ictclas.org
[3] http://chasen.org/~taku/software/TinySegmenter
[4] http://www.sogou.com/labs/dl/q-e.html

Run	$\mathcal{A}$	P(a|q)	I-rec @10	I-rec @20	I-rec @30	D-nDCG @10	D-nDCG @20	D-nDCG @30	D#-nDCG @10	D#-nDCG @20	D#-nDCG @30
uogTr-S-C-3	SQ	Eq. (1)	0.1682	0.2245	0.2422	0.1698	0.1796	0.1650	0.1690	0.2020	0.2036
uogTr-S-C-1	SQ	Eq. (3)	0.3210▲	0.4187▲	0.4533▲	0.3385▲	0.3670▲	0.3386▲	0.3297▲	0.3929▲	0.3960▲
uogTr-S-C-2	SQ	Tab. 3: SS	0.1753	0.2407	0.2640	0.1772	0.1691	0.1470	0.1763	0.2049	0.2055
uogTr-S-C-4	SQ	Eq. (6)	0.3176▲	0.4170▲	0.4533▲	0.3364▲	0.3662▲	0.3386▲	0.3270▲	0.3916▲	0.3960▲
uogTr-S-C-5	BR	Eq. (4)	0.4947▲	0.4947▲	0.4947▲	0.6598▲	0.4278▲	0.3309▲	0.5772▲	0.4613▲	0.4128▲
uogTr-S-J-1	AT	Eq. (2)	0.0113	0.0270	0.0270	0.0045	0.0077	0.0081	0.0079	0.0173	0.0175
uogTr-S-J-2	BR	Eq. (4)	0.4321▲	0.4321▲	0.4321▲	0.4071▲	0.2979▲	0.2629▲	0.4196▲	0.3650▲	0.3475▲

Table 2: Subtopic mining results. SQ, AT, and BR denote query aspects derived from the SogouQ query log, the ClueWeb09-JA anchor-text, and using the Bing related searches API, respectively.

Feat.	Description	Formula
QF	query frequency	$\frac{\|\{q' \in \mathcal{L} \| q' = a\}\|}{\|\mathcal{L}\|}$
KS	character-level similarity	$levenshtein(q, a)$
WS	word-level similarity	$\frac{\|terms(q, a)\|}{\|terms(q)\|}$
CS	click similarity	$\frac{\|clicks(q, a)\|}{\|clicks(q)\|}$
SS	session similarity	$\frac{\|sessions(q, a)\|}{\|sessions(q)\|}$

Table 3: Subtopic ranking features [10].

where $\mathrm{WS}(q, q')$ represents the fraction of words in q present in $q' \in \mathcal{L}$, $\mathrm{SS}(q', a)$ represents the probability of q' and a co-occurring in the same session, and $\mathrm{CS}(q', a)$ is the probability of q' and a leading to clicks on the same URL.

For the Japanese anchor-text log, the likelihood of an aspect a given the query q is estimated as:

$$\mathrm{P}(a|q) = \frac{1}{|\mathcal{L}|} \sum_{q' \in \mathcal{L}} \mathrm{WS}(q, q')\, \mathrm{CS}(q', a), \tag{2}$$

where $\mathrm{WS}(q, q')$ is as before and $\mathrm{CS}(q', a)$ is now the probability of q' and a being anchors for the same URL (i.e., the equivalent of clicks in an anchor-text log).

Besides directly considering the aspects mined from a query or anchor log as potential subtopic mining runs, we experiment with re-ranking the aspects mined from the Chinese query log with respect to multiple features, according to:

$$\mathrm{P}(a|q) = \sum_i w_i f_i(q, a), \tag{3}$$

where the features f_i are described in Table 3. In order to appropriately estimate the weight w_i for each of these features, we deploy a listwise learning-to-rank approach [3]. As training examples, we consider pairs $\langle q, \mathcal{Q} \rangle$, where q is a randomly selected query from SogouQ, and $\mathcal{Q}$ is a set of 'ideal' query aspects (also present in SogouQ), provided as related searches by the Bing API.[5] In total, we obtain 680 such training queries, with an average of 8.1 aspects each.

On top of this learned ranking, we perform a click-driven diversification of query aspects. Our intuition is that similar aspects will lead to clicks on similar URLs. Therefore, in order to maximise the coverage of clicked URLs among the selected aspects, we explicitly diversify these aspects with respect to their clicked URLs, by adapting our approach for document ranking diversification, described in Section 4.

[5] `http://msdn.microsoft.com/en-us/library/dd900818.aspx`

Finally, we also consider related searches provided by the Bing API as an additional run for the Chinese and Japanese subtopic mining subtasks. Note that performance figures for these runs are only meaningful for rank cutoffs smaller than or equal to 13, which is the maximum number of related searches provided by Bing for any given query. Additionally, since the provided related searches are not weighted, we employ a simple rank-based weighting function:

$$\mathrm{P}(a|q) = |\mathcal{A}| - rank(q, a), \tag{4}$$

where $\mathcal{A}$ is the list of related searches provided by Bing for the query q, and $rank(q, a)$ is the position of a in this list.

3.1 Subtopic Mining Runs

In total, we submitted five runs to the Chinese subtopic mining subtask:

- uogTr-S-C-3 comprises the aspects mined from SogouQ and ranked according to Eq. (1);
- uogTr-S-C-1 re-ranks the aspects in uogTr-S-C-3 using our data-driven approach (Eq. (3));
- uogTr-S-C-2 re-ranks the aspects in uogTr-S-C-3 by their likelihood of co-occurring with the initial query in a session (SS in Table 3);
- uogTr-S-C-4 diversifies the aspects in uogTr-S-C-1 according to their received clicks (Eq. (5));
- uogTr-S-C-5 comprises related searches obtained using the Bing API and ranked by Eq. (4).

In addition, we submitted two Japanese runs:

- uogTr-S-J-1 comprises anchor-text obtained from the Japanese portion of ClueWeb09 and ranked by Eq. (2);
- uogTr-S-J-2 comprises related searches obtained using the Bing API and ranked by Eq. (4).

3.2 Experimental Results

The performance of all submitted runs is shown in Table 2. Significance with respect to uogTr-S-C-3 (for Chinese) and uogTr-S-J-1 (for Japanese) is verified using the Wilcoxon signed-rank test. In particular, the symbols △ and ▲ denote significant improvements at the $p < 0.05$ and $p < 0.01$ levels, respectively. From the table, we observe the following:

- uogTr-S-C-5 (Chinese) and uogTr-S-J-2 (Japanese) outperform all other runs, showing that the Bing related searches provide an informal upper-bound;

- uogTr-S-C-3 shows a weak early performance compared to uogTr-S-C-5, suggesting that more refined re-ranking techniques are needed;

- uogTr-S-C-2 outperforms uogTr-S-C-3, showing that session similarity is a stronger feature than click similarity;

- uogTr-S-C-1 massively outperform uogTr-S-C-3 and approaches uogTr-S-C-5 at evaluation cutoff 30, showing that a learned model can effectively improve the mined subtopics;

- uogTr-S-C-4 differs only marginally from uogTr-S-C-1, suggesting that the click evidence is too sparse for diversification;

- uogTr-S-J-1 performs weakly, suggesting that more refined techniques are needed to mine effective subtopics from anchor-text.

4. DOCUMENT RANKING

In the document ranking subtask, we test the performance of our state-of-the-art xQuAD framework [5, 6, 7, 8, 9] for web search result diversification in non-English languages. Based on an initial ranking $\mathcal{R}$ for the query q, xQuAD iteratively builds a re-ranking $\mathcal{S}$ by selecting, at each iteration, a document $d^* \in \mathcal{R} \setminus \mathcal{S}$ such that:

$$d^* = (1 - \lambda)\,\mathrm{P}(d|q) + \lambda\,\mathrm{P}(d, \bar{\mathcal{S}}|q), \tag{5}$$

where $\mathrm{P}(d|q)$ denotes the probability of d being relevant given the query q and $\mathrm{P}(d, \bar{\mathcal{S}}|q)$ denotes the probability of d but none of the documents already selected in $\mathcal{S}$ being diverse given q. These two probabilities are mixed using the parameter λ, which implements a trade-off between promoting relevant and diverse documents [6]. By marginalising over the possible aspects of q, the probability $\mathrm{P}(d, \bar{\mathcal{S}}|q)$ can be further broken down as:

$$\mathrm{P}(d, \bar{\mathcal{S}}|q) = \sum_{a \in \mathcal{A}} \Big[\mathrm{P}(a|q)\,\mathrm{P}(d|q,a) \prod_{d_j \in \mathcal{S}} (1 - \mathrm{P}(d_j|q,a))\Big], \tag{6}$$

where $\mathrm{P}(a|q)$ denotes the importance of the aspect a given the query q, $\mathrm{P}(d|q,a)$ denotes the coverage of d given the query q and the aspect a, and $\mathrm{P}(\bar{\mathcal{S}}|q,a)$ denotes the novelty of any document satisfying a, based on the probability that none of the documents in $\mathcal{S}$ satisfy this aspect.

In our participation, the aspects $\mathcal{A}$ and the probability $\mathrm{P}(a|q)$ are directly obtained from some of the subtopic mining approaches described in Section 3. The exception is for Bing related searches, which are uniformly weighted in our document ranking runs, such that:

$$\mathrm{P}(a|q) = \frac{1}{|\mathcal{A}|}, \forall a \in \mathcal{A}. \tag{7}$$

The probabilities $\mathrm{P}(d|q)$ and $\mathrm{P}(d|q,a)$ are obtained via learning-to-rank, as described in Section 4.1. Finally, the diversification trade-off λ is automatically set either uniformly for all queries, or on a per-query basis, so as to perform a selective diversification, according to the ambiguity level of each query, as discussed in Section 4.2.

Document Features			Query Features		
Group	Feature	#	Group	Feature	#
WM	BB2	4	QPP	AvICTF*	1
	BM25	4		AvIDF*	1
	DPH*	4		AvPMI*	1
	LM	4		EnIDF*	1
	MQT	4		Gamma1*	1
	PL2	4		Gamma2*	1
DM	pBiL*	8		QueryFrequency*	1
	MRF	8		QueryScope*	1
LA	Absorbing*	1		TermCount*	1
	EdgeRecip	1		TokenCount*	1
	Inlinks	1	LOG	ClickCount	4
	Outlinks	1		ClickEntropy	1
	PageRank*	3		HostEntropy	1
URL	URLDigits	2		ReformCount	4
	URLComponents	3		SessionDuration	4
	URLLength*	3	CORR	ExtIntentCorrel*	12
	URLType*	1		IntIntentCorrel*	16
Grand Total		56	Grand Total		52

Table 4: Document and query features. All features are used in the Chinese runs; the Japanese runs use only features marked with a star (*).

4.1 Ranking Model

In order to obtain effective estimates of the relevance of a document to the initial query q (i.e., $\mathrm{P}(d|q)$) and of its coverage of multiple aspects of this query (i.e., $\mathrm{P}(d|q,a)$), we resort to machine learning. In particular, we leverage several document features traditionally used in the learning-to-rank literature and also shown to be effective in a diversification scenario [7, 8]. In total, we employ 56 features for the Chinese document ranking runs, as described in the left side of Table 4. These include standard weighting models (WM), term dependence models (DM), link analysis (LA) and URL features. For the Japanese runs, given our time constraints, we experimented with a smaller subset of these features, marked with a star (*) in Table 4.

As a learning algorithm, we use Metzler's Automatic Feature Selection (AFS) listwise learner, which has been shown to be effective for web search [7, 8]. In order to learn effective models, we use the 100 queries and relevance assessments from the adhoc task of the TREC 2009 and 2010 Web tracks. As a secondary investigation, we analyse whether our ranking models learned on English web search data generalise effectively to the Chinese and Japanese corpora.

4.2 Ambiguity Model

Recognising that not all queries are equally ambiguous, we investigate the effectiveness of our previously proposed approach to selectively diversify the search results [6]. In particular, given a test query q, our approach automatically predicts how to best set the trade-off λ between promoting relevance or diversity in the ranking, based upon an ambiguity model learned from training queries similar to q.

To enable our approach, we represent each query in a space of 52 features, summarised in the right portion of Table 4. These include query performance predictors (QPP), query log-based features (LOG), and correlation features (CORR). The first two feature groups have been used in our previous works [6, 8], while the last group is introduced here to measure the similarity between the rankings produced for the initial query and those produced for each of its identified aspects. All 52 features are used for the Chinese runs. For

	Rel.	Diversity				I-rec			D-nDCG			D#-nDCG		
Run	P(d|q)	$\mathcal{A}$	P(a|q)	P(d|q,a)	λ	@10	@20	@30	@10	@20	@30	@10	@20	@30
uogTr-D-C-6	L56	–	–	–	–	0.6527	0.7550	0.8090	0.4207	0.4381	0.4415	0.5367	0.5965	0.6253
uogTr-D-C-1	L56	BR	Eq. (7)	L56	Uni	0.6406	0.7458	0.7781	0.4252	0.4517	0.4331	0.5329	0.5987	0.6056
uogTr-D-C-2	L56	BR	Eq. (7)	L56	Sel	0.6600	0.7550	0.7931	0.4316	0.4658$^{\triangle}$	0.4600	0.5458	0.6104	0.6265
uogTr-D-C-3	L56	SQ	Eq. (3)	L56	Uni	0.6301	0.7430	0.7851	0.4480$^{\blacktriangle}$	0.4782$^{\blacktriangle}$	0.4698$^{\blacktriangle}$	0.5390	0.6106$^{\triangle}$	0.6274
uogTr-D-C-4	L56	SQ	Eq. (3)	L56	Sel	0.6474	0.7500	0.7911	0.4423$^{\triangle}$	0.4748$^{\blacktriangle}$	0.4758$^{\blacktriangle}$	0.5449	0.6124$^{\blacktriangle}$	0.6334$^{\triangle}$
uogTr-D-C-5	L56	SQ	Tab. 3: SS	L56	Uni	0.6624	0.7603	0.8028	0.4374$^{\triangle}$	0.4661$^{\blacktriangle}$	0.4632$^{\blacktriangle}$	0.5499$^{\triangle}$	0.6132$^{\blacktriangle}$	0.6330$^{\blacktriangle}$
uogTr-D-J-5	L12	–	–	–	–	0.6697	0.7701	0.8028	0.4034	0.4531	0.4477	0.5365	0.6116	0.6253
uogTr-D-J-1	L12	BR	Eq. (7)	L12	Uni	0.6845	0.7876	0.8224	0.4715$^{\blacktriangle}$	0.5138$^{\blacktriangle}$	0.5002$^{\blacktriangle}$	0.5780$^{\blacktriangle}$	0.6507$^{\blacktriangle}$	0.6613$^{\blacktriangle}$
uogTr-D-J-2	L12	BR	Eq. (7)	L12	Sel	0.6840	0.7835	0.8231	0.4673$^{\blacktriangle}$	0.5143$^{\blacktriangle}$	0.5014$^{\blacktriangle}$	0.5756$^{\blacktriangle}$	0.6489$^{\blacktriangle}$	0.6622$^{\blacktriangle}$
uogTr-D-J-3	L12	AT	Eq. (2)	L12	Uni	0.6702	0.7692	0.7972	0.4091	0.4493	0.4425	0.5397	0.6093	0.6198
uogTr-D-J-4	L12	AT	Eq. (2)	L12	Sel	0.6660	0.7752	0.8046	0.4088	0.4522	0.4502	0.5374	0.6137	0.6274

Table 5: Document ranking results. uogTr-D-C-6 is an unofficial baseline run. SQ, AT, and BR denote query aspects derived from the SogouQ query log, the ClueWeb09-JA anchor-text, and using the Bing related searches API, respectively, as described in Section 3. L56 and L12 denote the ranking models learned for Chinese and Japanese, respectively, as discussed in Section 4.1. UNI and SEL refer to the uniform and selective diversification regimes described in Section 4.2, respectively.

the Japanese runs, log-based features are left out, as we do not have access to a Japanese query log.

Different from our ranking models described in Section 4.1, it is unlikely that an ambiguity model would easily generalise across languages. Hence, we use the small set of 10 Chinese and 10 Japanese queries provided by the organisers as training data. Instead of learning an ambiguity model a priori, we deploy an instance-based learning approach using a k-nearest neighbour regression, as this was shown to perform well for English [6]. In our submissions, we test both uniform (based on all 10 training queries, regardless of the test query) and selective (based on the $k = 3$ most similar training neighbours of each test query) ambiguity models.

4.3 Document Ranking Runs

We submitted five runs to the Chinese document ranking subtask. A 6th (unofficial) run is included here as a baseline:

- uogTr-D-C-6 deploys our learned ranking models without diversification, as an unofficial baseline;
- uogTr-D-C-1 deploys xQuAD uniformly over uogTr-D-C-6, with query aspects given by Bing related searches (uogTr-S-C-5);
- uogTr-D-C-2 is similar to uogTr-D-C-1, except that diversification is performed selectively;
- uogTr-D-C-3 deploys xQuAD uniformly over uogTr-D-C-6, with query aspects given by SogouQ reformulations, ranked by our data-driven approach (uogTr-S-C-1);
- uogTr-D-C-4 is similar to uogTr-D-C-3, except that diversification is performed selectively;
- uogTr-D-C-5 deploys xQuAD uniformly over uogTr-D-C-6, with query aspects given by SogouQ reformulations, ranked by session similarity (uogTr-S-C-2).

Additionally, we submitted another five Japanese runs:

- uogTr-D-J-5 deploys our learned ranking models without diversification, as a baseline;
- uogTr-D-J-1 deploys xQuAD uniformly over uogTr-D-J-5, with query aspects given by Bing related searches (uogTr-S-J-2);
- uogTr-D-J-2 is similar to uogTr-D-J-1, except that diversification is performed selectively;
- uogTr-D-J-3 deploys xQuAD uniformly over uogTr-D-J-5, with query aspects given by related anchor-text (uogTr-S-J-1);
- uogTr-D-J-4 is similar to uogTr-D-J-3, except that diversification is performed selectively;

4.4 Experimental Results

The performance of all submitted runs is shown in Table 5. Significance with respect to uogTr-D-C-6 (for Chinese) and uogTr-D-J-5 (for Japanese) is verified using the Wilcoxon signed-rank test. As before, the symbols $\triangle$ and $\blacktriangle$ denote significant improvements at the $p < 0.05$ and $p < 0.01$ levels, respectively. From the table, we observe the following:

- uogTr-D-C-6 (Chinese) and uogTr-D-J-5 (Japanese) provide strongly performing baselines;
- uogTr-D-C-{1..5} modestly outperform uogTr-D-C-6 in terms of I-rec, and substantially in terms of D-nDCG, which attests the effectiveness of xQuAD for Chinese diversification;
- uogTr-D-J-{1..4} modestly outperform uogTr-D-J-5 in terms of I-rec, and substantially in terms of D-nDCG, which attests the effectiveness of xQuAD also for Japanese diversification;
- uogTr-D-C-{2,4} consistently outperform uogTr-D-C-{1,3}, showing that selective diversification is helpful for Chinese diversification;
- uogTr-D-J-{2,4} do not consistently outperform runs uogTr-D-J-{1,3}, showing that selective diversification is less helpful for Japanese diversification;
- uogTr-D-C-{3..5} generally outperform runs uogTr-D-C-{1,2}, showing both the promise of our data-driven approach to subtopic mining (uogTr-S-C-1) and the robustness of xQuAD to noisy aspects.

5. CONCLUSIONS

We have participated in both the subtopic mining and document ranking subtasks of the NTCIR-9 Intent task. In the subtopic mining subtask, we have introduced a novel data-driven approach to identify effective query aspects from a query log. Our approach is general and has shown promising results in a direct comparison to related queries provided by a commercial search engine. In the document ranking subtask, we have attested the effectiveness of our xQuAD framework for diversifying search results from non-English corpora. Our results have shown that the ideas underlying the framework are sound and generalise smoothly and effectively to Chinese and Japanese data, with some of our document ranking runs performing among the top ones.

6. REFERENCES

[1] V. Dang and B. W. Croft. Query reformulation using anchor text. In *Proc. of WSDM*, pages 41–50, 2010.

[2] R. McCreadie, C. Macdonald, I. Ounis, J. Peng, and R. L. T. Santos. University of Glasgow at TREC 2009: Experiments with Terrier—Blog, Entity, Million Query, Relevance Feedback, and Web tracks. In *Proc. of TREC*, 2009.

[3] D. A. Metzler. Automatic feature selection in the Markov random field model for information retrieval. In *Proc. of CIKM*, pages 253–262, 2007.

[4] I. Ounis, G. Amati, V. Plachouras, B. He, C. Macdonald, and C. Lioma. Terrier: a high performance and scalable information retrieval platform. In *Proc. of OSIR at SIGIR*, 2006.

[5] R. L. T. Santos, C. Macdonald, and I. Ounis. Exploiting query reformulations for Web search result diversification. In *Proc. of WWW*, pages 881–890, 2010.

[6] R. L. T. Santos, C. Macdonald, and I. Ounis. Selectively diversifying Web search results. In *Proc. of CIKM*, pages 1179–1188. ACM, 2010.

[7] R. L. T. Santos, C. Macdonald, and I. Ounis. Intent-aware search result diversification. In *Proc. of SIGIR*, pages 595–604, 2011.

[8] R. L. T. Santos, R. McCreadie, C. Macdonald, and I. Ounis. University of Glasgow at TREC 2010: Experiments with Terrier in Blog and Web tracks. In *Proc. of TREC*, 2010.

[9] R. L. T. Santos, J. Peng, C. Macdonald, and I. Ounis. Explicit search result diversification through sub-queries. In *Proc. of ECIR*, pages 87–99, 2010.

[10] F. Silvestri. Mining query logs: turning search usage data into knowledge. *Found. Trends Inf. Retr.*, 4(1–2):1–174, 2010.

[11] R. Song, M. Zhang, T. Sakai, M. P. Kato, Y. Liu, M. Sugimoto, Q. Wang, and N. Orii. Overview of the NTCIR-9 Intent task. In *Proc. of NTCIR*, 2011.