University of Essex at the TREC 2012 Session Track

M-Dyaa Albakour*, University of Glasgow
Udo Kruschwitz, University of Essex
dyaa.albakour@glasgow.ac.uk, udo@essex.ac.uk

1 Summary of Experiments

The primary goal of our participation in the Session track is to further evaluate our *anchor expansion* technique proposed in the previous year [1]. In particular, we aim to test the effectiveness of this approach on a more realistic dataset collected this year. SWIRL 2012 noted that there is still a large gap between the study of users and the study of IR algorithms [2], so the session data collected for this year's session track aimed at addressing some of the suggestions made in the SWIRL 2012 report .

We exploit the anchor logs to derive query expansions that are relevant to user information needs throughout the session. Anchor text has shown to be effective for a variety of information retrieval tasks. This include the diversity task in the TREC Web track [3]. Anchor text can be considered as a replacement to user queries as often web authors use similar labels to describe web pages to those used by searchers to find them [4]. Moreover, Dang and Croft have recently shown how anchor text can be used to simulate user sessions. They have considered all the anchor text pointing to the same document as queries in the same user session [5].

We derive query expansions from the anchor logs by using anchor logs as simulation of query logs to derive related terms or phrases that represent the user information needs in the session (RL2). We also use further interaction data to create a better model of the user interests. We consider the documents displayed throughout the session as good indicators of the user's interests which is a similar approach to the pseudo relevance feedback model, e.g. [6] (RL3). In addition, we consider only those clicked as useful indicators which is a similar approach to the implicit relevance feedback models, e.g. [7, 8] (RL4).

Following this, we submitted two runs to the Session track:

- **essexSAnchor**: This run uses the aforementioned anchor expansion technique to generate RL2, RL3 and RL4.

*This work was conducted while the author was working as a researcher at the University of Essex

Table 1: $nDCG$@10 of our runs and the TREC median. The arrows indicate improvement(↑), decline (↓) or identical (↔) against the previous results lists, the first arrow in a cell relates to RL1, the second arrow to RL2 and so on. Double arrows (⇑ / ⇓) indicates the comparison is statistically significant returning a two tailed t-test $value < 0.01$. The figure in bold is the top obtained score in either runs.

	RL1	RL2	RL3	RL4
TREC median	0.2455	0.1746	0.1901	0.2160
essexSAnchor	0.1941	⇑ 0.2204	⇑↑ 0.2265	⇑↑↑ **0.2307**
Wiki	0.1941	↓ 0.1899	↓↔0.1899	↓↔↔0.1899

- **essexSWiki**: This is our baseline run. It expands the current query of each session with the term "wikipedia". The rationale for this is that by doing so in Session Track 2010 [9], a statistically significant improvement over a bag-of-query-terms baseline was observed.

In both runs, we use the public Indri index[1] of the ClueWeb09 Category B dataset, which uses the Dirichlet's language modelling and supports query expansion with belief operators. In addition, the Waterloo spam rankings[2] are used to filter out spam documents from the rankings. We consider documents with scores of 70% or less as spam which is recommended by the creators of those rankings [10]. Finally, in both runs, the ranked list RL1 is generated by simply using the current query in the session.

Table 1 reports the performances of the our two runs along with the TREC median using $nDCG$@10. First, most of our results are above the TREC median. Second, the results show that our anchor expansion technique is capable of using the interaction data throughout the user session to improve the retrieval performance. We also see as more interaction data is available to the anchor expansion technique, the retrieval performance increases. In addition, these results are consistent with the previous year's with one exception where an improvement is now observed from $RL3$ to $RL4$. In other words, with what appears to be more realistic log data, we can now observe an improvement in performance when taking into account what users have looked at and not just what they have been presented. Finally, the baseline is not effective as it was in the previous year. This is a good indication of how realistic the task is. Simply returning Wikipedia articles about the topic would not necessarily satisfy the user's information needs.

Acknowledgements

This research is part of the AutoAdapt research project. AutoAdapt is funded by EPSRC grants EP/F035357/1 and EP/F035705/1.

[1] `http://boston.lti.cs.cmu.edu:8085/clueweb09/search/cataenglish/lemur.cgi`
[2] `http://durum0.uwaterloo.ca/clueweb09spam/`

References

[1] M-D. Albakour, U. Kruschwitz, B. Neville, D. Lungely, M. Fasli, and N. Nanas. University of Essex at the TREC 2011 Session Track. In *Proceedings of the 20^{th} Text Retrieval Conference (TREC 2011)*. NIST, 2012.

[2] James Allan, Bruce Croft, Alistair Moffat, and Mark Sanderson. Frontiers, challenges, and opportunities for information retrieval: Report from swirl 2012 the second strategic workshop on information retrieval in lorne. *SIGIR Forum*, 46(1):2–32, May 2012.

[3] N. Craswell, D. Fetterly, M. Najork, S. Robertson, and E. Yilmaz. Microsoft research at trec 2009: Web and relevance feedback tracks. In *Proceedings of the 18th Text REtrieval Conference (TREC)*. NIST, 2009.

[4] Nadav Eiron and Kevin S. McCurley. Analysis of anchor text for web search. In *SIGIR '03: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pages 459–460, New York, NY, USA, 2003. ACM.

[5] Van Dang and Bruce W. Croft. Query reformulation using anchor text. In *WSDM '10: Proceedings of the third ACM international conference on Web search and data mining*, pages 41–50, New York, NY, USA, 2010. ACM.

[6] J. Rocchio. Relevance Feedback in Information Retrieval. In *The SMART Retrieval System*, pages 313–323, 1971.

[7] D. Kelly and N. J. Belkin. Reading time, scrolling and interaction: Exploring implicit sources of user preferences for relevance feedback during interactive information retrieval. In *Proceedings of the 24^{th} Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2001)*, pages 408–409. ACM, 2001.

[8] F. Radlinski and T. Joachims. Query chains: learning to rank from implicit feedback. In *Proceeding of the 11^{th} ACM SIGKDD International Conference on Knowledge Discovery in Data Mining (SIGKDD 2005)*, pages 239–248. ACM, 2005.

[9] Deirdre Lungely, M-Dyaa Albakour, and Udo Kruschwitz. The use of domain modeling to improve performance over a query session. In *Proceedings of the ECIR'11 workshop on Information Retrieval Over Query Sessions, SIR'11*, 2011.

[10] G. V. Cormack, M. D. Smucker, and C. L. Clarke. Efficient and Effective Spam Filtering and Re-ranking for Large Web Datasets. *Information Retrieval*, 14:441–465, 2011.