

Exploring folksonomies for adaptive query expansion

Fabio Gasparetti

Department of Computer Science and Automation
Artificial Intelligence Laboratory
Roma Tre University
Via della Vasca Navale, 79 - 00146 Rome, Italy
gaspare@dia.uniroma3.it

Abstract. Web is growing at an astonishing rate, making it necessary to devise effective methods for helping users find what they are looking for. Adaptive query expansion (QE) allows users to better define their search domain by supplementing the original query with additional terms related to their preferences and information needs [3].

While Automatic QE has a long history in the information retrieval community, there are a very few attempts to make this technique adaptive to user current needs (e.g., [6, 1, 7]). Most of the Automatic QE approaches are based on the following assumption: documents classified higher by an initial search contain many useful terms that can help discriminate relevant documents from irrelevant ones. Despite the large number of studies, a crucial issue is that the expansion terms identified through traditional methodologies from the pseudo-relevant documents may not be all useful [2]. One of the failure reasons of the query expansion has been identified in the lack of relevant documents in the local collection. Consequently, some works advance the use of an external resource for query expansion in order to improve the effectiveness of query expansion, such as thesaurus (e.g., WordNet) [4], Wikipedia [8], and search engine query logs [5]. Recently, several authors have focused on social annotations in various tasks, largely motivated by their increasing availability through many Web-based applications.

The proposed approach¹ is an extension of the traditional QE techniques, which rely on the computation of two-dimensional co-occurrence matrices. Our system makes use of three-dimensional co-occurrence matrices, where the added dimension is represented by semantic classes (i.e., categories comprising all the terms that share a semantic property) related to the folksonomy extracted from social bookmarking services such as delicious, Digg, and StumbleUpon.

The whole procedure of adaptation is completely transparent to the user, as it takes place in an implicit way based on his profile. The user profile is created and dynamically updated using the information related to visited pages and corresponding search queries. The system analyzes the input queries and, if they actually reflect the interests already shown by

¹ This is a joint work with my colleagues Claudio Biancalana, Alessandro Micarelli and Giuseppe Sansonetti.

the user in previous searches, it returns different QEs involving different semantic fields. The output of the system is structured in different blocks categorized through keywords, thus helping the user decide which result is most relevant to him. A deep evaluation on different datasets and real users shows that our system outperforms some well-known approaches in the literature.

Keywords: adaptive query expansion, folksonomy, search engine, information retrieval

References

1. Bouadjenek, M.R., Hacid, H., Bouzeghoub, M., Daigremont, J.: Personalized social query expansion using social bookmarking systems. In: Proceedings of the 34th international ACM SIGIR conference on Research and development in Information. pp. 1113–1114. SIGIR '11, ACM, New York, NY, USA (2011), <http://doi.acm.org/10.1145/2009916.2010075>
2. Cao, G., Nie, J.Y., Gao, J., Robertson, S.: Selecting good expansion terms for pseudo-relevance feedback. In: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval. pp. 243–250. SIGIR '08, ACM, New York, NY, USA (2008), <http://doi.acm.org/10.1145/1390334.1390377>
3. Carpineto, C., Romano, G.: A survey of automatic query expansion in information retrieval. *ACM Comput. Surv.* 44(1), 1 (2012)
4. Collins-Thompson, K., Callan, J.: Query expansion using random walk models. In: Proceedings of the 14th ACM international conference on Information and knowledge management. pp. 704–711. CIKM '05, ACM, New York, NY, USA (2005), <http://doi.acm.org/10.1145/1099554.1099727>
5. Cui, H., Wen, J.R., Nie, J.Y., Ma, W.Y.: Query expansion by mining user logs. *IEEE Trans. on Knowl. and Data Eng.* 15, 829–839 (July 2003), <http://dl.acm.org/citation.cfm?id=1435677.858986>
6. De Meo, P., Quattrone, G., Ursino, D.: A query expansion and user profile enrichment approach to improve the performance of recommender systems operating on a folksonomy. *User Model. User-Adapt. Interact.* 20(1), 41–86 (2010)
7. Lin, Y., Lin, H., Jin, S., Ye, Z.: Social annotation in query expansion: a machine learning approach. In: Proceedings of the 34th international ACM SIGIR conference on Research and development in Information. pp. 405–414. SIGIR '11, ACM, New York, NY, USA (2011), <http://doi.acm.org/10.1145/2009916.2009972>
8. Xu, Y., Jones, G.J., Wang, B.: Query dependent pseudo-relevance feedback based on wikipedia. In: Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval. pp. 59–66. SIGIR '09, ACM, New York, NY, USA (2009), <http://doi.acm.org/10.1145/1571941.1571954>