Snippet-Based Relevance Predictions for Federated Web Search

Thomas Demeester[1], Dong Nguyen[2], Dolf Trieschnigg[2], Chris Develder[1], and Djoerd Hiemstra[2]

[1] Ghent University - iMinds, Ghent, Belgium
{tdmeeste,cdvelder}@intec.ugent.be
[2] University of Twente, Enschede, The Netherlands
{d.nguyen,d.trieschnigg,d.hiemstra}@utwente.nl

Abstract. How well can the relevance of a page be predicted, purely based on snippets? This would be highly useful in a Federated Web Search setting where caching large amounts of result snippets is more feasible than caching entire pages. The experiments reported in this paper make use of result snippets and pages from a diverse set of actual Web search engines. A linear classifier is trained to predict the snippet-based user estimate of page relevance, but also, to predict the actual page relevance, again based on snippets alone. The presented results confirm the validity of the proposed approach and provide promising insights into future result merging strategies for a Federated Web Search setting.

Keywords: Federated Web search, snippets, classification, relevance judgments.

1 Introduction

The actual influence of result snippets on the overall efficiency of search engines has been largely understudied, since until recently there was no dedicated test collection that contains actual result snippets as well as the resulting web pages from a wide variety of sources. This gap was filled by the dataset introduced in [4], designed for research on Federated Web Search (FWS) and made available to researchers. It contains a large amount of sampled data from over a hundred diverse online search engines, and relevance judgments of snippets and pages for the results from these resources in response to the 50 reference topics [3].

The goal of this paper is to analyze how well result snippets from a specific origin can be used to predict the relevance of the corresponding page, given a query, and how well these can be cast into a single merged list. Our research is motivated by the question whether in FWS, merging the results from various resources into a single ranked list, could be done based on the snippets alone. Another issue is whether caching large amounts of result snippets, instead of a small number of full pages, would improve the accuracy of fast FWS systems, e.g., organized in a peer-to-peer setting.

P. Serdyukov et al. (Eds.): ECIR 2013, LNCS 7814, pp. 697–700, 2013.

A main problem in Federated Information Retrieval (FIR) and related to the goal of this paper, is result merging, i.e., ordering the results retrieved from several resources into a single ranked list [6]. Classification-based methods have already been used in FIR, but mostly focusing on predicting the relevant resources, for example in [1] for verticals. The snippets in our collection allow applying classification techniques for relevance predictions on the result level. In this paper we focus on predicting binary relevance and use a maximum entropy classifier [5].

We will describe the classifier designed to make snippet-based relevance predictions, shortly cover the extracted snippet features, and report classification efficiency for snippet and page predictions. We will also report the precision of the top-30 results, ranked according to the classifier output, and show that even with simple features we clearly outperform a round-robin merging of results from three of the strongest search engines.

2 Inferring Relevance from Snippets

The FWS collection presented in [4] contains several levels of relevance judgments, but here we will only consider strict binary relevance levels, for which snippets were judged to be definitely relevant[1], and pages were indicated as highly relevant or better. In [3] it was shown that even for the best resources only about 2/3 of all 'definitely relevant' snippets corresponded with a highly relevant result page. A classifier trained on snippet labels would hence not be able to overcome this intrinsic gap in relevance when used for the prediction of the page labels. Instead, we need to directly focus on page relevance, but still based on the information contained in the snippets alone, as motivated above. We will show that a binary linear classifier, trained on simple snippet features, is able to predict page relevance with a comparable accuracy as for snippet predictions, despite the apparent mismatch in relevance.

A number of the used resources contain very few highly relevant pages [3], due to the nature of the query set, which has been designed for the 2010 TREC Web Track [2]. Therefore, we will only consider the resources from the following verticals: General Web Search, Multimedia, Encyclopedia/Dictionaries, Blogs, Books & Libraries, News, and Shopping. In total, we retain 50 out of the original 108 resources, which together provided almost 90% of all highly relevant result pages from the collection. No further resource selection is performed.

In our experiments using the maximum entropy classifier, we applied 10-fold cross validation: for each run we trained on the results from 45 out of the total of 50 topics, and tested on the remaining 5. Results were averaged over each of the test topics.

[1] With *snippet relevance*, we actually mean the snippet-based estimate of the page relevance (or, in annotation terms, 'would you click this snippet?').

2.1 Snippet Feature Extraction

The snippet-based features used for classification, fall into three groups. The first group, denoted as 'constant', contains an overall constant feature and a binary feature for each of the considered resources. Considering only these features corresponds to assigning a prior probability of relevance to each resource, according to the number of highly relevant results found in the training set. These vary widely, with the largest general web search engines on top. The second group contains 27 features directly calculated from each resource's snippets, without requiring snippets from other resources. This distinction is important, e.g., for result merging purposes, where snippets from only a few resources would be available. These features include binary features for the considered snippet's rank in the result list (we only crawled the top 10 results for each resource), features indicating the presence of query terms in the title, summary and location fields of the query (see [4]), and the length of these fields. The combination of these features together with the constant features is indicated as 'local'. Finally, 8 global features are extracted, calculated from the whole set of snippets for the considered topic. These include further length-based features for the different snippet fields, but relative to the average length in the considered set, and tf-idf scores for the snippet title and summary in response to the query (without stemming). All features together are indicated with 'all'.

2.2 Results

In web search, precision is typically considered more relevant than recall, we thus set the relevance cut-off for our classifier on a predicted probability of relevance of 0.5, with relatively high precision, and sufficient recall. Table 1 shows the mean classifier precision and recall, both for the prediction of snippet and page relevance. Using all features, the prediction of page relevance performs only slightly lower than for snippets. Note that no resource had on average more than 5 out of 10 highly relevant pages (cfr. the zero scores using only the constant features), whereas the snippet judgments were more optimistic. Also, calculating the global features did not yield much improvement in terms of precision, but allowed retrieving more relevant results.

Next, we create a merged result list by ranking the results according to their predicted probability of relevance. We compare this with a baseline using a round-robin ranking of the three major search engines, i.e., Google, Yahoo!, and Bing. Fig. 1 shows the precision in the top-30 of this ranked list, together with the

Table 1. Snippet-based prediction accuracy for snippet and page relevance

	snippet relevance		page relevance	
features	precision	recall	precision	recall
constant	0.52	0.04	0.00	0.00
local	0.65	0.27	0.61	0.15
all	0.67	0.37	0.62	0.23

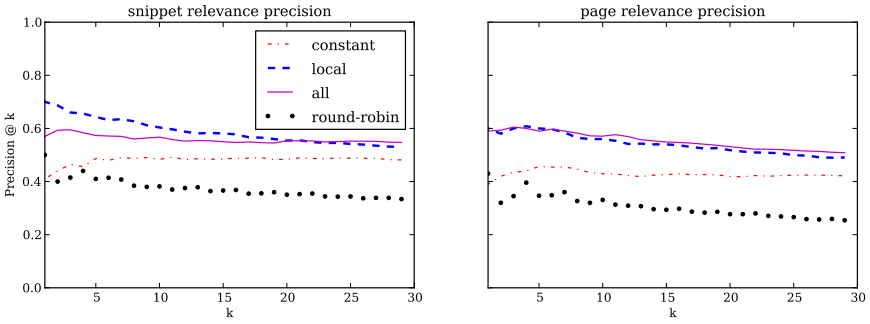

Fig. 1. Precision as a function of the rank k for the 30 highest ranked results as predicted for all resources, compared with the round-robin result for Google, Yahoo!, and Bing

baseline. The precision for the predicted results is considerably higher than for the round-robin baseline, even for the constant features (i.e., the mere ranking of resources that best explains the training labels), and the precision in page predictions is again not much lower than for snippets.

3 Conclusions

We demonstrated that in a Federated Web Search setting, result page relevance predictions are possible, based on snippet features alone. A detailed analysis in terms of the number of resources that provide relevant results, which features contributed the most to the results, etc., is proposed as future work, together with full-fledged result merging research for the current setting, based on snippet classification.

Acknowledgments. This research was cofunded by iMinds (Interdisciplinary institute for Technology) in Flanders, and the Netherlands Organization for Scientific Research, NWO, grants 639.022.809 and 640.005.002.

References

1. Arguello, J., Callan, J., Diaz, F.: Classification-based resource selection. In: CIKM 2009. ACM Press, New York (2009)
2. Clarke, C.L.A., Craswell, N., Soboroff, I., Cormack, G.V.: Overview of the TREC 2010 Web Track. In: TREC, pp. 1–9 (2010)
3. Demeester, T., Nguyen, D., Trieschnigg, D., Develder, C., Hiemstra, D.: What Snippets Say about Pages in Federated Web Search. In: Hou, Y., Nie, J.-Y., Sun, L., Wang, B., Zhang, P. (eds.) AIRS 2012. LNCS, vol. 7675, pp. 250–261. Springer, Heidelberg (2012)
4. Nguyen, D., Demeester, T., Trieschnigg, D., Hiemstra, D.: Federated Search in the Wild: the Combined Power of over a Hundred Search Engines. In: CIKM 2012 (2012)
5. Nigam, K., Lafferty, J., Mccallum, A.: Using Maximum Entropy for Text Classification. In: IJCAI 1999 Workshop on Information Filtering (1999)
6. Shokouhi, M., Li, L.: Federated Search. Foundations and Trends in Information Retrieval 5(1), 1–102 (2011)