Dynamic Search in Specialized Domains

Felipe Moraes
CS Dept, UFMG
Belo Horizonte, MG, Brazil
felipemoraes@dcc.ufmg.br

Rodrygo L. T. Santos
CS Dept, UFMG
Belo Horizonte, MG, Brazil
rodrygo@dcc.ufmg.br

Nivio Ziviani
CS Dept, UFMG & Kunumi
Belo Horizonte, MG, Brazil
nivio@dcc.ufmg.br

1 INTRODUCTION

The need for exploration commonly arises in professional search settings such as in medical, legal, patent, military intelligence, and academic search, but also in personal searches such as in travel planning or personal health research [1]. Exploratory searches often involve complex sessions, demanding multiple interactions between the user and a search system. Along the interactive process, the system must dynamically adapt to each feedback provided by the user in order to improve the understanding of the user's need and the usefulness of the subsequently retrieved documents.

Research on exploratory search has been supported by several initiatives. The Text REtrieval Conference (TREC) have hosted related research tracks on interactive search, search within sessions, search for task completion and, more recently, dynamic search in specialized domains. The latter problem, embodied by the TREC Dynamic Domain track,[1] is the focus of this paper. Given an initial query, a dynamic search system must improve its understanding of the user's information need through a series of interactions. In each interaction, the user may provide the system with feedback on the relevance of specific passages of the retrieved documents with respect to one or more aspects underlying his or her information need. The system must then choose to either provide the user with further documents or end the interactive process. An effective system should be able to satisfy as many query aspects as possible (to maximize user satisfaction) with as few interactions as possible (to minimize user effort).

2 EFFECTIVE DYNAMIC SEARCH SYSTEMS

A dynamic search system must cope with four key problems: (i) produce an initial sample of candidate documents given the user's query and the domain of interest; (ii) decide whether the user's information need has been satisfied and eventually stop the interactive process; (iii) leverage the user's feedback to learn an improved aspect model; (iv) produce an enhanced ranking given the learned aspect model. Several attempts have been made to produce dynamic search systems that could effectively tackle these problems. Nevertheless, for the domains considered in the TREC 2015 and 2016 Dynamic Domain track, even the reportedly most effective system in each domain shows only marginal improvements compared to vanilla ad-hoc search baselines, which leverage no user feedback.

In this paper, we aim to better understand the challenges involved in building effective dynamic search systems. To this end, we isolate each of the aforementioned problems as a separate component of a dynamic search system as pictured in Figure 1. In particular, using data from the TREC 2015-2016 Dynamic Domain track, we aim to answer the following research questions:

- $Q1$. How does the initial document sample impact the effectiveness of a dynamic search system?
- $Q2$. What is the impact of feedback modeling on the system's knowledge of the aspects underlying the user's query?
- $Q3$. How do improved coverage estimates impact the system's ability to dynamically adapt its ranking strategy?
- $Q4$. What is the impact of early and late stopping strategies on the attained gain-effort trade-off?

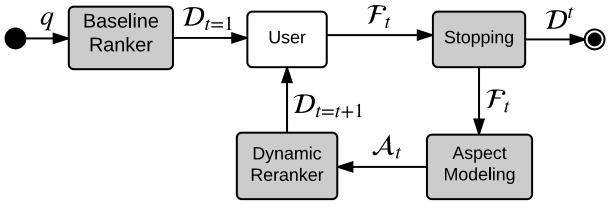

Figure 1: Flow diagram of a typical dynamic search system.

3 CONTRIBUTIONS

Our main contribution in this paper is the investigation of the role of different components on the effectiveness of a dynamic search system for specialized domains. Through a comprehensive analysis, we found that a high-precision baseline ranker may improve dynamic search at early interactions, whereas a high-recall baseline ranker tends to favor later interactions. Moreover, mishandling the user's feedback on individual passages associated with an aspect or on entire aspects may lead to decreased effectiveness. Likewise, we demonstrated the need for accurately estimating the coverage of each retrieved document with respect to each query aspect, particularly for queries with fewer aspects, which seem inherently harder to improve. Finally, we found that early stopping strategies achieve a better gain-effort trade-off compared to late stopping strategies, which highlights the challenge of promoting effective exploration in this task.

REFERENCES

[1] Gary Marchionini. 2006. Exploratory Search: From Finding to Understanding. *Commun. ACM* (2006).

[2] Felipe Moraes, Rodrygo L. T. Santos, and Nivio Ziviani. 2017. On effective dynamic search in specialized domains. In *Proceedings of the 3rd ACM International Conference on the Theory of Information Retrieval*. ACM, Amsterdam, The Netherlands. DOI: http://dx.doi.org/10.1145/3121050.3121065

*This is an extended abstract of Moraes et al. [2].

[1] http://trec-dd.org/