Diversity, intent, and aggregated search

de Rijke, M.

DOI
10.1145/2682862.2684462

Publication date
2014

Document Version
Final published version

Published in
ADCS : proceedings of the 19th Australasian Document Computing Symposium

Citation for published version (APA):
Diversity, Intent, and Aggregated Search

Maarten de Rijke
University of Amsterdam, Amsterdam, The Netherlands
derijke@uva.nl

1. BACKGROUND

Diversity, intent and aggregated search are three core retrieval concepts that receive significant attention. In search result diversification one typically considers the relevance of a document in light of other retrieved documents. The goal is to identify the probable “aspects” of an ambiguous query, retrieve documents for each of these aspects and make the search results more diverse. By doing so, in the absence of any knowledge of users’ context or preferences, the chance that the user will find at least one of these results to be relevant to their underlying information need is increased. Those probable “aspects” of a query may refer to lexical ambiguity (e.g., flash – Adobe Flash, flash light, flash gordon, flash airlines, flash mob, . . . ) or to intentional ambiguity (e.g., pizza – how to make one, where to buy one, images, nutritional value, background, restaurant, . . . ). The automatic discovery of query intent has become an active research area, with a range of observational and algorithmic studies as outcomes. Understanding the likely intents behind a query can help search engines to automatically route the query to the corresponding vertical search engines so as to obtain particularly relevant results, thus greatly improving user satisfaction. In aggregated search the task is to search and assemble information from a variety of sources and to organize the resulting material within a single interface. The result page of a modern search engine often goes beyond a simple ranked list. Many specific intents are addressed by aggregated search solutions: especially presented documents, often retrieved from specific sources, that stand out from the regular organic search results.

2. RECENT ADVANCES

Diversity, intent, and aggregated search give rise to significant research challenges, both algorithmically and in terms of evaluation. In the talk I will highlight recent developments and point out directions for future work. In particular, concerning search result diversification I will run through a new perspective on the problem by casting it as a data fusion problem, following [7] D. Odijk, Burscher, R. Vliegenthart, and M. de Rijke. Automatic thematic content analysis: Finding frames in news. In SocInfo2013. ACM, November 2013.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.

Copyright is held by the owner/author(s).

ADCS ’14, Nov 27–28 2014, Melbourne, VIC, Australia
ACM 978-1-4503-3000-8/14/11.
http://dx.doi.org/10.1145/2682662.2684462.


