

What Makes Book Search in Social Media Complex?

Marijn Koolen¹ Toine Bogers³ Antal van den Bosch⁴ Jaap Kamps^{1,2}

¹ Institute for Logic, Language and Computation, University of Amsterdam, The Netherlands

² Archives and Information Studies, University of Amsterdam, The Netherlands
{marijn.koolen,kamps}@uva.nl

³ Department of Communication and Psychology, Aalborg University Copenhagen, Denmark
toine@hum.aau.dk

⁴ Centre for Language Studies, Radboud University, The Netherlands
a.vandenbosch@let.ru.nl

This is a compressed abstract based on [2]. Real-world information needs are generally complex, yet almost all research focuses on either relatively simple search based on queries or recommendation based on profiles. It is difficult to gain insight into complex information needs from observational studies with existing systems; potentially complex needs are obscured by the systems' limitations. The general aim of this paper is to investigate whether explicit book search requests in social media can be used to gain insight in complex information needs that cannot be solved by a straightforward look-up search. We analyse a large set of annotated book requests from the LibraryThing discussion forums. We investigate 1) the comprehensiveness of book requests on the forums, 2) what relevance aspects are expressed in real-world book search requests, and 3) how different types of search topics are related to types of users, human recommendations, and results returned by retrieval and recommender systems.

We focus on LibraryThing¹ (LT), a popular social cataloguing site with 1.9 million members, which also offers a popular discussion forum to its users for discussing books. The forum is also used for book discovery: thousands of LT members use the forum to receive or provide recommendations for which books to read next. From the forums we can derive rich statements of requests, including explicit statements on the context of use and the context of the user, with example books and 'ground truth' human recommendations. We study this in the context of the INEX Social Book Search Track [1], which builds test collections around the book requests posted on the LT discussion forums. A non-random sample of 2,646 forum threads were annotated on whether the initial message is a *book request* or not, yielding 944 topics (36%) containing a book request. The requests were annotated with the relevance aspect(s) they express, based on set of 8 aspects: content-criteria (e.g. topic, genre), accessibility (reading level, length), familiarity (e.g. similar to previous reading experiences), novelty, engagement, metadata and socio-cultural resonance, with known-item searches labeled as a separate category. The books mentioned by other LT members in the thread were annotated as positive, neutral, or negative suggestions.

¹<http://librarything.com/>, last accessed October 23, 2015.

We found that the LT forum requests are comprehensive, with the majority containing multiple relevance aspects. The two dominating aspects are the *content* of the book (74% of requests) and looking for *familiar* reading experiences (36%), while other aspects are more oriented toward the reading context. In the majority of requests based on *familiarity* the requester provides example books to guide others in their recommendations. The *familiarity* aspects are also mostly used when searching fiction, whereas *content* aspects are less genre-specific. Finally, *content* aspects are more used by active users with large personal catalogues, where *familiarity* and contextual aspects are more typical of less active users with smaller catalogues. Suggestions for content and familiarity aspects also differ in terms of book popularity, with less popular books being suggested for content aspects than for familiarity aspects. The combination of content, context, and examples in a search request is a form of querying that is not supported by any current systems.

Retrieval systems can effectively use the content aspects of the search requests, and recommender systems can pick up signals in the requester's catalogue. With standard systems for retrieval (language model) and recommendation (K nearest neighbour) we demonstrated on a set of 680 requests and associated relevance judgements that a linear combination performs better than the individual approaches, in particular for topic groups where context and familiarity play a role. This suggests that the request type has an important role to play in the design of book discovery systems.

We highlighted the diversity of complex book search requests, and observed a mixture of content and context going beyond currently existing systems. Similar rich profiles and contextual information are available in many modern search scenarios. This is an important first step toward the development of novel information access systems that blend traditional search and (hybrid) recommendation approaches into a coherent whole.

Acknowledgments

This research was supported by the Netherlands Organization for Scientific Research (README project, NWO VIDI # 639.072.601; ExPoSe project, NWO CI # 314.99.108).

References

- [1] M. Koolen, T. Bogers, J. Kamps, G. Kazai, and M. Preminger. Overview of the INEX 2014 social book search track. In *Working Notes for CLEF 2014 Conference, Sheffield, UK, September 15-18, 2014.*, volume 1180 of *CEUR Workshop Proceedings*, pages 462–479. CEUR-WS.org, 2014.
- [2] M. Koolen, T. Bogers, A. van den Bosch, and J. Kamps. Looking for books in social media: An analysis of complex search requests. In *Advances in Information Retrieval - 37th European Conference on IR Research, ECIR 2015. Proceedings*, pages 184–196, 2015.