Mining, ranking and recommending entity aspects

abstract

Reinanda, R.; Meij, E.; de Rijke, M.

Publication date
2015

Document Version
Final published version

Proceedings

Proceedings of the 14th Dutch-Belgian Information Retrieval Workshop

Mining, Ranking and Recommending Entity Aspects (Abstract)

Ridho Reinanda[†]
r.reinanda@uva.nl

Edgar Meij[‡]
emeij@yahoo-inc.com

Maarten de Rijke[†]
derijke@uva.nl

[†] University of Amsterdam, Amsterdam, The Netherlands
[‡] Yahoo Labs, London, United Kingdom

ABSTRACT

With the proliferation of mobile devices, an increasing amount of available structured data, and the development of advanced search result pages, modern-day web search is increasingly geared towards entity-oriented search [1, 7, 9]. A first step and common strategy to address such information needs is to identify entities within queries, commonly known as *entity linking* [6]. Semantic information that is gleaned from the linked entities (such as entity types, attributes, or related entities) is used in various ways by modern search engines, e.g., for presenting an entity card, showing actionable links, and/or recommending related entities [2, 4, 5].

Entities are not typically searched for on their own, however, but often combined with other entities, types, attributes/properties, relationships, or keywords [9]. Such query completions in the context of an entity are commonly referred to as entity-oriented intents or entity aspects [8, 11]. In this paper we study the problem of mining and ranking entity aspects in the context of web search. In particular, we study four related tasks in this paper: (1) identifying entity aspects, (2) estimating the importance of aspects with respect to an entity, (3) ranking entity aspects with respect to a current query and/or user session, and (4) leveraging entity aspects for query recommendation.

The first step in identifying entity aspects involves extracting common queries in the context of an entity and grouping them based on their similarity. We perform this process offline and investigate three matching strategies for clustering queries into entity aspects: *lexical*, *semantic*, and *click-based*. Gathering such entity aspects can already be valuable on its own since they can be used to, e.g., discover bursty or consistent entity intents or to determine entity type-specific aspects [8].

In the next step we rank the obtained entity aspects for each entity in a query-independent fashion using three distinct strategies. This provides us with a mechanism to retrieve the most relevant aspects for a given entity on its own, which, in turn, can be used to, e.g., summarize the most pertinent information needs around an entity or to help the presentation of entity-oriented search results such as customized entity cards on SERPs [1].

The third task that we consider is aspect recommendation. Given an entity and a certain aspect as input, recommend related aspects. This task is motivated by the increasing proliferation of entity-oriented interface elements for web search that can be improved by, e.g., (re)ranking particular items on these elements. Recommending aspects for an entity can also help users discover new and serendipitous information with respect to an entity. We consider two approaches to recommend aspects: *semantic* and *behavioral*. In the semantic approach, relatedness is estimated from a semantic representation of aspects. The behavioral approach is based on the "flow" of aspect transitions in actual user sessions, modeled using an adapted version of the query-flow graph [3].

We perform large-scale experiments on both a publicly available and a commercial search engine's query log to evaluate our proposed methods for mining, ranking, and recommending entity aspects, as well as for recommending queries. We perform contrastive experiments using various similarity measures and ranking strategies. We evaluate the quality of the extracted entity aspects by manually evaluating the generated clusters. Since manually evaluating aspect rankings for entities is not straightforward, we resort to automatic evaluation based on adjacent query pairs. A similar automatic evaluation strategy is also employed to evaluate aspect recommendations.

We find that entropy-based methods achieve the best performance compared to maximum likelihood and language modeling on the task of entity aspect ranking. Concerning aspect recommendation we find that combining aspect transitions within a session and semantic relatedness give the best performance. Furthermore, we show that the entity aspects can be effectively utilized for query recommendation.

This work was presented at SIGIR 2015 [10].

References

[1] N. Balasubramanian and S. Cucerzan. Topic pages: An alternative to the ten blue links. In *IEEE-ICSC 2010*, 2010.
[2] R. Blanco, B. B. Cambazoglu, P. Mika, and N. Torzec. Entity recommendations in web search. In *ISWC '13*, 2013.
[3] P. Boldi, F. Bonchi, C. Castillo, D. Donato, A. Gionis, and S. Vigna. The query-flow graph: Model and applications. In *CIKM '08*, 2008.
[4] L. Hollink, P. Mika, and R. Blanco. Web usage mining with semantic analysis. In *WWW '13*, 2013.
[5] T. Lin, P. Pantel, M. Gamon, A. Kannan, and A. Fuxman. Active objects: Actions for entity-centric search. In *WWW '12*, 2012.
[6] E. Meij, W. Weerkamp, and M. de Rijke. Adding semantics to microblog posts. In *WSDM 2012*, 2012.
[7] P. Pantel and A. Fuxman. Jigs and lures: Associating web queries with structured entities. In *ACL '11*, 2011.
[8] P. Pantel, T. Lin, and M. Gamon. Mining entity types from query logs via user intent modeling. In *ACL'12*, 2012.
[9] J. Pound, P. Mika, and H. Zaragoza. Ad-hoc object retrieval in the web of data. In *WWW '10*, pages 771–780, 2010.
[10] R. Reinanda, E. Meij, and M. de Rijke. Mining, ranking and recommending entity aspects. In *SIGIR 2015: 38th international ACM SIGIR conference on Research and development in information retrieval*. ACM, August 2015.
[11] X. Yin and S. Shah. Building taxonomy of web search intents for name entity queries. In *WWW '10*, 2010.