Task-Based Support in Search Engines

Darío Garigliotti
Aalborg University

Darío Garigliotti is a postdoctoral researcher at Aalborg University, Denmark, where he works in collaboration with Prof. Katja Hose and Assoc. Prof. Johannes Bjerva, in the areas of information extraction, information retrieval and natural language processing. He completed his doctoral degree in Computer Science at University of Stavanger, Norway, under the supervision of Prof. Krisztian Balog. The website at `https://dariogarigliotti.github.io/` contains more information about Garigliotti's work.

The research conducted by Garigliotti focused on utilizing and extending methods and techniques from semantic search within an information access paradigm that aims to support users in achieving their tasks. More specifically, to enhance search engines with functionalities for recognizing the underlying tasks in searches and providing support for task completion. The work presented in this thesis is organized in three grand themes: entity type information for entity retrieval, entity-oriented search intents, and task-based search. Alongside the theoretical and empirical contributions, a number of resource contributions were developed, including several corpora and test collections, and a knowledge base of entity-oriented search intents.

Web search has become a key technology on which people rely daily for getting information about almost everything. The evolution of the search experience has also shaped the expectations of people about it. Many users seem to expect today's web search engines to behave like a kind of "wise interpreter," capable of understanding the meaning behind a search query, realizing its current context, and responding to it directly and appropriately. Search by meaning, or semantic search, encompasses a large portion of information retrieval (IR) research devoted to study more meaningful representations of the information need expressed by the user query. Entity cards, direct displays, and verticals are examples of how major commercial search engines have indeed responded to user expectations, capitalizing on query understanding. That is, giving meaning to information pieces within the query and within the searched items, by mapping them to elements in some knowledge structure of reference. Semantic search can then be seen as a rich toolbox, encompassing multiple techniques to recognize essential knowledge units in the information need, identify them uniquely in the underlying knowledge repositories, and exploit them to address a particular aspect of query understanding [Balog 2018]. The pivotal component that sparked this evolution trend is the increased availability of structured data published in knowledge repositories and knowledge bases, now primary sources of information for entity-oriented search. The whole research presented in this thesis centers around query understanding: identifying target entity types, recognizing and building a knowledge base of entity-oriented search intents, and providing support for task-based search by the means

of query suggestions or task recommendations.

Search is usually performed with a specific goal underlying the query. In many cases, this goal consists of a nontrivial task to be completed. Current search engines support a small set of basic tasks, and most of the knowledge-intensive workload for supporting more complex tasks is left to the user. Task-based search can be viewed as an information access paradigm that aims to enhance search engines with functionalities for recognizing the underlying tasks in searches and providing support for task completion. The research presented in this thesis focuses on utilizing and extending methods and techniques from semantic search in the next stage of the evolution of search engines, namely, to support users in achieving their tasks.

The work presented in this thesis can be grouped in three grand themes. The first one is *entity type information for entity retrieval.* Types—semantic classes grouping multiple entities—are a distinctive property of entities. This grouping characteristic makes types capable to generalize information of many entities. In the context of entity retrieval, a system in place could try to identify the type of entities that the query is seeking. This type-based information could then be combined with the types assigned to each entity in a reference knowledge base, in order to improve the ranking of results. The nature of type information brings about a set of challenges for its usage. The primary objective of this research is then formulated as the following research question: How can entity type information be utilized in ad-hoc entity retrieval? We conduct a systematic evaluation and analysis of methods for type-aware entity retrieval, in terms of three main dimensions [Garigliotti and Balog 2017b]. Also, we revisit the problem of hierarchical target type identification, present a state-of-the-art supervised learning method [Garigliotti et al. 2017], and analyze the usage of automatically identified target entity types for type-aware entity retrieval [Garigliotti et al. 2019].

The second grand theme is *entity-oriented search intents*, that is, understanding what is asked for in queries revolving around entities, by observing patterns in the way that users express their search queries. In particular, we focus on studying those queries that typically consist of (one of the names of) an entity and possibly additional words that refine the intent behind the query. We address the following research question: What do entity-oriented queries ask for, and how can they be fulfilled? We propose a categorization scheme for entity-oriented search intents, and study the distributions of entity intent categories per entity type [Garigliotti and Balog 2018b]. Our follow-up step is building a structured repository or knowledge base of entity-oriented search intents, which could be leveraged in various applications. This representation involves modeling search intents in a structured fashion, as well as devising methods for populating this knowledge base. Hence, the following research question is of our interest: How can we build a knowledge base of entity-oriented search intents? We devise an approach for automatically assigning intent categories to type-level query refiners, and a clustering stage of refiners which express the same underlying intent [Garigliotti and Balog 2018a]. Additionally, we represent each intent uniquely in a knowledge repository, and map to it the facts we acquire about its categorization and refiners. We evaluate performance both component-wise and end-to-end, and demonstrate that our approach is able to generate high-quality data.

The last grand theme is *task-based search*, as a mechanism to assist the user with articulating her information need. In particular, we are firstly interested in generating query

suggestions to support task-based search. Many search scenarios are driven by a specific target or goal, which is often complex and knowledge-intensive, making the user to issue a sequence of queries to eventually complete her underlying task. A way to formally frame this objective is by aiming for the generation, upon an initial query, of a ranked list of query suggestions that cover all the possible subtasks related to the task the user is trying to achieve. We ask the following research question: How can we generate query suggestions for supporting task-based search? To address this question, we design a probabilistic generative framework for task-based query suggestion [Garigliotti and Balog 2017a]. More specifically, we propose an architecture for extracting keyphrases from a variety of sources and generating query suggestions from them. This architecture encompasses components to select documents from particular information sources, extract keyphrases from the documents, and combine the keyphrases with the initial query into query suggestions. Using a test collection from a dedicated benchmark campaign, we propose a principled approach for estimating each of these components. Additionally, we address further related questions, such as answering whether a unified method can produce suggestions in two scenarios, query completions and query refinements. A possible extension of this strategy would be to recommend specific tasks to users, based on their search queries, such as planning a holidays trip or organizing a birthday party. Our main inquiry here can be expressed with this research question: How can we recommend tasks based on search queries and missions? We introduce the problems of query-based task recommendation and mission-based task recommendation, build a dedicated test collection, and establish respective methods as suitable baselines for the introduced problems.

REFERENCES

BALOG, K. 2018. *Entity-Oriented Search*. The Information Retrieval Series, vol. 39. Springer.

GARIGLIOTTI, D. AND BALOG, K. 2017a. Generating query suggestions to support task-based search. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '17. ACM, 1153–1156.

GARIGLIOTTI, D. AND BALOG, K. 2017b. On type-aware entity retrieval. In *Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval*. ICTIR '17. ACM, 27–34.

GARIGLIOTTI, D. AND BALOG, K. 2018a. IntentsKB: A knowledge base of entity-oriented search intents. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. CIKM '18. ACM, 1659–1662.

GARIGLIOTTI, D. AND BALOG, K. 2018b. Towards an understanding of entity-oriented search intents. In *Advances in Information Retrieval - Proceedings of the 40th European Conference on IR Research, ECIR 2018*. Springer, 644–650.

GARIGLIOTTI, D., HASIBI, F., AND BALOG, K. 2017. Target type identification for entity-bearing queries. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '17. ACM, 845–848.

GARIGLIOTTI, D., HASIBI, F., AND BALOG, K. 2019. Identifying and exploiting target entity type information for ad hoc entity retrieval. *Information Retrieval Journal 22,* 3, 285–323.