

# Predicting Death In Game of Thrones

Soheil Gharatappeh

March 26, 2020

## Main Paper

In [4] Zhiltsov et al. discuss entity retrieval in the web of data (ERWD), using multi-fielded entity representation. The idea is to incorporate term dependencies into the task of document retrieval in a collection of structured documents. It is known that retrieval models, such as bigram, ordered and un-ordered n-grams, that use dependencies in query terms are more accurate in case of longer queries. Retrieval models such as Markov Random Field (MRF) represent a query using a graph of dependencies between the query terms and employ bigram, ordered or un-ordered base relationships between the terms, and calculates the score of each document as a linear combination of potential functions. Also, Sequential Dependence Model (SDM), which is based on MRF, assumes sequential dependencies for the query terms in the form of unigram, ordered or un-ordered bigram. These methods improve the quality of entity representation. However, in order to improve the retrieval models, Fielded Sequential Dependence Model (FSDM) is proposed, which generalizes and guarantees sequential dependence, especially in structured documents. It is also suggested that using a five-field entity representation scheme we can improve the quality of entity representation and the semantic relation between them.

This approach is evaluated in comparison to available benchmarks for ERWD such as MLM-CA (mixture of language models coordinate accent) and SDM-CA, and other well-known methods such as LM, BM25, PRMS, etc. by using four different query sets. This work is evaluated in terms of MAP, Precision at 10, Precision at 20, and b-pref. The results suggest an overall improvement in almost all of the cases across different query sets.

## 1st Paper

In [3], Kotov et al. study Query Expansion; a very common technique for improving web search results. The idea is to exploit the graph-based representation of ConceptNet knowledge base in order to enrich the lexico-semantic relations between terms. Query expansion, especially in case of difficult queries, can be improved using related concepts obtained from ConceptNet. The related concepts enable us to represent complex and multi-step inferences between concepts more efficiently. From the whole set of related expansion queries, a subset of more related ones can be chosen automatically using different Machine Learning or heuristic tools in order to increase the efficiency of the retrieval problem. The results suggest that employing learning-based expansion methods leads to improved retrieval results by adding truly related concepts around the difficult query.

## 2nd Paper

Garigliotti in [1] talks about entity retrieval, the task of returning entity from a knowledge base. The idea is to improve the performance of entity retrieval by employing information about the type of entities. They studied various type taxonomy frameworks from different knowledge bases such as Wikipedia, Dbpedia, Yago and Freebase. Both term based and type based similarities are employed in order to formulate a probabilistic retrieval approach. This formulation and a "target type oracle" that always gives the correct target type for a given query is used to evaluate the retrieval model against a term-based baseline approach in terms of MAP metric. The results suggest that the type aware retrieval performed best when used with a large type taxonomy such as Wikipedia with very specific types.

## 3rd Paper

Hasibi et al. in [2] proposed an the novel idea of using entity linking in the task of entity retrieval. The idea is to extend term-based entity retrieval model using a probabilistic approach that links entities in the query. This approach is based on the MRFs, makes it capable of encompassing a number of existing retrieval models. The new component of entity matching, named *Entity Linking incorporate Retrieval* (ELR), can be considered as an extension to MRF retrieval model. One of the key elements of ELR is incorporating entity annotations information into entity linking. TAGME is used to achieve this goal and to annotate the query terms. This work is evaluated using standard retrieval test collection of 500 heterogeneous queries, in terms of MAP. The results show a consistent improvements over all standard language models and state-of-the-art ad-hoc entity retrieval.

## Application to GoT Project

The quality of entity retrieval in a query-search scenario can be improved using the techniques explained in the papers. Some of these ideas are, using term dependencies, multi-fielded entity representation, probabilistic based language models (ordered, unordered, unigram, bigram). But, in GoT project, we do not have a query. Unless, we want to create a query out of each name in the form of "is  $j$ character $j$  dead?". Then, we can use the idea of multi-fielded entity representation in order to increase the accuracy of the retrieval process. But, I suspect this method only works if we have a big knowledge base with multiple node for each entity. But, one of the biggest challenges that we have in this project is with unlinkables. So, my guess is the improvements after applying this method is not going to be significant.

## References

- [1] Dario Garigliotti and Krisztian Balog. On type-aware entity retrieval. *Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval - ICTIR '17*, 2017.
- [2] Faegheh Hasibi, Krisztian Balog, and Svein Erik Bratsberg. Exploiting entity linking in queries for entity retrieval. In *Proceedings of the 2016 ACM International Conference on the Theory of Information Retrieval*, ICTIR '16, page 209–218, New York, NY, USA, 2016. Association for Computing Machinery.

- [3] Alexander Kotov and ChengXiang Zhai. Tapping into knowledge base for concept feedback: Leveraging conceptnet to improve search results for difficult queries. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, WSDM '12, page 403–412, New York, NY, USA, 2012. Association for Computing Machinery.
- [4] Nikita Zhiltsov, Alexander Kotov, and Fedor Nikolaev. Fielded sequential dependence model for ad-hoc entity retrieval in the web of data. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '15, page 253–262, New York, NY, USA, 2015. Association for Computing Machinery.