

Toward explainable embeddings-based and Pseudo relevance Feedback query expansions in Information Retrieval

M2R internship proposal: 1 person, 5 months.

Supervisors :

- Catherine Berrut, LIG
- Philippe Mulhem, LIG
- Jean-Pierre Chevallet, LIG

Many elements may be used to automatically expand users queries in a information retrieval system.

Among them, the documents of the corpus themselves are used by the "pseudo relevance feedback" process. Such process works well, but explaining why one expansion is achieved is not straightforward at all, as the expansions models are complex [Vaidyanathan 2015]. Other approaches tend to expand queries using neural networks learning-based (word embeddings) [Almasri et al. 2016], but such learning has many parameters and the expansion may contain words that do not appear in the corpus of documents. Others [Kuzi et al. 2016] take a part of both worlds by integrating words embeddings and relevance feedback techniques. In any case, there might be a large gap between the initial user query and the expended one, and none of these approaches above try to tackle the explanation of the expansion.

The goal of this work is to study the features of such expansions, and to deduce some ways to handle, at least partially, explanations about the expansions. The problem is largely open, and such explanations may benefit from cross-explanations between PRF and embeddings.

The expected work comprises three steps:

1. state of the art of query expansion, relevance feedback and pseudo relevance feedback, query expansions using words embeddings
2. proposal of a model of explaination for PRF expansions and embedding-based expansions
3. experiments on one or several classical corpora (Web, Health)

Bibliographic references:

[Kuzi et al. 2016] S. Kuzi, A. Shtok, and O. Kurland. 2016. Query Expansion Using Word Embeddings. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management/ (CIKM '16)*. ACM, New York, NY, USA, 1929-1932. DOI: <https://doi.org/10.1145/2983323.2983876>

[Almasri et al. 2016] M. AlMasri, C. Berrut and J.-P. Chevallet , A comparison of deep learning based query expansion with pseudo-relevance feedback and mutual information - ECIR 2016

[Vaidyanathan 2015] R. Vaidyanathan. A Study on Retrieval Models and Query Expansion using PRF. International Journal of Scientific and Engineering Research. 6. 2015.