

Relevance Feedback Search Based on Automatic Annotation and Classification of Texts

Rafael Leal, Joonas Kesäniemi, Mikko Koho and Eero Hyvönen

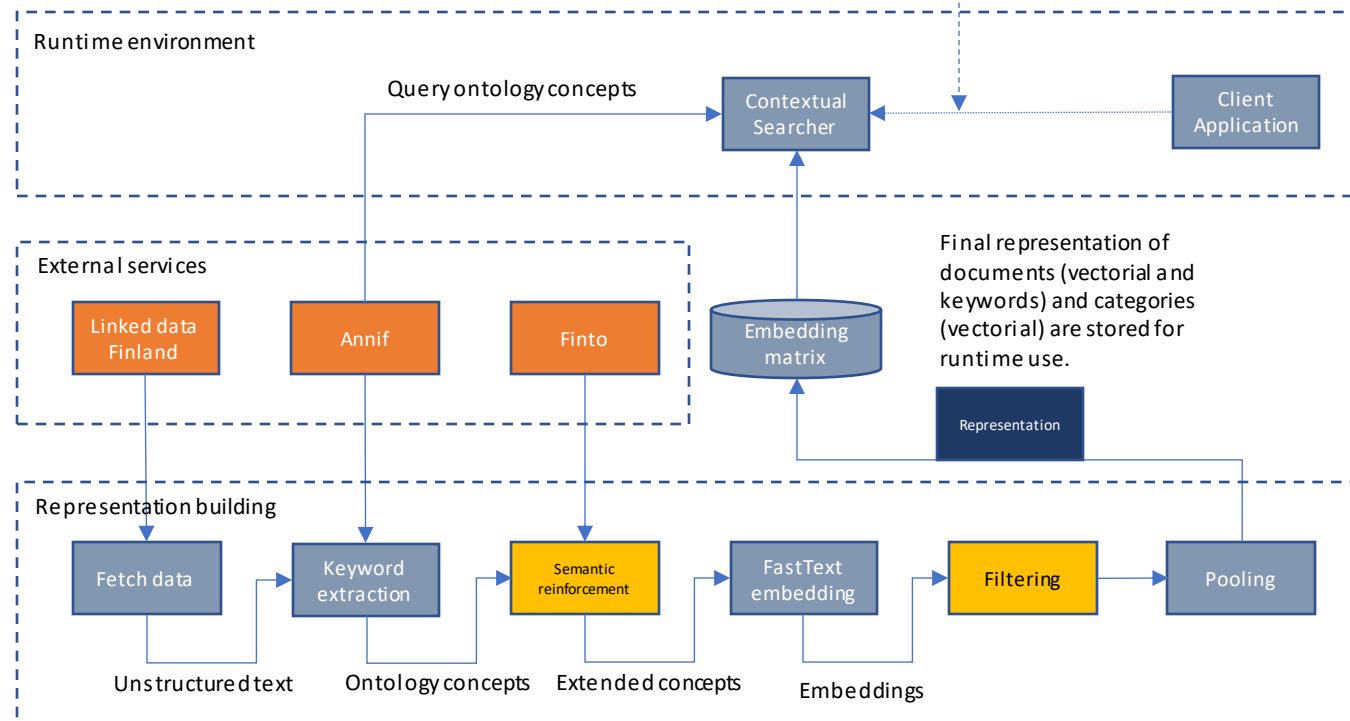
HELDIG – Helsinki Centre for Digital Humanities (University of Helsinki, Finland) and Semantic Computing Research Group – SeCo (Aalto University, Finland)

Problem

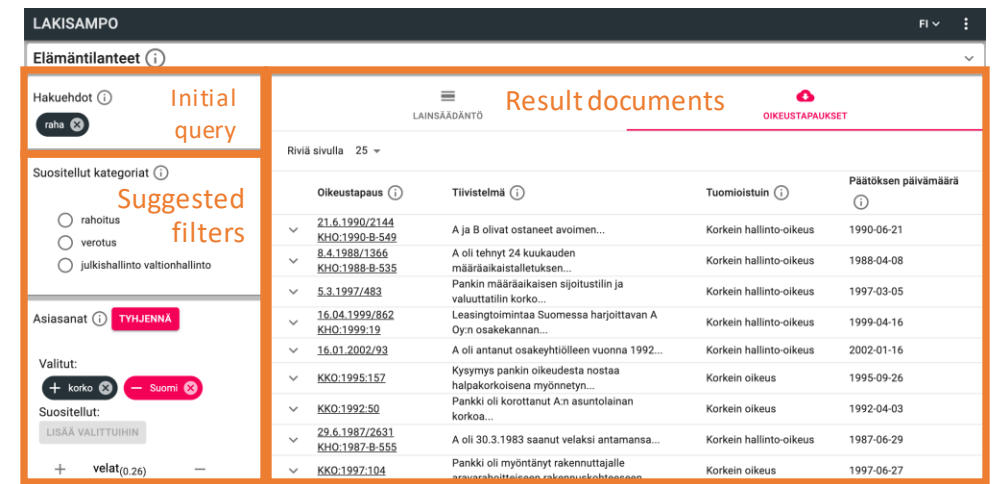
How to help users find appropriate search results in a domain they do not necessarily master?

Solution

- Relevance Feedback to help guide the user towards a satisfying search result
- Keyword (positive/negative) and category candidates for the user to choose
- Semantically-oriented: keyword extraction, word embeddings and unsupervised classification to store text representations and calculate results



Use case – LawSampo



The proposed solution was implemented as part of the LawSampo semantic portal, which is powered by a harmonized Linked Data knowledge graph about the Finnish legislation.

The Relevance Feedback functionality is part of the contextual search view that guides the user by suggesting document categories and relevant keywords to narrow down the results matching the initial search query.

Evaluation of the classification task

Two custom datasets (Yle and Minilex) and three different measures (F1, MRR and mean rank):

- **Yle dataset:** F1 = 0.742, MRR = 0.832 and Mean rank = 0.67
- **Minilex dataset:** F1 = 0.572, MRR = 0.705 and Mean rank = 1.496

Top prediction classified correctly in **57–74%** of the cases and among the top 3 predictions in **80–90%** of the cases. This variation may be due to the length of the documents and the quality of the category labels.