

# Planning report for thesis at IDA

## Author

Sijin Cheng, Computer Science

## Preliminary title

A Behavior-based Optimization of a Search Query for Patents

## Problem description

To catch the important trends in technology or materials, the company need to keep a continuous watch on the patent activities within their field. This is normally done via a search alert that signals the user when any new patents relevant to the search query are published. lamIP Platform is an innovative tool for patent search and patent management. lamIP's platform has a proprietary database with more than 100 million individual patent from around the world and they are adding up to 100.000 new patent publications every week. The platform provides excellent patent search possibilities. and with lamIP Search Alert, the customer can define specific searches in order to capture the most up-to-date, relevant competitive Intelligence for the business.

However, there are several factors influence the effect of the search alert now. The first is that choosing the right words to search with is difficult so that the search query cannot accurately locate the user's intent, the reason of it is that the structure of patent documents is complex and the use of a specific terminology is not always commonly known or used. The second, technologies change, new competitors enter the market, new materials appear, and consumer behaviour might change, too, which may be difficult to be observed by the set Search Alert. It is therefore desirable to continuously update and optimize the search query for a search alert, to adapt to observed changes in the response of the search alert owner to newly captured patents. In this paper, we are trying to optimize the existing search query by giving interactively query suggestions by learning the user behaviour model and present a ranked list of the patents identified using the new query.

## Approach

Find the approaches that might take to decide which new query to present

- 1) Attempt to extract key phrases from search results which user scored
- 2) Give the weight to phrases based on different scores (stars)
- 3) Use K-means algorithm to clustering the phrases and find the cluster center point C1
- 4) Process C1 as the new query Q1

Find the candidate patents

- 1) Search candidate patents P1 based on new query Q1
- 2) Attempt to extract key phrases from search results in the trash box

- 3) Use k-means algorithm to clustering the trash results and find the center point C2
- 4) Process C2 as the trash query Q2
- 5) Search candidate patents P2 based on Q2
- 6) Determine the final candidate patents P equals P1 minus P2

Find the approaches that ranked list of the patents using the new query

- 1) Calculating the similarity of the candidate patents and the new query Q1
- 2) Rank the candidate patents based on the similarity

Evaluation

- 1) Accuracy: precision, recall, f-measure
- 2) Calculate SSE (Sum of squared errors) using the new search results and center point C1 and C2.

## Literature base

[1]–[8]

- [1] M. Allahyari *et al.*, “A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques,” 2017.
- [2] K. Wagaj and P. S. Pingale, “A Literature Survey on Keyword Query Processing in Web Search Engines,” vol. 7, no. 2, pp. 471–473, 2016.
- [3] D. Roy, D. Ganguly, M. Mitra, and G. J. F. Jones, “Representing Documents and Queries as Sets of Word Embedded Vectors for Information Retrieval,” 2016.
- [4] R. C. Balabantaray, C. Sarma, and M. Jha, “Document Clustering using K-Means and K-Medoids,” *arXiv1502.07938 [cs]*, 2015.
- [5] F. O. Isinkaye, Y. O. Folajimi, and B. A. Ojokoh, “Recommendation systems: Principles, methods and evaluation,” *Egypt. Informatics J.*, vol. 16, no. 3, pp. 261–273, 2015.
- [6] T. Ruotsalo, G. Jacucci, P. Myllymäki, and S. Kaski, “Interactive intent modeling,” *Commun. ACM*, vol. 58, no. 1, pp. 86–92, 2014.
- [7] Z. Lu, W. J. Wilbur, J. R. McEntyre, A. Iskhakov, and L. Szilagyi, “Finding query suggestions for PubMed,” *AMIA Annu. Symp. Proc.*, vol. 2009, pp. 396–400, 2009.
- [8] S. Bhatia, D. Majumdar, and P. Mitra, “Query suggestions in the absence of query logs,” *Proc. 34th Int. ACM SIGIR Conf. Res. Dev. Inf. - SIGIR '11*, p. 795, 2011.

# Time plan

beamgent

Printed: 02/16/2018

