CIKM'2014
2014 ACM International Conference on Information and Knowledge Management

Reviews For Paper

Track	Information Retrieval
Paper ID	206
Title	A Study of Query Reformulation for Patent Prior Art Search with Partial Patent Applications

Masked Reviewer ID: Assigned_Reviewer_1

Review:

Question	
Overall Rating	Accept
Top 3 Strengths	The paper deals with patent IR. A study of query reformulation for patent prior art search is performed where parts of the patents like title, abstract, claims, description can be used as queries. Two approaches of query reformulation are used: query expansion and query reduction. A term overlap analysis points out the mismatch between queries and documents and this should be alleviated by query reformulation. In chapter 3 evaluation principles and experimental setup with use of the CLEF-IP 2011 and 2010 datasets are shown. Based on previous approaches seven query expansion methods were implemented but none of them achieved significant improvements over the baseline. Concerning the adaption of specific options some term selection methods improved performance but this is influenced by the queries. Concerning query pruning it seems to be helpful to reduce the set of query terms. There are some very interesting findings which can be deducted from the experiments.
Top 3 Weaknesses	The authors mention that they work with new query expansion and reduction frameworks (MMRQE & MMRQR). Please point out more clearly what is new? They seem very much inspired by usual techniques. In my opinion it is not clear that one special query expansion technique will improve performance. It seems to depend on the type of terms and queries. What principles can we learn for these experiments for further work? Formal: Please have a look on the last paragraphe of 2.1. The figures are very small (e.g. Fig. 4) and therefore hard to read.
Detailed Comments	I like the paper because it gives insight into a big experiment series in the patent domain. I think work in this context is very important because the current solutions are far from being satisfactory. In terms of recall and precision no real solutions exist.
Author feedback needed?	No
What specific feedback do you like the authors to provide?	No

Masked Reviewer ID: Assigned_Reviewer_2

Review:

Question	

Overall Rating	Accept
Top 3 Strengths	1) The paper is very clear in presenting the methods applied for QE and QR, 2) comprehensiveness, the paper neatly presents all the necessary background information and also the details that are useful to understand better the motivation and the objectives of this research work. 3) the comparative analysis of existing and novel methods for query expansion and reduction explores different facets of the overall problem and the methods applied for patent prior-art search
Top 3 Weaknesses	1) I think the experimental setting should have been better designed without applying the IPC filtering step (sentence: "We also used the patent classification... as suggested in previous works."). The authors specifically aim to address the prior-art search problem in the context of an incomplete patent application. In this context, it is most likely that IPC codes have not been thoroughly considered or even more than this have not been assigned yet.
Detailed Comments	This paper focuses on prior art search with queries that represent unfinished patent applications. This is an interesting approach and could be a very useful feature of prior-art patent search systems since evaluating the patentability of an idea at the earliest possible stages will normally be time and money saver. Both query expansion and reduction that exploit patent structure and term diversification are examined and they present a comprehensive comparative analysis of existing and novel methods for query expansion and reduction in patent prior-art search. Overall the paper presents a very useful study of how query can re-formulated (using QE or QR) when conducting prior-art patent search. "...between queries an relevant documents." -> "...between queries and relevant documents." (d is missing)
Author feedback needed?	No
What specific feedback do you like the authors to provide?	No

Masked Reviewer ID: Assigned_Reviewer_3

Review:

Question	
Overall Rating	Neutral
Top 3 Strengths	Extensive experimentation
Top 3 Weaknesses	The paper introduces query expansion methods, but does not compare it with other state-of-the-art methods The evaluation over patent retrieval is not compared with stat-of-the-art in this domain The improvement, if exists, is small and incremental
	The paper introduces a method (or two variations thereof) for query reformulation. The methods are tested on patent prior art search. Although the authors test multiple variations over the approach, it fails to compare the query reformulation approach with state-of-the-art approaches to query reformulation. At the same time, it also fails to compare the results with state-of-the-art prior art search. The

Detailed Comments	experiments test many variations that include so many specialized parameters and settings that I am left wondering if the approach is actually significant, even if it were to be completely original. It appears to make a small difference. The authors are advised to separate the testing of query expansion from the testing of prior art search. Each can and should be tested in isolation against state of the art methods. As it is, the only system tested against for pseuso relevance is Rocchio (and it has numerous variations in any case.)
Author feedback needed?	No
What specific feedback do you like the authors to provide?	No