Semantic-based medical records retrieval via medical-context aware query expansion and ranking.

ABSTRACT

Efficient retrieval of medical records involves contextual understanding of both the query and the records contents. This will enhance the searching effectiveness beyond merely keyword matching and is assisted by analyzing its semantics notion such as by the utilization of the MeSH thesaurus .The query is annotated and expanded by information from the deep medical contextual understanding.This is because typically medical records contain medical terminologies which may not be included in the user query but is important for accurate search hit. Besides, the terminologies have synonyms which should be utilized for richer and expanded query.The main contribution of the paper is the semantic-based retrieval technique by utilizing context-aware query expansion and search ranking method. Medical domain is chosen as a proof of concept and a medical record retrieval application was developed.The source of medical records are obtained from the ImageCLEF 2010 dataset which also houses a series of evaluation campaign such as photo annotation, robot vision and Wikipedia retrieval. This paper addresses the following problems: (i)semantic-based query expansion technique which increase the content awareness ability,(ii) MeSH- manipulated indexer which entails medical terminologies and their synonym,(iii) adoption of extended Boolean matching to measure similarity between query and documents, and (iv)ranking method which prioritizes matched expanded query size.The results were measured using precision, recall and mean average precision (MAP) score.Comparing against other approaches, our method has several achievements including; (i) more efficient access of MeSH thesaurus through the manipulated indexer compared to its original form; (ii) enrichment of query expansion using synonym term can improve mean average precision (MAP) value as opposed to standard query expansion; (iii) our comprehensive ranking method achieved high recall. According to MAP score we are in the top five run system amongst submitted run systems in ImageCLEF2010 medical task.

Keyword: Clinical record retrieval; Semantic; Query expansion; Ranking method.