Abstract

Many celebrities and experts from various fields may have been referred by not only their personal names but also by their aliases on web. Aliases are very important in information retrieval to retrieve complete information about a personal name from the web, as some of the web pages of the person may also be referred by his aliases. The aliases for a personal name are extracted by previously proposed alias extraction method. In information retrieval, the web search engine automatically expands the search query on a person name by tagging his aliases for complete information retrieval thereby improving recall in relation detection task and achieving a significant mean reciprocal rank (MRR) of search engine. For the further substantial improvement on recall and MRR from the previously proposed methods, our proposed method will order the aliases based on their associations with the name using the definition of anchor texts-based co-occurrences between name and aliases in order to help the search engine tag the aliases according to the order of associations. The association orders will automatically be discovered by creating an anchor texts-based co-occurrence graph between name and aliases. Ranking support vector machine (SVM) will be used to create connections between name and aliases in the graph by performing ranking on anchor texts-based co-occurrence measures. The hop distances between nodes in the graph will lead to have the associations between name and aliases. The hop distances will be found by mining...
the graph. The proposed method will outperform previously proposed methods, achieving substantial growth on recall and MRR.

References


Computer Science

Index Terms
Information Retrieval

Keywords
Anchor Text Mining  Graph Mining  Word Co-occurrence Graph