Named Entity Search in Enterprise

Mohan Yang
UCLA CS Department
yang@cs.ucla.edu

Tao Cheng
Microsoft Research
taocheng@microsoft.com

Manoj Syamala
Microsoft Research
manojsy@microsoft.com

ABSTRACT

Documents in enterprise usually contain various named entities (e.g., people name, product name, company name, etc). However, today's enterprise search engines don't support searching over these named entities in documents. Toward building an efficient and accurate named entity search prototype for enterprise search, we studied co-occurrence based approaches for named entity search. The basic assumption behind these approaches is that words co-occurring with entity provide related information about this entity. Our first approach builds a profile for each entity using text co-occurring with that entity, and searches entities using their profiles. Our second approach finds related entities to a keyword query using online aggregation of windows, where a window contains all co-occurring words with an entity within an interval. We demonstrate the effectiveness of our prototype over a 3TB corpus from the Microsoft Web.