



Related Entity Finding Task

Goal

our goal is to construct logical hierarchical sitemap to integrate the hierarchical information into existing search engine to get the relevant pages which have the results residing in tables or lists especially by exploiting site level knowledge

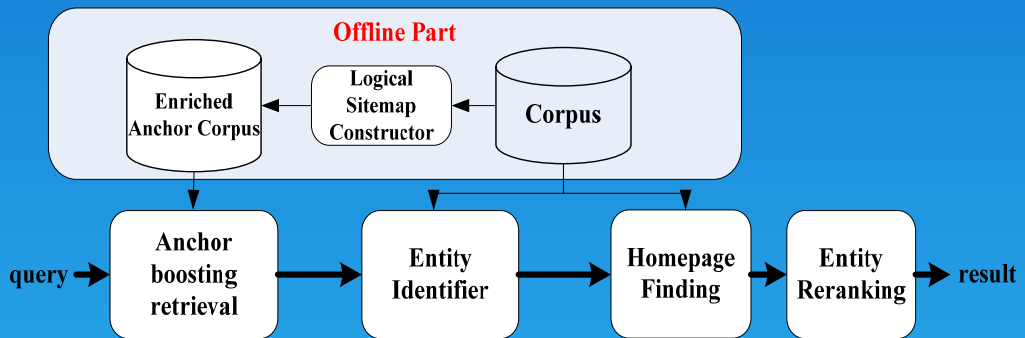
Step 1. Query Topic Parsing

Use Open Ephyra to analyze question syntactically and semantically with the explicitly specified target entity type.

Step 2. Named Entity Identification

Employ OpenEphyra's NETagger to complete named entity identification.

System Overview



Step 3. Related Entity Candidates Ranking

$$Er = \sum_i P_i \left(\sum_{e \in E} occurrence(e) \right)$$

Apply the formula to rank related entity candidates. P refers to Web page which the entities are extracted from. E refers to the set of extracted entities. Pi(Occurrence(e)) refers to the occurrence of e extracted from Web page Pi.

Step 4. Entity Homepage Finding and Ranking

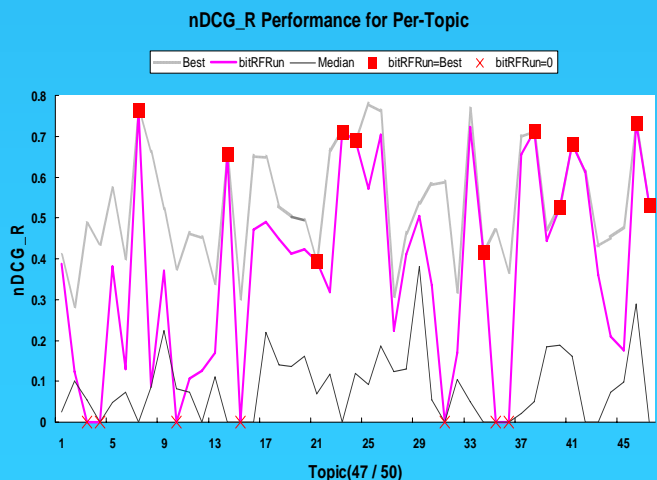
The procedure is to identify the extracted entities' homepages. We consider five sources for finding homepages ordered by confidence.

Step 5. Logical Hierarchical Sitemap

Develop a method to extract hierarchical information in pages, construct a logical sitemap and collapse hierarchical information into a categorized string to index an enriched anchor repository



Experimental Results



E.g. <http://www.mdfarmbureau.com/>