

Apparatus and Method for Disambiguating a Plurality of Entity Mentions in a Document

Abstract

Apparatus and method for disambiguating a plurality of entity mentions in a document, the method comprising: parsing a document to obtain a plurality of entity mentions; determining matched entity mention or mentions by matching each of the plurality of entity mentions with each entity in a database; determining a score for each matched entity mention based on number and type of links the matched entity mention has with other matched entity mentions; determining the matched entity mention with highest score as a hub entity, which is deemed as a disambiguated entity mention. If there are more than one matched entity mentions, the method calculates, for each matched entity mention, a weighted sum of weight values for one or more entities that are n hops away from the corresponding matched entity in the database, wherein the weight value of the more than one entities was predetermined based on entity type.

(Figure 1)