Author Name Disambiguation for Biomedical Literature

Authors : Parthiban Srinivasan

Abstract : PubMed provides online access to the National Library of Medicine database (MEDLINE) and other publications, which contain close to 25 million scientific citations from 1865 to the present. There are close to 80 million author name instances in those close to 25 million citations. For any work of literature, a fundamental issue is to identify the individual(s) who wrote it, and conversely, to identify all of the works that belong to a given individual. Due to the lack of universal standards for name information, there are two aspects of name ambiguity: name synonymy (a single author with multiple name representations), and name homonymy (multiple authors sharing the same name representation). In this talk, we present some results from our extensive work in author name disambiguation for PubMed citations. Information will be presented on the effectiveness and shortcomings of different aspects of successful name disambiguation such as parsing, validation, standardization and normalization.

Keywords : disambiguation, normalization, parsing, PubMed
Conference Title : ICDMKM 2016 : International Conference on Data Mining and Knowledge Management
Conference Location : Madrid, Spain
Conference Dates : March 24-25, 2016