

The Role of Named Entity Recognition for Information Extraction

Authors : Girma Yohannis Bade, Olga Kolesnikova, Grigori Sidorov

Abstract : Named entity recognition (NER) is a building block for information extraction. Though the information extraction process has been automated using a variety of techniques to find and extract a piece of relevant information from unstructured documents, the discovery of targeted knowledge still poses a number of research difficulties because of the variability and lack of structure in Web data. NER, a subtask of information extraction (IE), came to exist to smooth such difficulty. It deals with finding the proper names (named entities), such as the name of the person, country, location, organization, dates, and event in a document, and categorizing them as predetermined labels, which is an initial step in IE tasks. This survey paper presents the roles and importance of NER to IE from the perspective of different algorithms and application area domains. Thus, this paper well summarizes how researchers implemented NER in particular application areas like finance, medicine, defense, business, food science, archeology, and so on. It also outlines the three types of sequence labeling algorithms for NER such as feature-based, neural network-based, and rule-based. Finally, the state-of-the-art and evaluation metrics of NER were presented.

Keywords : the role of NER, named entity recognition, information extraction, sequence labeling algorithms, named entity application area

Conference Title : ICALELT 2023 : International Conference on Applied Linguistics and English Language Teaching

Conference Location : Los Angeles, United States

Conference Dates : October 30-31, 2023