

Algorithm name detection in computer science papers

In this project, you have to extract unique algorithm names from a big corpus of published computer science research/journal/conference articles. Then try to locate or identify the original work where it was first published. Finally categorize or cluster the different algorithm names into various subdomains of computer science. You will use context based learning and semisupervised learning techniques.

Tasks:

Identify algorithm names in individual research papers, using modified Named Entity Recognition techniques,

Filter out noise like university names, places, author names, names of datasets etc,

Create a seed list of true positives and falsepositives for algorithm names,

Use algorithm name cooccurrence context information and sentence level context, information to get representations for each algorithm name,

Given a new named entity, use some metric to identify whether this is an algorithm or not use information from previous step to determine,

Classify / cluster algorithms into several categories.