AP Project Report

DBLP Query Engine

Project Contribution:

Roll No.	Name	Contribution
2015085	Sanidhya Singal	GUI, Backend Coding
2015063	Pranav Nambiar	Parsing, Backend Coding

Entity Resolution:

We are using the data in the <www> tag of the dblp.xml file given. It contains author names that are known by various aliases (synonyms). We map the given aliases with a tree-map to the original primary name. When we encounter a name, we check if it is an alias and if it is, we replace the alias with the primary name. Similarly, we look for the initials of an author and try to match them with the author's name.

Prediction:

We store the number of papers published by each author in a tree of trees. When we are given a range of years to work upon, to predict the number of papers published, we take the segment of tree specified and apply linear regression on that segment and predict our answer based on that segment.