

DBLP Project Report

Motivation: Out of DBLP and FMS, we chose DBLP because of its complexity. And it is doubtless we have learned a lot by doing this project.

Use Case and class Diagram (LAB 7):

[Diagrams \(LAB 7\)](#)

Disclaimer : *several optimizations has been done since then to increase efficiency and reduce query time.*

Native Libraries used:

1. SAX Parser

Entity resolution:

In entity resolution , firstly we are splitting the string by spaces and storing them into an array . So , after splitting what we got is first name , middle name and last name separately . Now we are checking if the size of both the array formed is same then we check the same indices if they satisfy the condition that is either they must be directly equal or one of them must be the initial of another. If the size of both the array formed is not equal then i check if every string of small length must satisfy the condition of larger string ,if they do so then they are same otherwise different .

Prediction :

We sort of used knn algorithm for prediction .we used latest 10% of data of what we are provided to predict how much more publishes can be done after year limit. Basically, it is specialised mean of the given data that has been used for prediction of total number of publishables.

Contributions:

Saurabh Kumar : 2015088

1. GUI (except for query3)
2. Parser
3. Query1 and Query2
4. Exception Handling
5. Doxygen Documentation.

Prashant : 2015072

1. Query3
2. GUI for Query3
3. Entity resolution
4. Use Case and Class Diagram for project(lab 7)