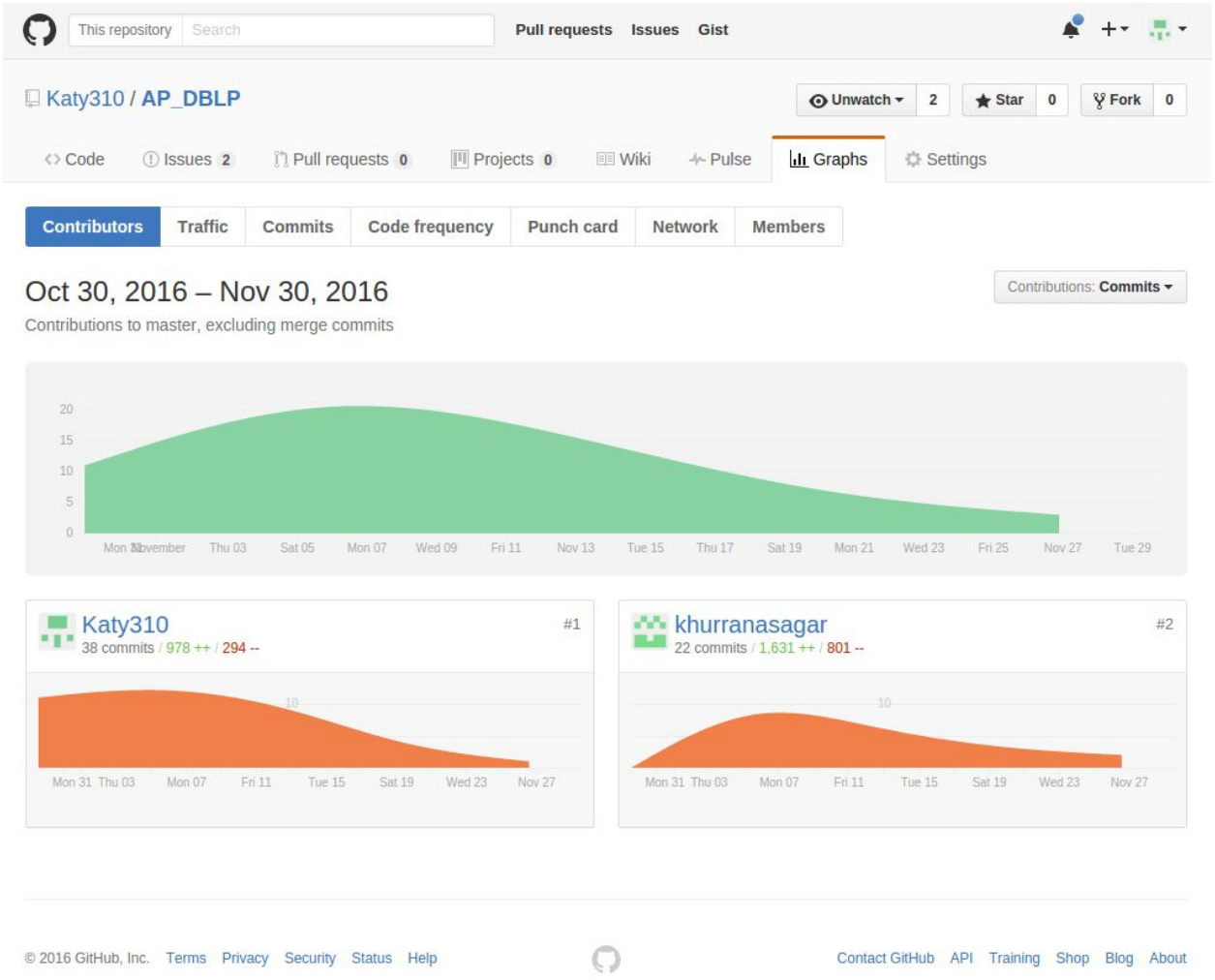
DBLP Query Engine

Individual Contributions



Katyayani Singh

- GUI.java
    1. Used JSplitPane to create the required UI
    2. Managed interactions between all screens
- Person.java

Sagar Khurana

- Parser.java
    1. Used Stax Parser to parse the file
    2. Handled queries
- Publication.java

Entity Resolution

We created a HashMap which stores the author name as the key and the number of publications as value without resolving the author names.

We have an ArrayList of authors which contains objects of Person type. Each Person object has a collection for author names which are defined by reading each entity of www homepage records.

For Query 1, if the name input by the user exists in the ArrayList of Persons, then we search for the names present in Publication entities of dblp.

For Query2, we go through each person record and fetch the names of the authors. For each author, we fetch the number of Publications from the HashMap.

Prediction

We used linear regression to predict the number of publications of the following year.

The regression equation for y on x is: ***y = bx + a*** where ***b*** is the slope and ***a*** is the intercept (the point where the line crosses the y axis)

$$b = \frac{\sum d_x d_y}{\sum d_x^2} \qquad a = \bar{y} - b\bar{x}$$

Now calculate $\sum d_x^2 = \sum x^2 - \frac{(\sum x)^2}{n}$

Calculate $\sum d_y^2 = \sum y^2 - \frac{(\sum y)^2}{n} = 0.$

Calculate $\sum d_x d_y = \sum xy - \frac{\sum x \sum y}{n}$ (this c