**Project Topic**

Develop an algorithm to detect and suggest the tags to be specified to a question posted on Stack Overflow based on the content present in it using natural language processing.

**Team Members**

* + Praseed M
  + Nizamudheen T I
  + Paul M David
  + Joby John
  + Kiran K B
  + Nandu A

**Description**

Stack Overflow is something which every programmer use one way or another. Each month, over 50 million developers come to Stack Overflow to learn, share their knowledge, and build their careers. It features questions and answers on a wide range of topics in computer programming. The website serves as a platform for users to ask and answer questions, and, through membership and active participation, to vote questions and answers up or down and edit questions and answers in a fashion similar to a wiki or Digg. As of April 2014 Stack Overflow has over 4,000,000 registered users, and it exceeded 10,000,000 questions in late August 2015. Based on the type of tags assigned to questions, the top eight most discussed topics on the site are: Java, JavaScript, C#, PHP, Android, jQuery, Python and HTML.

**Data Source**

The data required for training and testing our algorithm can be accessed from the link provided below.

<https://www.kaggle.com/c/facebook-recruiting-iii-keyword-extraction/>

[Data size: 3GB](https://www.kaggle.com/c/facebook-recruiting-iii-keyword-extraction/)

**Methodology**

We intend to implement the algorithm in python programming language using different types of classifier algorithms to pinpoint the most effective and efficient one to enable in our program.

Database management will be achieved with hadoop to maximize resource utilization.

The extracted data will later be preprocessed to optimize it for the specific algorithm.

**Output Description**

The output of the algorithm will be in the form of tags assigned to each question by the algorithm.