Predicting Tags for Stack-overflow Questions

**Introduction**

In Stack-overflow website while a user asks a question, one will be given an option to choose at most 5 tags regarding what he is asking. User has to select these tags manually. Tags helps the website to show appropriate questions to those who are interested in that particular topic *[Tags define a topic of interest. For Eg: java – about java language, multithreading – about multithreading independent of the language one is using etc.]*, thus targeting the right people who could answer it. As the user is in control of the tags, he may choose tags which may or may not be relevant to a question. New users may not know how to tag a question to reach maximum audience or sometimes a user may choose a tag which is relevant but there could be a tag which is more appropriate which he didn’t know existed *[There are nearly 40K tags in Stack-overflow]*. So, to help a user in making his question to reach wide & appropriate audience and to increase the probability of getting it answered we can predict tags while a user is writing it. Thus making the user aware of tags available and giving him a chance to select could help the website community in total. Our project is trying to address this issue by predicting the tags for a given question.

Our approach has two models. One is to use a SVM with Quadratic Kernel (a supervised learning method) and other to use Clustering (an unsupervised learning approach).

**Problem Definition and Algorithm**

Task Definition: If given a question which is about to be posted our program will