

# Recommendation System for StackOverflow Community

Arpita Sheth (110422828)

Shreya Bhatia (110619150)

## Introduction

Questions Everywhere? We need Answers! Stack Overflow is a question answer community very widely used for answering questions targeting developers and software engineers. It has accumulated large volumes of knowledge through the voluntary services of people across the globe. By analysing this data, we propose to recommend to the user, similar solved queries and posts which have the most relevant answers.

## Motivation

### *Faster!*

People need help and it is time saving if the question has been already asked and answered by someone else. We propose to provide an interactive platform that recommends similar previously answered posts to the user. We also propose to present the visuals of most recent statistics in the Stack Overflow data to help identify the current trends and popular topics.

## Goals

- Allow users to enter query via an interactive platform and get recommendations for their queries.
- Create a dashboard to find previously answered similar posts and find the most similar query with the best answer.
- Identify and visualise the most relevant statistics and trends about the Stack Overflow questions.

## Proposed Methods

- Use D3.js and Flask for visualising the recommendation system results and trends.
- Topic Modelling, LDA - For extracting topics from questions, answers and comments.
- Text Analysis - using tag matching along with query matching and distance similarity metrics.

## Data Specifics

Stack Exchange publicly provides Stack Overflow data in XML format. The data set contains XML documents : badges.xml, comments.xml, posts.xml, users.xml, votes.xml. Posts.xml.

Posts.xml contains actual content of posts, user\_id, post\_type etc.

We will be focusing on questions in academia, programming areas.

## Evaluation Method

We will be testing and calculating accuracy of the final model on a pre-tagged test data.

## References

- [1] Stack Overflow public data, Retrieved from: <https://archive.org/details/stackexchange>
- [2] A. Barua, S. W. Thomas and A. E. Hassa. An analysis of topics and trends in Stack Overflow. 2012
- [3] S. Wang, D. Lo and L. Jiang. An Empirical Study on Developer Interactions in StackOverflow. SAC '13
- [4] M. Asaduzzaman, A. S. Mashiyat, C. K. Roy and K. A. Schneider. Answering questions about unanswered questions of stack overflow. MSR '13