**Code Search**

**IR Project**

**>> Team Details**

1. Nishant Shinde - S20210010161
2. Omkar Khairnar - S20210010120

**>> Tasks**

1. Data Collection  
   – Web crawling ([stackoverflow](https://stackoverflow.com/questions/tagged/javascript))
2. Preprocessing
3. Creating Inverted Index  
   – Posting list for all terms in data
4. TF-IDF  
   – Calculation of IDF using inverted index  
   – Calculation of log-term frequency for terms w.r.t documents  
   – Calculation of TF-IDF of document terms by multiplying above two calculations
5. Query processing  
   – Preprocessing query  
   – Calculation of TF-IDF of query
6. Finding Similarity Coefficient  
   – Use of TF-IDF of documents and query to find similarity coefficient with cosine-similarity function  
   – Retrieval of the document to display from document\_ids which are ranked top 20
7. Building UI  
   – Creation of UI with Django  
   – Implementation of function to send query and receive documents  
   – Implementation of function to send feedback and receive Precision Recall curve
8. Evaluation  
   – Displaying Precision-Recall curve based on the user feedback.