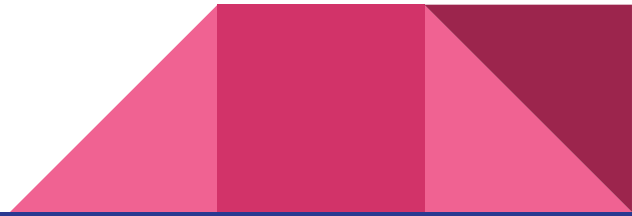
Ranking Based Question Answering System

Agenda

- Introduction
- Related Work
- Proposed Technique and Algorithm
- Implementation
- Work Update



Introduction

Question Answering system is an automated computer system that answers queries questioned by humans in natural language.

Related Work

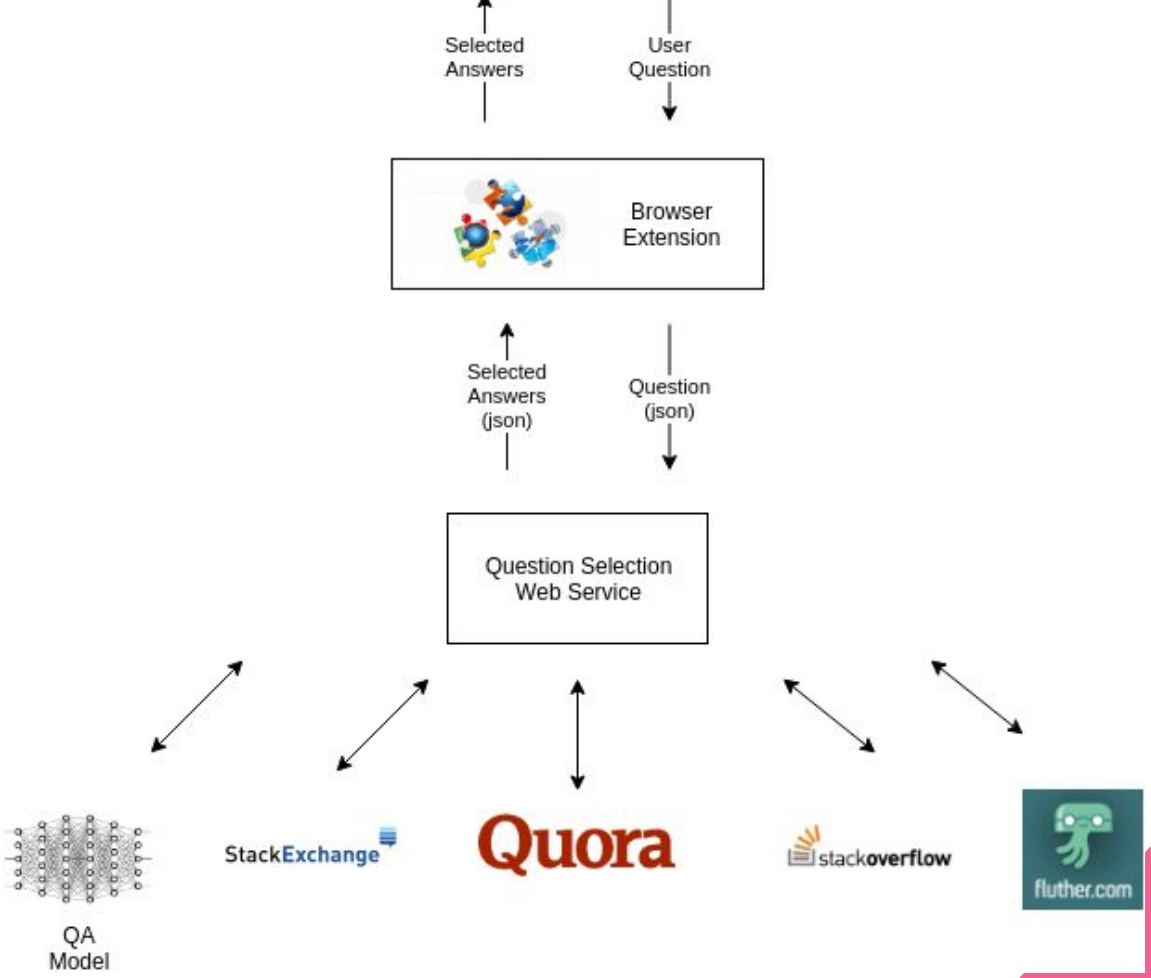
Many user based question answering systems available on the web:

- Quora
- Stack Overflow
- Stack Exchange
- Yahoo Answers
- Flutter

Proposed Technique

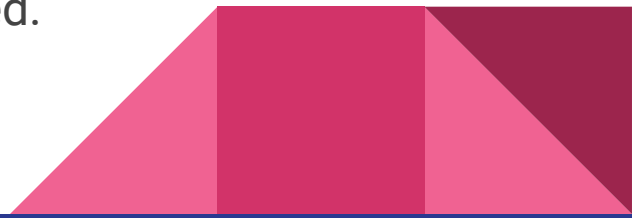
The architecture of the system comprises of 4 modules:

1. Chrome/Firefox Extension
2. Question Selection Web Service
3. Question Answering System
4. Question Classification Model


Selected Answers
User Question
Browser Extension
Selected Answers (json)
Question (json)
Question Selection Web Service
QA Model
StackExchange
Quora
stackoverflow
fluther.com

Dataset

- The Yahoo! Answers topic classification dataset is a human labelled dataset constructed using 10 largest main categories.
- Each class contains 140,000 training samples and 6,000 testing samples. Therefore, the total number of training samples is 1,400,000 and testing samples 60,000 in this dataset.
- From all the answers and other meta-information, only the best answer content and the main category information were used.



Classes for classification

1. Society & Culture
2. Science & Mathematics
3. Health
4. Education & Reference
5. Computers & Internet

Classes for classification (Contd.)

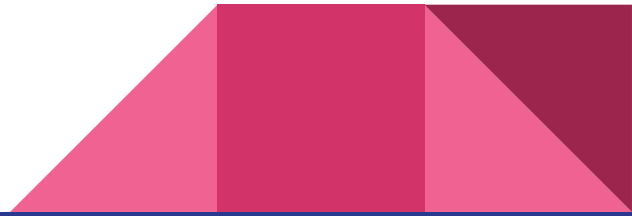
6. Sports

7. Business & Finance

8. Entertainment & Music

9. Family & Relationships

10. Politics & Government

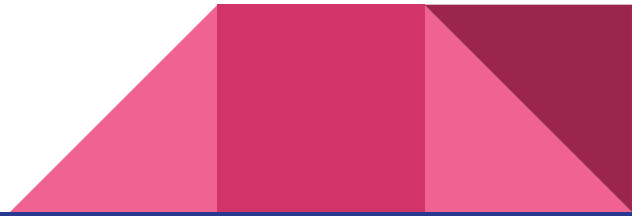


Implementation Details

1. Tokenization
2. Stop words removal
3. Lemmatizing with NLTK
4. Measuring the Cosine Similarity

Question Classification - Approach

1. Text Exploration
2. Text Cleaning
3. Obtaining POS Tags, Identifying Named Entities, Lemmas, Syntactic Dependency Relations and Orthographic Features.
4. Using the obtained properties as Features.
5. Using a Linear SVM model on the engineered features.



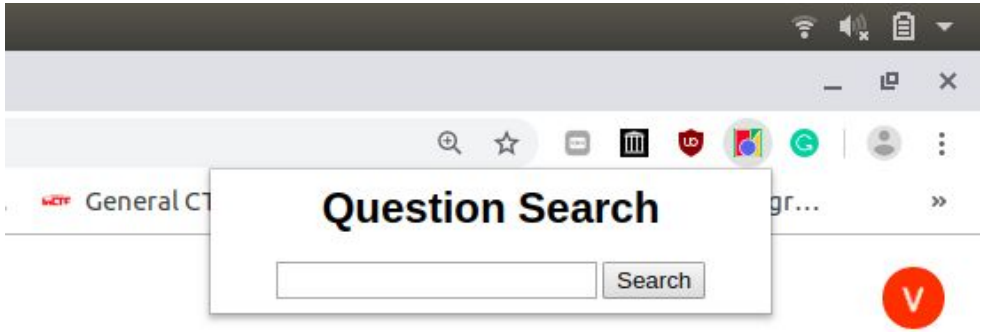
Model

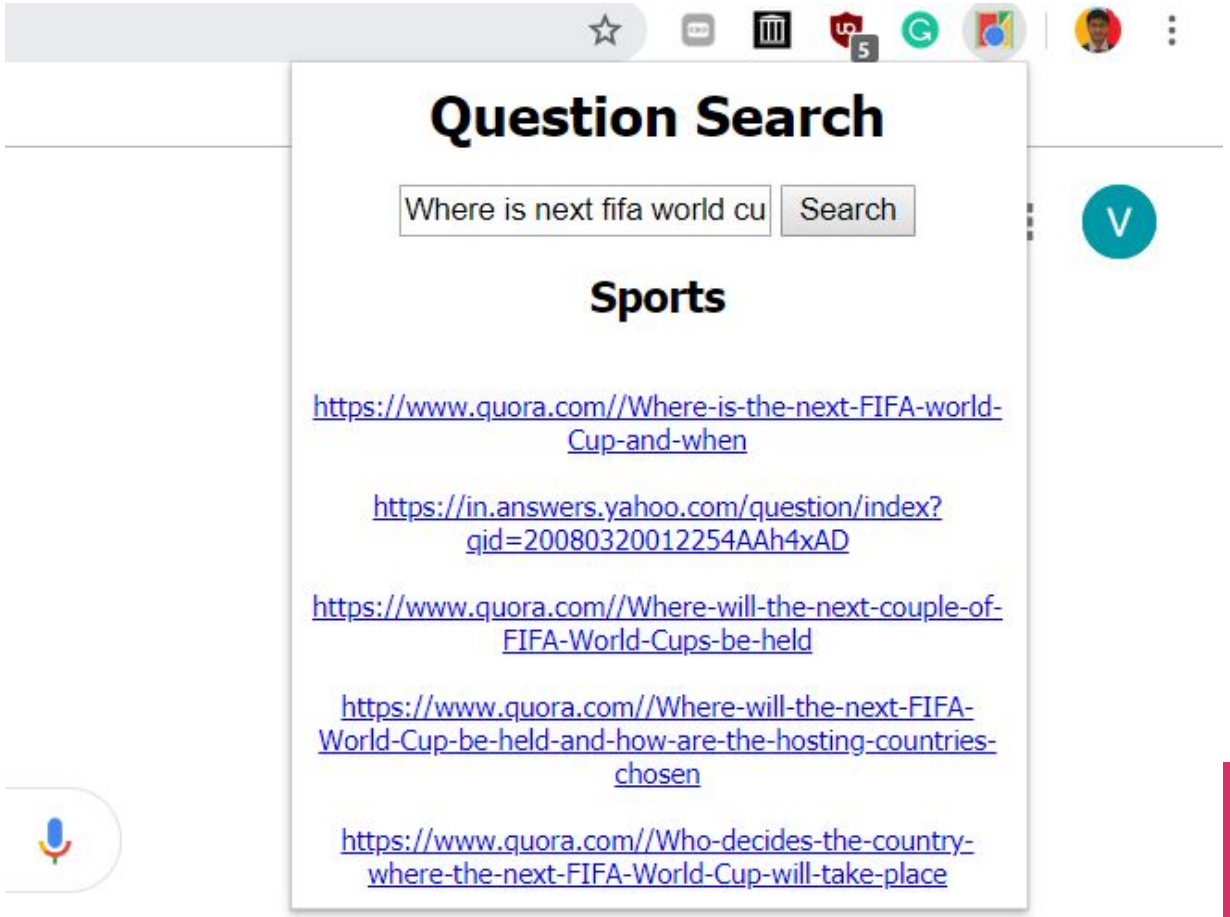
Linear Support Vector Machine Classifier

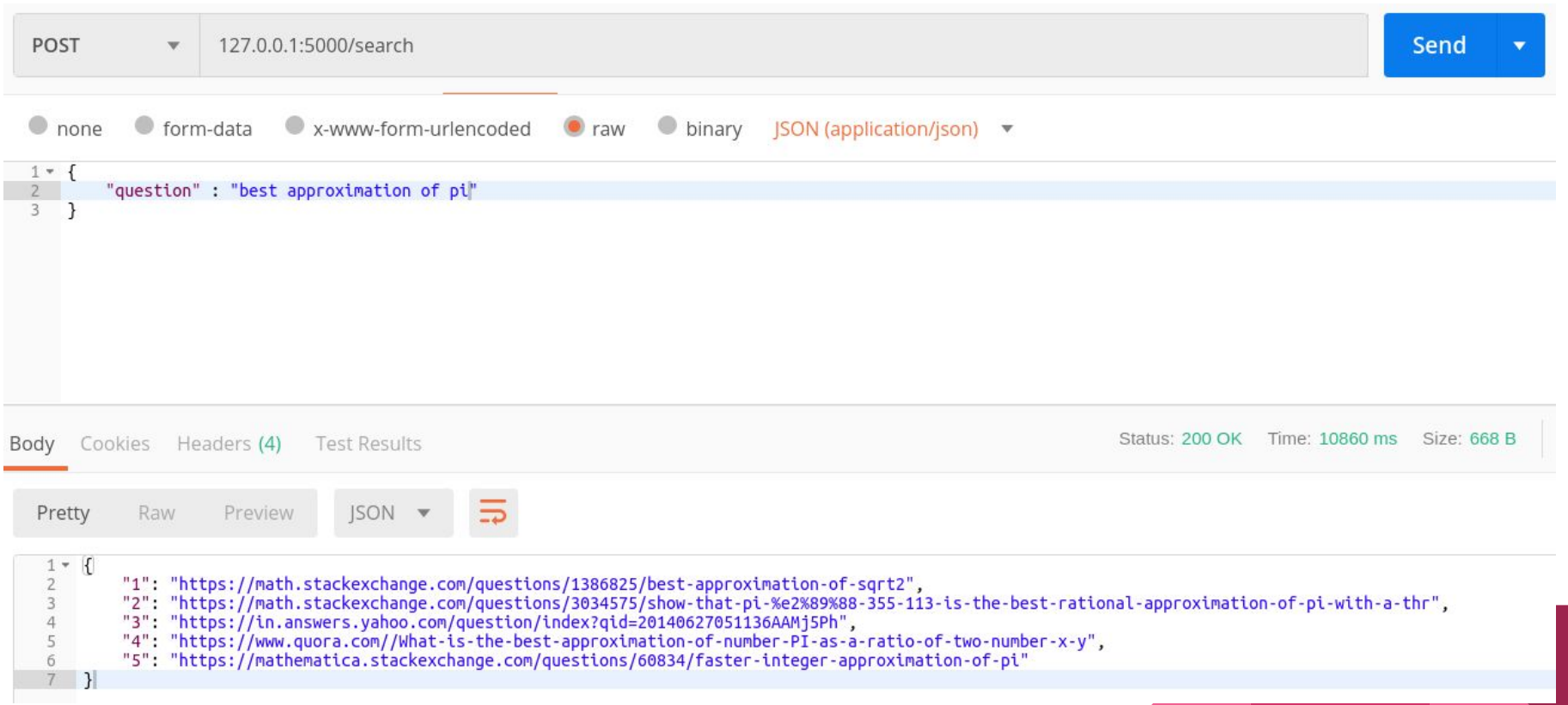
Features used: Named Entity Recognition + Lemmas + POS Tags

Accuracy: 66.316%

Implementation

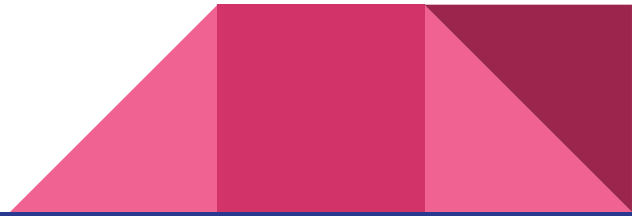



Question Search
Where is next fifa world cu
Search
Sports
https://www.quora.com//Where-is-the-next-FIFA-world-Cup-and-when
https://in.answers.yahoo.com/question/index?qid=20080320012254AAh4xAD
https://www.quora.com//Where-will-the-next-couple-of-FIFA-World-Cups-be-held
https://www.quora.com//Where-will-the-next-FIFA-World-Cup-be-held-and-how-are-the-hosting-countries-chosen
https://www.quora.com//Who-decides-the-country-where-the-next-FIFA-World-Cup-will-take-place


POST
127.0.0.1:5000/search
Send
none
form-data
x-www-form-urlencoded
raw
binary
JSON (application/json)
{
"question" : "best approximation of pi"
}
Body
Cookies
Headers (4)
Test Results
Status: 200 OK
Time: 10860 ms
Size: 668 B
Pretty
Raw
Preview
JSON
{
"1": "https://math.stackexchange.com/questions/1386825/best-approximation-of-sqrt2",
"2": "https://math.stackexchange.com/questions/3034575/show-that-pi-%e2%89%88-355-113-is-the-best-rational-approximation-of-pi-with-a-thr",
"3": "https://in.answers.yahoo.com/question/index?qid=20140627051136AAMj5Ph",
"4": "https://www.quora.com//What-is-the-best-approximation-of-number-PI-as-a-ratio-of-two-number-x-y",
"5": "https://mathematica.stackexchange.com/questions/60834/faster-integer-approximation-of-pi"
}

Work Update

1. Developed Chrome Web Extension
2. Developed Question Selection Web Service
3. Selected most relevant Questions from QA systems using the developed web service.
4. Question similarity matching using cosine similarity
5. Applied Classification techniques to find the domain of the asked question



Future Goals

1. Deploying and Integrating Context based QA model with the above developed system
2. Applying deep learning techniques such as Siamese Manhattan LSTM for similarity matching of questions

THANK YOU

- Vishal Mittal (2017A7PS0080P)