PROJECT SYNOPSIS



# **Title**: Multi domain question answering system

# **Guide**: Ms. P D More

# **Experts**:

* Ms. V A Mishra
* Ms. K H Wanjale

# **Team:**

|  |  |  |
| --- | --- | --- |
| **Krishna Purohit** | **422030** | [**purohitkrishnar@gmail.com**](mailto:purohitkrishnar@gmail.com) |
| **Akash Varma** | **423064** | [**akash.varma095@gmail.com**](mailto:purohitkrishnar@gmail.com) |
| **Aanshi Verma** | **422068** | [**aanshiverma03@gmail.com**](mailto:purohitkrishnar@gmail.com) |
| **Shaaz Shaikh** | **422048** | [**shaaz.shaikh@viit.ac.in**](mailto:purohitkrishnar@gmail.com) |

# 

# OBJECTIVE

This project is our effort to develop a multi-domain question answering system based on natural language processing which can answer concise and accurate answers of questions related to multiple major domains such as politics, business and economics, science and technology, sports and entertainment. To put a upper bound on the scope of the project, we have decided to take only those topics in consideration which are trending in the news.

# TECHNICAL DETAILS

**Database and it’s design**:

Database used will be mysql to store metadata and path for json file for retrieving answer. It will be updated under some specific time so that user will be able to get information about current trends in exact manner. Information will be divided in different files and in a particular file contents related to some specific topic will be stored. Metadata will contain information like from which link the data was taken and when was it taken. The database will contain keywords and path of relevant documents. It will stored in the format as, keywords and the corresponding path of document which contains the keywords.

**Language and libraries** :

Python will be used to write programming logic for question answer system. SpaCy is used to prepare text for deep learning. It interoperates seamlessly with TensorFlow, PyTorch, scikit-learn, Gensim and the rest of Python's AI ecosystem. With SpaCy, you can easily construct linguistically sophisticated statistical models for a variety of NLP problems. SpaCy uses an object-oriented approach. Parsing some text returns a document object, whose words and sentences are represented by objects themselves. Each of these objects has a number of useful attributes and methods, which can be discovered through introspection.

**User Interface**:

Django Web Framework will be used to design UI. Django is a high-level Python Web framework that encourages rapid development and clean, pragmatic design.

# INNOVATIVENESS AND USEFULNESS

Recently there is a tendency of growth of information requests to different search engines. The purpose of such requests is to find information about the event, phenomenon, date, place, reason, consequence or object. Most of the requests of this category are defined as a question, using such language constructions as “what is it”, “who is it”, “how is it”. To solve such problems as formations of an answer to a question, the question-answer systems are used. This system will be able to answer questions based on current trending topics. There are different domains from which the trending topics are selected like politics, sports, business etc. The main purpose of this system is to provide to the user a short answer in a natural language, instead of the set of snippets on which the user needs to collect independently information and on its basis to form the answer to the question. The projects that were closed domain focused only on a particular domain and its knowledge, the system we are proposing is closed domain but the contents of domain change with time ,this is help users get the latest information and it will not be obsolete after certain time period.

# CURRENT STATUS OF DEVELOPMENT

We have done the literature survey and analysed requirements of the project.The database design and the way data has to be stored is finalized. Data gathering part is well researched and websites are figured out which will be used to construct knowledge base. The framework to be used for designing front end and User Interface is decided i.e. Django framework. We are still working on and exploring the algorithms to be used in the projects and that will be completed soon. We’ve not started with the implementation part yet.

# COMPETITIVE ADVANTAGE

# The advantage of this system against the existing systems is that it will provide answers to the latest things that are trending. It will be dynamic and will change the contents depending on the trending topics, so in that way it will never be obsolete and be useful to users.