**MVSR Engineering College, Hyderabad - 501 510 **

**Department of Computer Science**

**and Engineering**

**Project Title :** Question Answering System in Telugu

**Batch.No:**

2451-18-733-128 G. Sreehith Reddy

2451-18-733-129 A.V.S.Shirish

2451-18-733-136 D.Ganesh

Under the guidance of

**Name** : Venkata Ramana Bhattula

**Designation** : Assistant Professor

**Problem Statement :**

● 60% Indian language internet users stated limited language support and content to be largest barrier for adoption of online services.

● Question answering systems have been extended in recent years to encompass additional domains of knowledge. But, there is no advancement in native languages.

● So, building Question answering system in a native language like telugu would be very much useful

**Objective :**

****

● Construction of a robust program that predicts relevant answer for the given question according to the context.

**Project scope :**

● Conversion of Telugu language into encoded format using BERT and giving the prediction of the task.

● Presently our project is limited to prediction of answer from given raw text of minimum size but could be modified by supplying documents of large sizes.

**Applications :** 

****

● Personal Assistants

● Search Engines

● Chat Bots

● E-Learning

**Requirements :** 

● Related Python libraries 

● Pretrained cross lingual NLP model

● Tokenizer 

● High-end RAM and GPU enabled machine

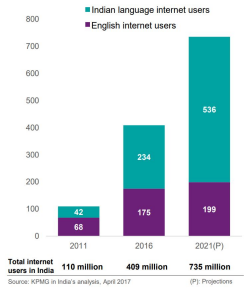
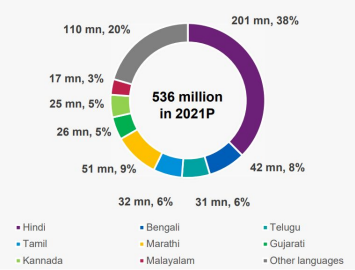
**Challenges :**

****

● Telugu Language Corpus is limited over the internet. ● Model Selection.

● Ambiguity

**Indian Language Sources on Internet :**

****

**A Taxonomy of QA Systems:**

| **Class** |
| --- |
| 1 |

**Type Example**

Factual Q : భారతదేశ రాజధాని ఏమిటి? A : భారతదేశ రాజధాని నూ్యు ఢిȄ్లో

2 Simple Reasoning Q : Ȏč క్రీటీస్ ఎలా చనిపč యాడు? A : ȇషం కారణంగా చనిపč యాడు.

3 Boolean ( True / False )

Q : సూర ్యుడు పȋశ్చిమాన ఉదయిȎా్తు డా? A : తప­్పు

4 Multiple Choice Q: ఇంద్రధనసు్సుక¡ ఎని్న రంగల¡ ఉనా్నయి? 1) 4 2) 5 3) 6 4) 7

5 Opinion Based Q : 2018 ఎని్నకల¡ నా్యుయంగా జరిగాయా?

**BERT ( BiDirectional Encoder Representations From Transformers )**

● Recently (2018) published paper by researchers at Google AI .

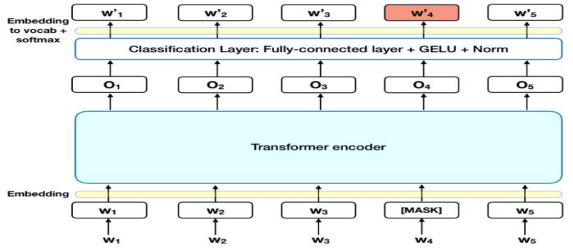
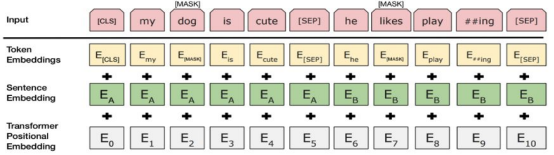
● Makes use of Transformer , a mechanism that learns contextual relation between words in a text.

● Has two mechanisms which includes, encoder that reads input and decoder that produces prediction of the task.

● As the name suggests it is Bidirectional which means reads texts sequentially (left- to-right or right-to-left)

● Many variants of BERT have been reproduced like roBERTa by Facebook, DistilBERT, IndicBERT.

**Working of BERT :**

****

**Working Example :**

**Passage** :

M.V.S.R 1981లĐ ఉȎా్మినియాయూనివరి్సుటీకిఅనుబంధంగా ఎంవీఎస్ ఆర్ ఇంజినీరింగ్ కళాశాల Ȏా్థు పించబడింది. ప్రసు్తు తం, కళాశాలలĐ MBAతో పాటు సిȇల్, CSE, ECE, EEE, IT, మ¸కానికల్ మరియ ఆటోమొబ·ౖల్ ఇంజనీరింగ్ లĐ B.E పč్ర గా్రీ మ్ ల¡, CSE, ECE మరియ మ¸కానికల్ లలĐ PG పč్ర గా్రీ మ్ ల¡ ఉనా్నయి. ఈ కళాశాలను ఉȎా్మినియా ȇశ్వȇదా్యులయం పిహÁచ్ డిఅవార ్డు క¡ దారితీసేపరిశోధనలను కొనȎాగించేకేంద్రంగా గరి్తుంచింది. CSE, ECE మరియ మ¸కానికల్ ఇంజనీరింగ్ ȇభాగాలలĐ డిగ్రీ. మొత్తుం మీద ఏడాదికి 1000 మంది ȇదా్యుర ్థు ల¡ అడి్మిషన్ తీసుక¡ంటార . ఈ కళాశాల రాష్ట్రంలĐని ప¿్ౖరైవేట్ ఇంజినీరింగ్ కళాశాలలĐ్లోమొదటిది, మరియ కళాశాల పా్ర రంభం నుండి అదు్భుతమ¸ైన మరియ ఆదర్శపా్ర యమ¸ైన ప్రమాణాలను నిరా్ధా రిసు్తు ంది, ఇదికళాశాలను రాష్ట్రంలĐని అత§్యున్నత మరియ కోర క¡నే కళాశాలలలĐ ఒకటిగా నిలిపింది.

**Question:**

M.V.S.RలĐ PG కోర ్సుల¡ ఏȇ అందించబడతాయి?

**Answer:**

M.V.S.R లĐ CSE, ECE మరియ మ¸కానికల్ PG కోర ్సుల¡ అందించబడతాయి.

**Q & A**

****