

README

Instruction for running the code (In the Folder **code**):

- Download the folder present [here](#)
- Include the absolute path to the downloaded folder in line 6 of the file "transformer.py":
 - `question_model = T5ForConditionalGeneration.from_pretrained("Absolute Path to the T5_Squad_V1_Model")`
- Execute these steps:
 - create a virtual environment
 - `pip3 install --upgrade cython`
 - `pip3 install numpy`
 - `pip install --upgrade pip`
 - `pip3 install -r requirements.txt`
- Run the file main.py:
 - `python3 main.py`

CSV Data Files (In the Folder **data**):

1. ObjQA Dataset:

- **PAA_Dataset** - Contains the extracted questions from *Google's People Also Ask Section* for around 1800 objective questions with each objective question containing at max 12 PAA questions. We are open-sourcing this dataset to provide a variety of questions which can be used to build data-driven approaches.
- **Annotations_Clustering** - Contains the annotated question set generated by the "*Clustering and Rule-based template*" approach.
- **Annotations_PAA-T5** - Contains the annotated question set generated by the "*PAA and T5*" approach.
- **Rule-based approach** - Contains the annotated question set generated by the existing method, *Rule-based approach*
- **T5-Transformer** - Contains the annotated question set generated by the existing method, *T5-Transformer*

2. Mcq Dataset:

- **McqData_proposedSystem** - Contains the annotated question set generated by the proposed system.

- **McqData_RuleBasedBaseline** - Contains the annotated question set generated by the existing method, *Rule-based approach*
- **McqData_T5** - Contains the annotated question set generated by the existing method, *T5-Transformer*