This folder contains source code for **CRB-CRS**:

We have two main directories at the root level. The first named **‘src’** contains the complete Python source code for CRB-CRS. The second directory, named ‘**data**,’ contains the data for both movie recommendations and dialogs.

In order to reproduce the results of this paper, the Python scripts under the directory ‘**src**’ should be executed in the given order.

1. The file named ‘**PrepareTrainingData.py**’ is used to parse ReDial training data into a text file based on each dialog between the seeker and human recommender.
2. The file named ‘**PreprocessSentences.py’** is used to preprocess the data parsed in the previous step. This piece of script outputs a text file containing training data after prepossessing, as mentioned in the manuscript.
3. The file named **‘RetrieveCandidatesByHistory.py**’ is used to train the TF-IDF model and retrieve candidate responses. Precisely, we retrieve candidate responses based on four different parameter values of dialog history for the user utterance, as mentioned in the manuscript. The approaches for outliers pruning and ranking of the retrieved set of candidate responses are also implemented in the same file.
4. The file named ‘**Cacululate\_MLE\_Probs.py**’ is used by the script in the file ‘RetrieveCandidatesByHistory’ to calculate the bigram probabilities in our ranking technique.
5. The file named ‘**IntegrateRecommendations.py**’ is used for the purpose of the integration of both movie recommendations and metadata (genre, explanation, etc.) information.
6. The remaining three .py files starting from ‘*Recommender’* are recommender approaches used to compute and integrate recommendations in the retrieved responses.

The description of the content in the ‘**data** directory is as follows.

1. There are three subdirectories under this directory named ‘**DialogData**’ and ‘**RecommendersItemData**’ and ‘**redial\_dataset**’.
2. Under the ‘**DialogData**’ directory, we have all the dialogs’ data files used or exported using python scripts, as previously mentioned. This data is used to train and retrieve candidate responses.
3. Under the ‘**RecommendersItemData**’ directory, we have the MovieLens rating dataset, which is used to compute recommendations and integrate metadata information (e.g., genres, actors, etc.) in the retrieved responses. For the purpose of the model’s offline initialization, we converted the original MovieLens data into further files.
4. The directory named redial\_dataset.zip contains the original dataset split in both training and test sets.

The file named ‘**KBRD & KGSF.pdf**’ contains the details on how to access the source code of both the DeepCRS and KBRD systems.

The file named ‘**Requirementes.txt**’ includes all the Python libraries to install in order to reproduce the results.

Note: The root folder named ‘**RB-CRS Code**’ can be imported as a standalone Python project.