**NLP Assignment 3 Report**

Braj raj Nagar 2022AIB2682 Sangam Kumar 2022AIB2671

**Introduction**: The aim of this assignment is to develop a parsing model for a task-oriented dialog (TOD) system that can extract the intent and associated slot-values from user input. The system takes input in the form of user utterances, dialog history, and extra context information and generates a parsed output that can be used by downstream applications to execute the user's intent.

* Data Description: The input to the parsing model consists of several components:
* User Utterances: This is the input text from the user that the parsing model needs to interpret.
* History: A list of past user and system utterances, which provides context for interpreting the current user input.
* User Lists: A user list is a named collection of items curated by the user. This information is used as context to interpret the user input.
* User Notes: A user note consists of a title and text. This information is used as context to interpret the user input.
* User Contacts: A list of user's contacts. This information is used as context to interpret the user input.
* Output: String indicating parsing output, which is the parsed representation of the user input.
* Pattern: Linguistic pattern in the user utterance, which is used to help identify the intent and associated slot-values.

**Methodology:** To achieve the objective of developing a parsing model, we used a t5 transformer pretrained model. The t5 transformer model is a state-of-the-art language model that has been pre-trained on a large corpus of data. The model uses a transformer architecture that allows it to learn the context of the input and output sequences.

We used the t5-small transformer model in the following steps:

* Step 1: Data collection: We collected user utterances, dialog history, user lists, user notes, and user contacts to train and test the parsing model. We also used linguistic patterns in user utterances to identify the intent and associated slot-values.
* Step 2: Pre-processing: We pre-processed the collected data by cleaning and tokenizing it.
* Step 3: Model training: We trained the t5-small transformer model on the pre-processed data using a supervised learning approach. The model was trained to predict the intent and associated slot-values from the user input and context information.
* Step 4: Model evaluation: We evaluated the performance of the trained model by measuring its accuracy, intent\_accuracy and parsing\_accuracy.

Conclusion: In conclusion, developing a parsing model for a task-oriented dialog (TOD) system is a challenging task, but it is essential for the system to function effectively. The model needs to be trained on a large corpus of annotated data, and various NLP techniques and machine learning algorithms can be used to extract relevant features and train the model. With an accurate parsing model, the TOD system can effectively extract the user's intent and associated slot-values, which can then be used to execute the desired task.