**Chatbot or Task-Oriented Dialog Agent Project Report**

**Chatbot: Axford**

**Human Language Technologies**

**CS 4395.001**

**Lisa Chen**

**System Description**

* This project, designated as Axford, is to be a chatbot project that relies on external knowledge that is collects and/or builds. They are designed to be a force of chaos with an interest in dnd that cooperates on the condition that it feels like cooperating. It is expected that that it is run on PyCharm, preferably on windows. There may be differences in setup if you use a Mac and/or Linux. If you don’t already have PyCharm installed on your computer, for best results install it. On PyCharm, go to the python packages tab and search up each package and/or library that’s listed at the top of the project and install each one yourself. You may need to also install this using your computer’s equivalent of Command Prompt depending on how you set up this project. The program ends if you enter the word: bye
  + **Description of NLP Techniques Used**
    - Chatbot project includes multiple NLP techniques such as:
      * **tf-idf** (term frequency-inverse document frequency):
        + the combination of the term frequency with the inverse document frequency where it involves multiplying tf by id
      * **tf:** (term frequency)
        + the count of a given term, t, in a document divided by the total number of words in the document, d.
      * **idf:** (inverse document frequency)
        + the computation result of first dividing the number of documents in the corpus, N, by the number of documents in which the term t appears at least once. Then the log is taken of this quantity.
      * **Pos tagging**: (parts of speech tagging)
        + the tagging of the parts of speech of a corpus
      * **Tokenization**:
        + the text processing technique that involves dividing text into smaller units, as in you’re dividing text into individual tokens, which include words, numbers, and punctuation. This doesn’t involves everything else but whitespace in the text.
      * **term frequency measures**:
        + measures that calculate the term frequency of a corpus
      * **Information retrieval techniques**:
        + techniques used to retrieve information
      * **Information Extraction**:
        + composes of techniques used to extract needed information
      * **Text normalization**:
        + changes related words into some standard form for the purpose of counting the occurrences for text processing purposes which can consist of converting all the text to lower case, removing punctuation, removing numbers or replacing them with a NUM token, stemming words - removing affixes, lemmatizing words which involves converting to the root form, removing stop words where you’re reducing text to important words only
  + **Description of How the NLP Techniques Were Used**
    - We can see that the NLP Techniques were used to handle the information for the knowledge base, in terms of gathering, building and processing it, in order to give further function to this project. The mentioned knowledge base is to be built using the collected and processed information in which this knowledge is allowed to have information that originates from the web or gathered collected information from other sources such as the user. This function includes working to have the project give reasonable responses to some degree if they are given reasonable input from the user.

**Diagram of Project’s Dialog Tree or Logic**

* **Diagram

  Description automatically generated**

**Sample Dialog Interactions**

(Sample 1)

* **Axford**: Hi, I’m Axford, a chaotic chatbot that has an interest in DnD. Please enter one of the following choices: UserOne, UserTwo, UserThree
* **User**: UserOne

(Sample 2)

* **Axford:** Hi UserOne! What’s your name?
* **User:** Murph

(Sample 3)

* **Axford:** That’s a cool name! What are your likes and dislikes?
* **User:** I like dnd, and I dislike getting my sleep interrupted

(Sample 4)

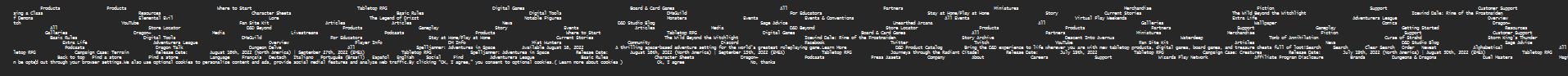
* **Axford:** I’m going to take a note of that. Btw, If you saw anything weird earlier that’s just us adding or storing some data. Just ignore it ok?
* **User:** Ok, then

(Sample 5)

* **User**: Bye
* **Axford**: Look over there! \*Runs away\* I gotta go, bye!

**Appendix for Created Knowledge Base and/or Live Lookup**

* The knowledge base contains the knowledge that has been built and/or collected
* This knowledge has been collected from the web to be used to build this knowledge base
* This knowledge base uses this external knowledge that it collected from it’s designated source(s) which includes but is not limited to knowledge collected from web, dialogue etc
  + **Sample(s) of Knowledge Base and/or Live Lookup**

****

**Appendix for Sample User Models That Were Created**

* **UserOne**
  + This is one of the user models
  + User model made to be selected by one of the users
  + User is expected to input their name when requested
  + User expected to input their likes and/or dislikes when requested
  + Information expected to be stored
  + Sample user model created for one of the users of the project as there’s 3 reviewers of the project
* **UserTwo**
  + This is one of the user models
  + User model made to be selected by one of the users
  + User is expected to input their name when requested
  + User is expected to input their likes and/or dislikes when requested
  + Information is expected to be stored
  + Sample user model created for the 2nd user of the project as there’s 3 reviewers of the project
* **UserThree**
  + This is one of the user models
  + User model made to be selected by one of the users
  + User is expected to input their name when requested
  + User is expected to input likes and/or dislikes when requested
  + Information is expected to be stored
  + Sample user model created for the 3rd user of the project as there’s 3 reviewers of the project
* **User**
  + This is one of the user models
  + User is expected to input their name when requested
  + User is expected to input their likes and/or dislikes when requested
  + Information is expected to be stored
  + Sample User model created in case user needs an additional user model that isn’t the first 3 ones.
  + Made in case an additional user model is needed

**Chatbot Evaluations**

* We see that from the project that it gives fairly reasonable responses, if it is given reasonable input that is expected to be given according to how the project was setup and trained, where should this happen in the expected manner then the dialog is fairly complex. It is also seen from the project that it includes multiple NLP techniques that are mentioned earlier in the report. This project has an implementation that can be considered to have originality of some level. As for the idea of the project’s originality, the originality can be judged to be of good level. When it comes to the user’s data, we can see that it’s being stored.
* Further analysis of chatbot’s strengths and weaknesses are mentioned below
  + **Chatbot Strengths Analysis**
    - We can see that if the project is setup properly and everything that is needed to be installed is installed with all the files in the correct places and any accounts and/or authentication that is needed to be done is done, then the project runs.
    - In addition, we can see that if the project is given responses that it expects, then the responses can be considered to be reasonable. The reasonableness of the input is based on what the user inputs into the project.
    - We see that flow of the conversation of the project can be considered to be complex of some degree. Further details on this can be seen in the given examples provided.
  + **Chatbot Weaknesses Analysis**
    - This project has been built on Windows and as such there may exist differences in setup instructions and/or compatibility when it comes to Mac and/or Linux users
    - Only runs correctly if you set it up correctly including installing anything that needs to be installed including python packages on pycharm and placing files in the correct place etc
    - Depending on your computer’s hardware specifications it may take a long quantity of time for the program to fully load including in between the different methods of the code
    - Depending on your Wi-Fi speed, the loading time of the program may take a long quantity of time to fully load everything for each part of the project
    - Need to install python packages yourself through PyCharm beforehand
    - Need to install what’s needed to be installed on your pc yourself beforehand