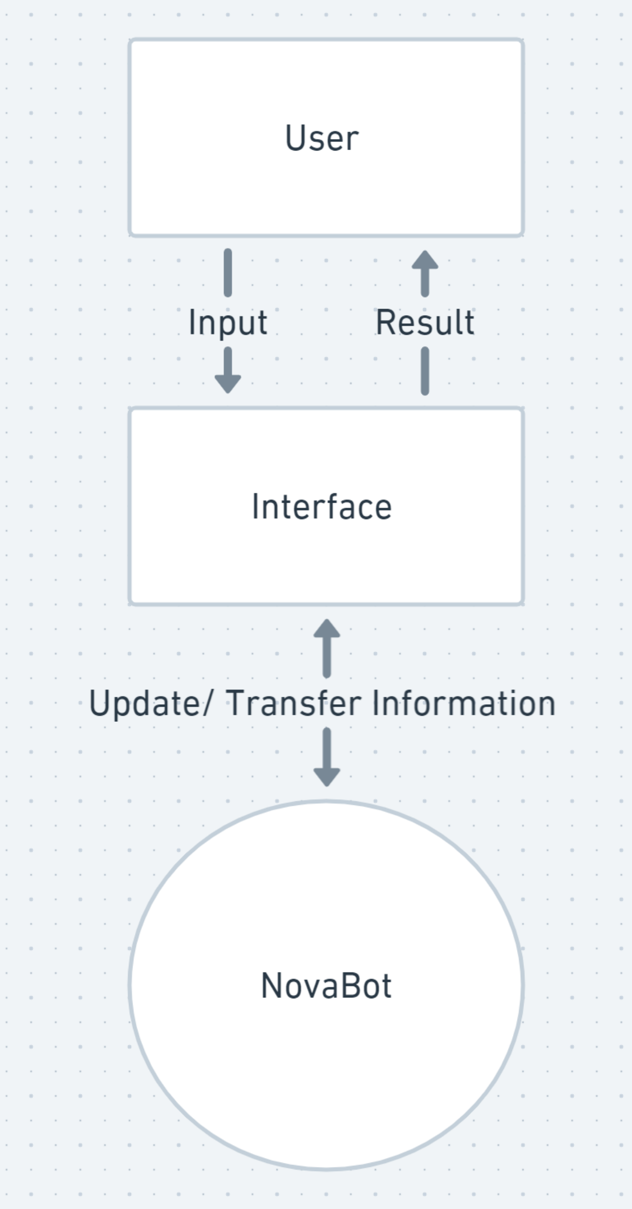
COSC 310 - Software Engineering: Chatbot

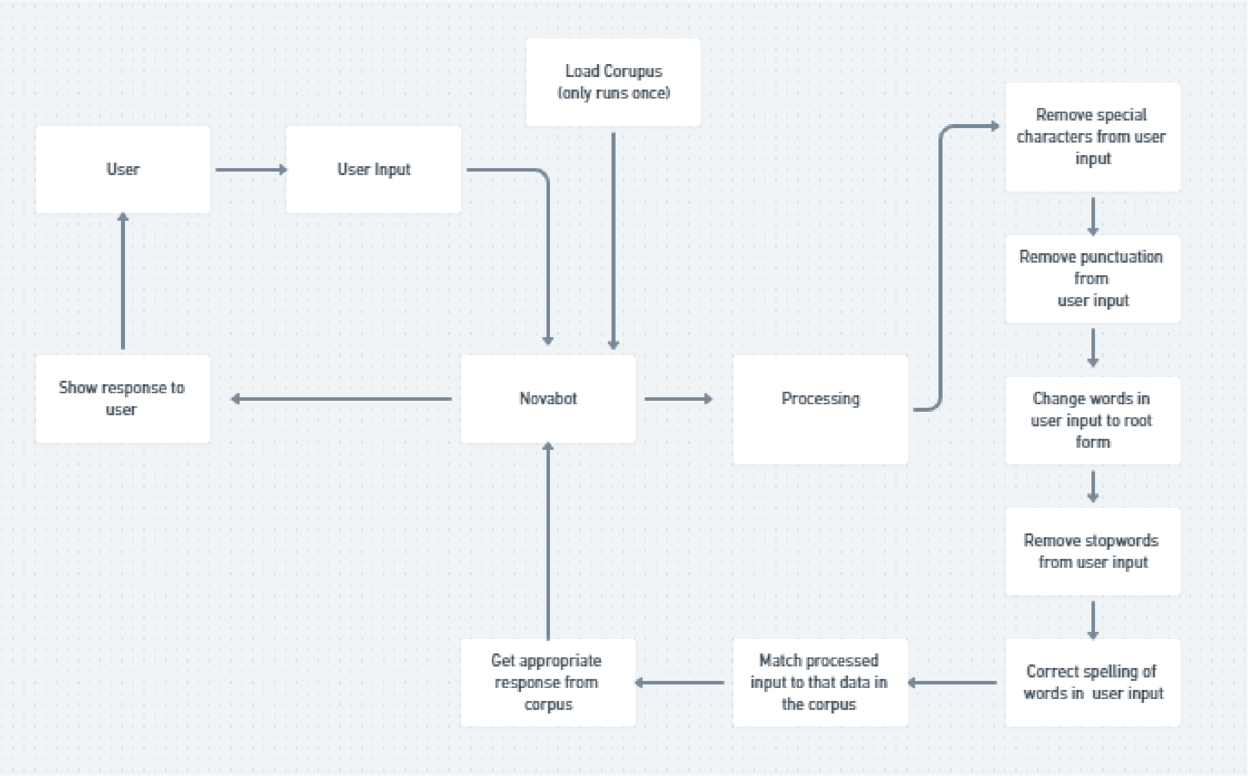
Haneen Abu-Hijleh, Ravi Bullock, William Chen,Peizhou Duan, Ilya Yereferenko

Level 0 Diagram:



The user will be inputting questions through the interface, and the sentences will be processed by the NovaBot and produce the correct answer for the user through the interface. The process will be cycled.

Level 1 Diagram:

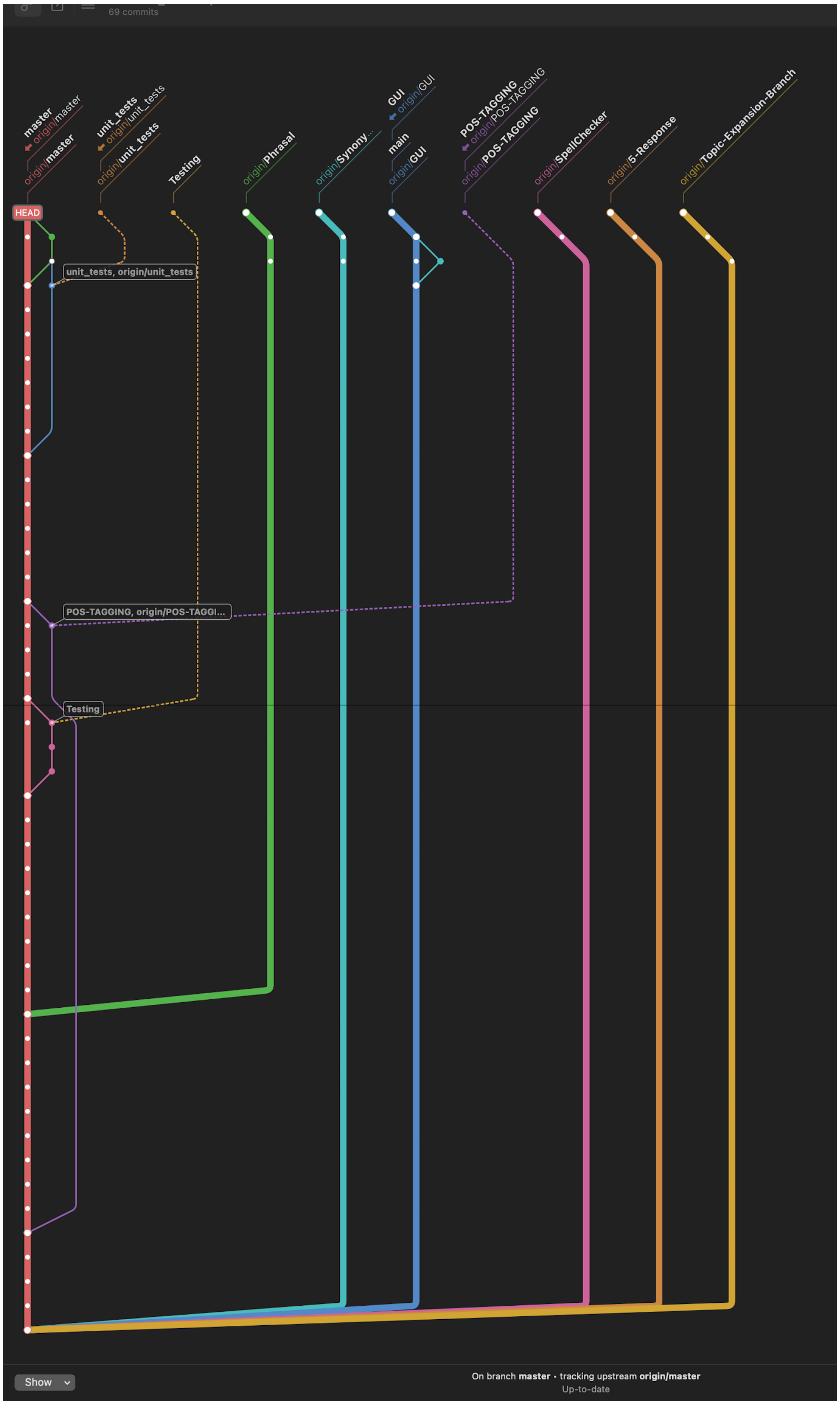

Load Corupus
(only runs once)
User
User Input
Remove special characters from user input
Remove punctuation from user input
Change words in user input to root form
Remove stopwords from user input
Correct spelling of words in user input
Show response to user
Novabot
Processing
Get appropriate response from corpus
Match processed input to that data in the corpus

The first thing that happens is that the corpus is loaded. The corpus contains all the questions and answers that NovaBot learns from, and is run to prepare NovaBot to answer questions. When a user asks a question, NovaBot accepts the input, then removes any special characters, removes punctuation, reduces the words asked to their roots, removes stopwards such as "and" or "a", then fixes any spelling errors, and then finally matches the resulting words to the corpus and extracts an appropriate response. NovaBot then outputs the result back to the user, and the user can once again ask a new question.

Submission of your GitHub repository. Graph showing different features developed on a separate branch and the commits made in the repository.

Github Repo: https://github.com/COSC-310-Group-18/COSC310_A3

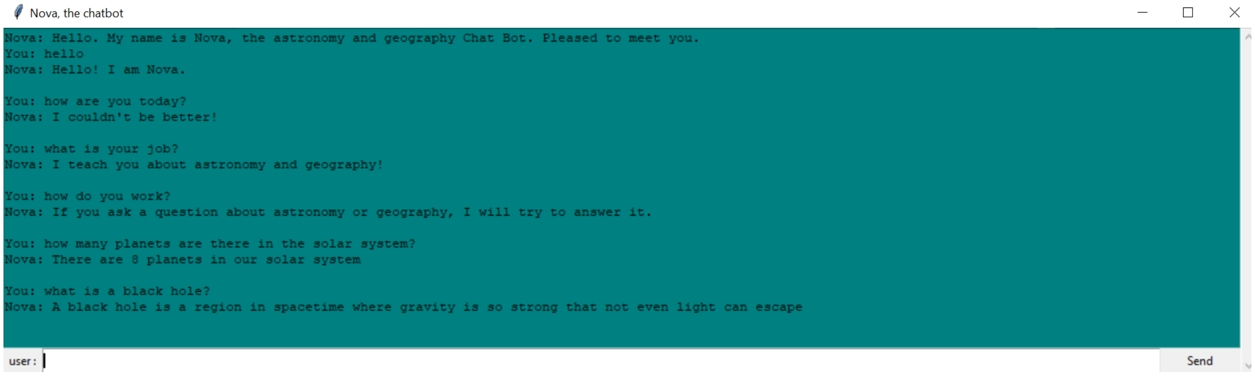
Graph of Branches and commits:



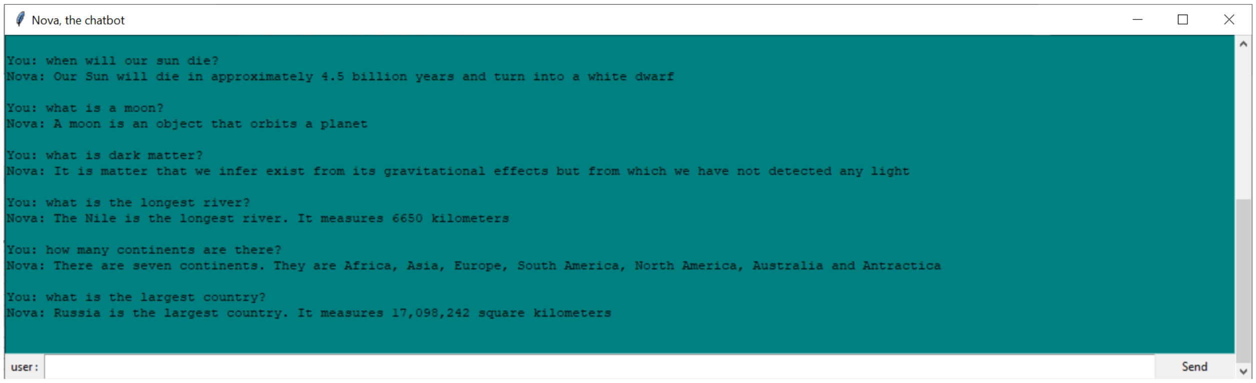
On branch **master** • tracking upstream **origin/master**
Up-to-date

5 points: Include sample output in your project report. Have one dialogue (at least 30 turns) that show a good or feasible conversation -- ensure your new features are demonstrated! Document a list of limitations of your program, and have at least two short dialogues that show when your agent is not able to handle the conversation properly.

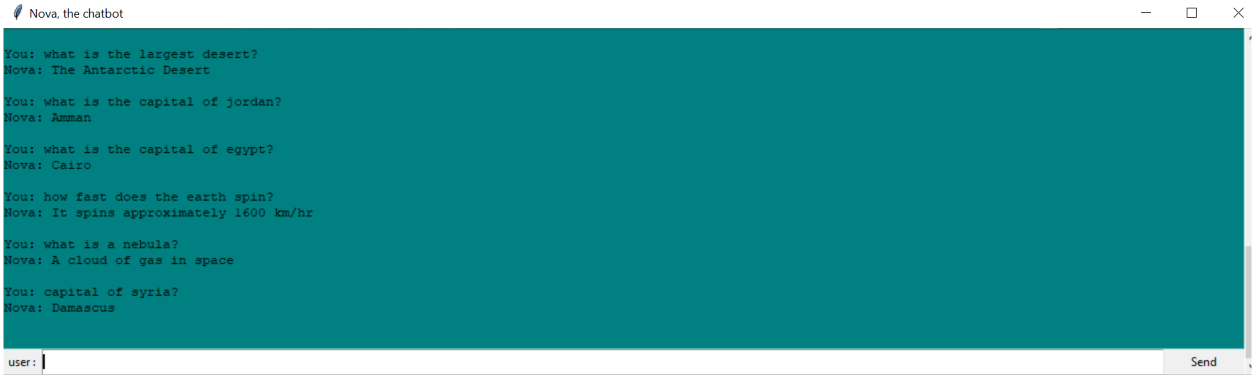
Good conversation:



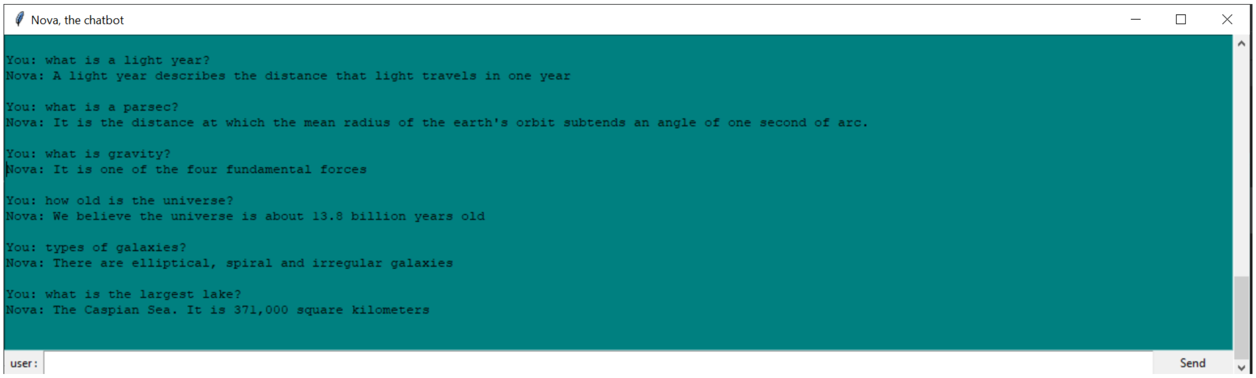
```
Nova: Hello. My name is Nova, the astronomy and geography Chat Bot. Pleased to meet you.
You: hello
Nova: Hello! I am Nova.

You: how are you today?
Nova: I couldn't be better!

You: what is your job?
Nova: I teach you about astronomy and geography!

You: how do you work?
Nova: If you ask a question about astronomy or geography, I will try to answer it.

You: how many planets are there in the solar system?
Nova: There are 8 planets in our solar system

You: what is a black hole?
Nova: A black hole is a region in spacetime where gravity is so strong that not even light can escape
```



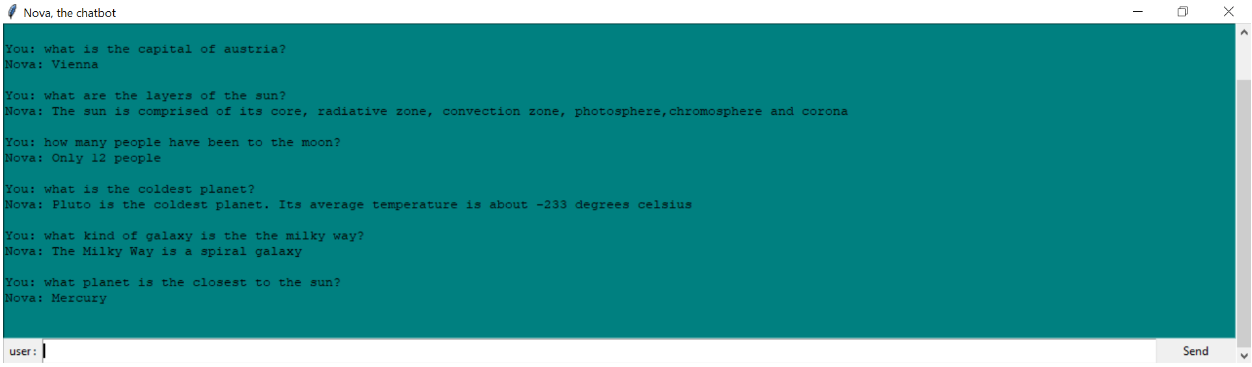
```
You: when will our sun die?
Nova: Our Sun will die in approximately 4.5 billion years and turn into a white dwarf

You: what is a moon?
Nova: A moon is an object that orbits a planet

You: what is dark matter?
Nova: It is matter that we infer exist from its gravitational effects but from which we have not detected any light

You: what is the longest river?
Nova: The Nile is the longest river. It measures 6650 kilometers

You: how many continents are there?
Nova: There are seven continents. They are Africa, Asia, Europe, South America, North America, Australia and Antractica

You: what is the largest country?
Nova: Russia is the largest country. It measures 17,098,242 square kilometers
```



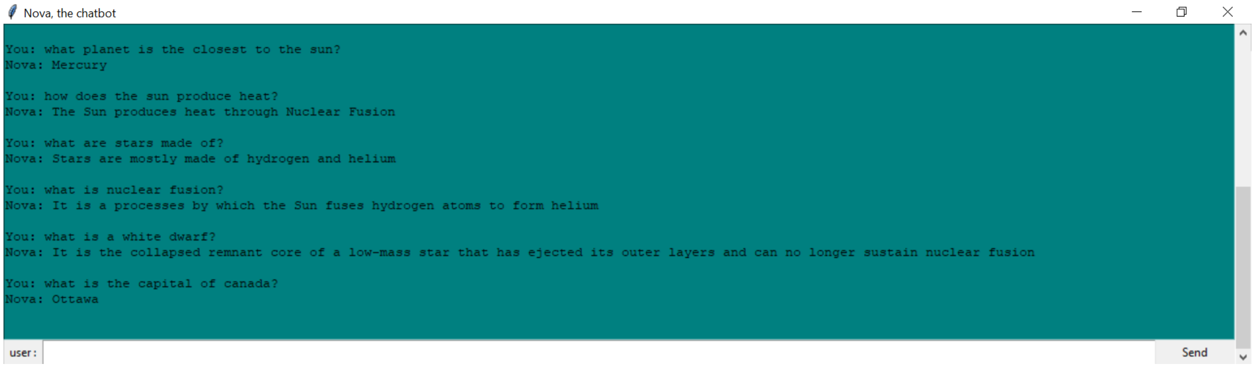
```
You: what is the largest desert?
Nova: The Antarctic Desert

You: what is the capital of jordan?
Nova: Amman

You: what is the capital of egypt?
Nova: Cairo

You: how fast does the earth spin?
Nova: It spins approximately 1600 km/hr

You: what is a nebula?
Nova: A cloud of gas in space

You: capital of syria?
Nova: Damascus
```



```
You: what is a light year?
Nova: A light year describes the distance that light travels in one year

You: what is a parsec?
Nova: It is the distance at which the mean radius of the earth's orbit subtends an angle of one second of arc.

You: what is gravity?
Nova: It is one of the four fundamental forces

You: how old is the universe?
Nova: We believe the universe is about 13.8 billion years old

You: types of galaxies?
Nova: There are elliptical, spiral and irregular galaxies

You: what is the largest lake?
Nova: The Caspian Sea. It is 371,000 square kilometers
```

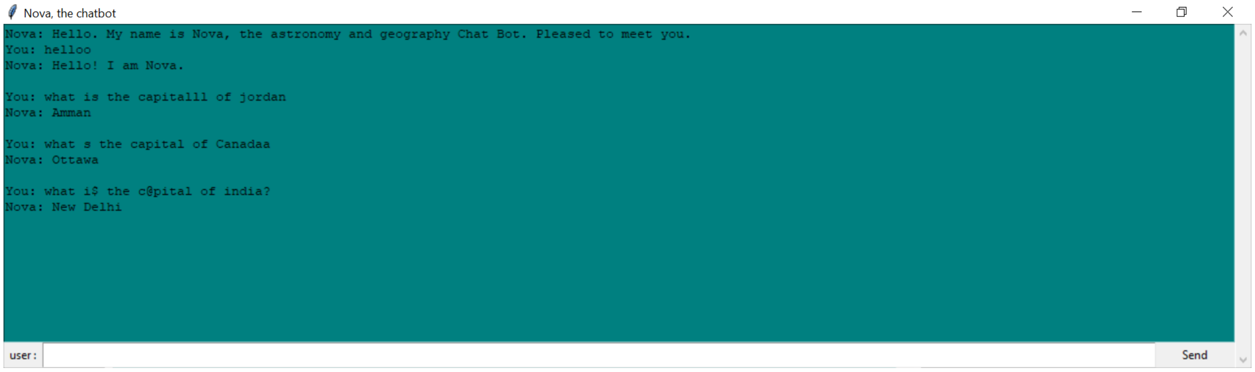


```
You: what is the capital of austria?
Nova: Vienna

You: what are the layers of the sun?
Nova: The sun is comprised of its core, radiative zone, convection zone, photosphere,chromosphere and corona

You: how many people have been to the moon?
Nova: Only 12 people

You: what is the coldest planet?
Nova: Pluto is the coldest planet. Its average temperature is about -233 degrees celsius

You: what kind of galaxy is the the milky way?
Nova: The Milky Way is a spiral galaxy

You: what planet is the closest to the sun?
Nova: Mercury
```



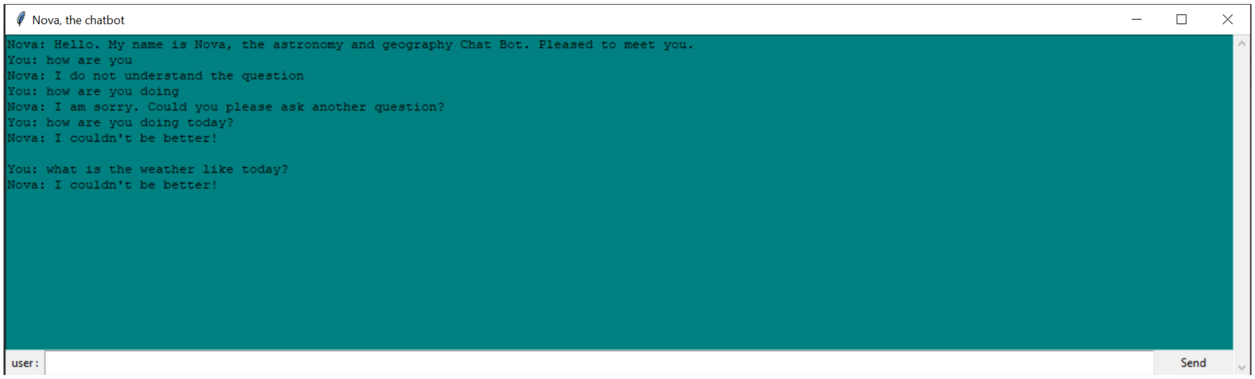
```
You: what planet is the closest to the sun?
Nova: Mercury

You: how does the sun produce heat?
Nova: The Sun produces heat through Nuclear Fusion

You: what are stars made of?
Nova: Stars are mostly made of hydrogen and helium

You: what is nuclear fusion?
Nova: It is a processes by which the Sun fuses hydrogen atoms to form helium

You: what is a white dwarf?
Nova: It is the collapsed remnant core of a low-mass star that has ejected its outer layers and can no longer sustain nuclear fusion

You: what is the capital of canada?
Nova: Ottawa
```

Demonstrating our spelling and phrasal (to accommodate special characters) new features:



```
Nova: Hello. My name is Nova, the astronomy and geography Chat Bot. Pleased to meet you.
You: helloo
Nova: Hello! I am Nova.

You: what is the capitalll of jordan
Nova: Amman

You: what s the capital of Canadaa
Nova: Ottawa

You: what i$ the c@pital of india?
Nova: New Delhi
```

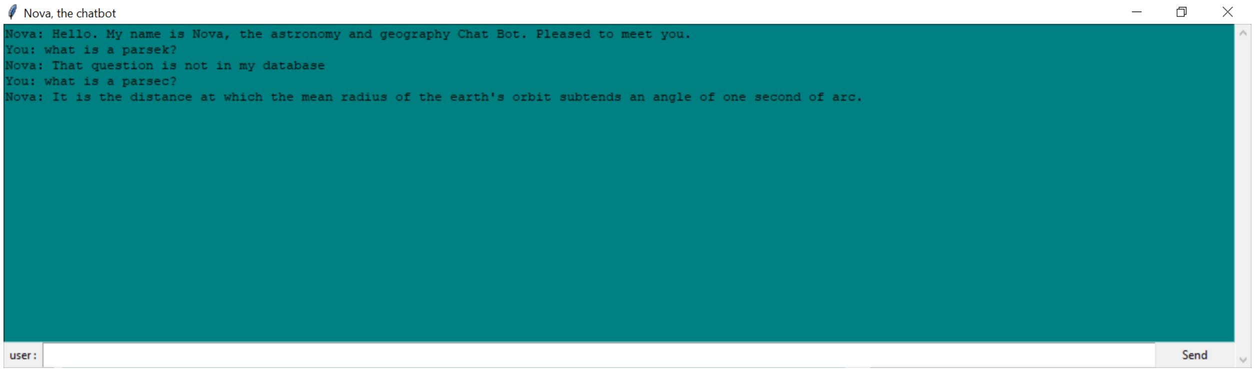
Bad conversation #1:



```
Nova: Hello. My name is Nova, the astronomy and geography Chat Bot. Pleased to meet you.
You: how are you
Nova: I do not understand the question
You: how are you doing
Nova: I am sorry. Could you please ask another question?
You: how are you doing today?
Nova: I couldn't be better!

You: what is the weather like today?
Nova: I couldn't be better!
```

Because of the way we remove stopping words, some sentences become almost impossible to respond to, such are “how are you?” as the entire sentence is composed of stopping words.

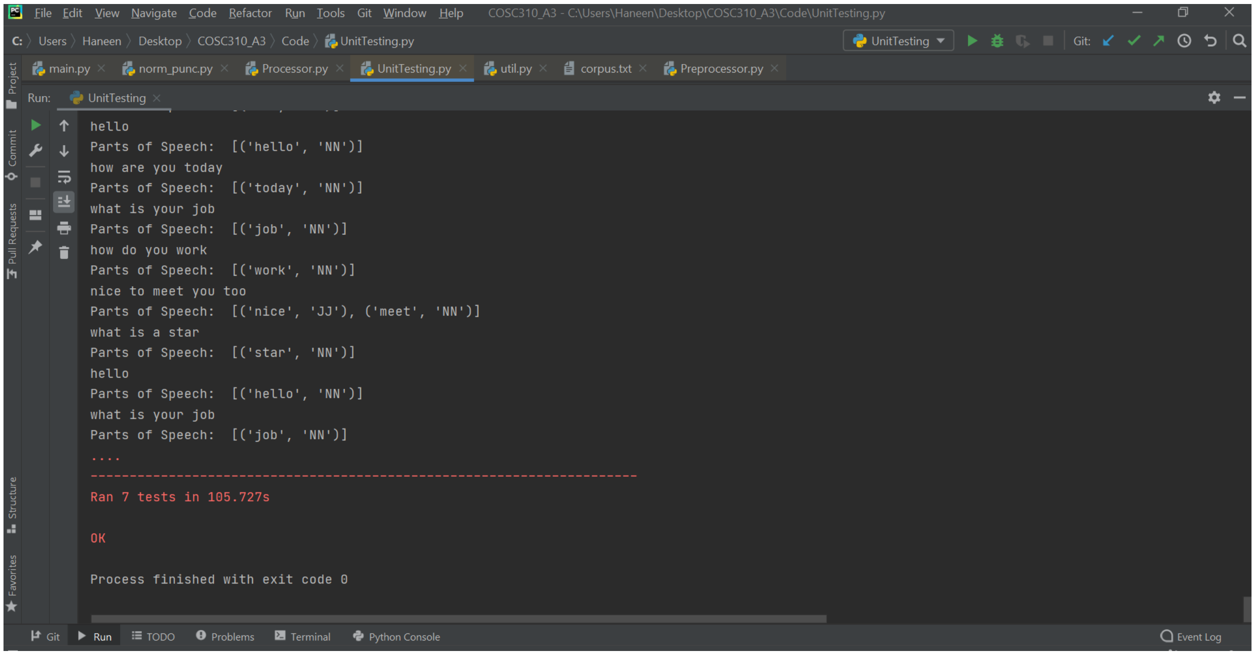
Bad conversation #2:



```
Nova: Hello. My name is Nova, the astronomy and geography Chat Bot. Pleased to meet you.
You: what is a parsek?
Nova: That question is not in my database
You: what is a parsec?
Nova: It is the distance at which the mean radius of the earth's orbit subtends an angle of one second of arc.
```

Even though we did add spell checking functionality, Nova does not account for all of the common spelling errors that the user maybe make (for example, our spell checker turns parsek into parse instead of turning it into parsec)

Passing the unit tests:


File Edit View Navigate Code Refactor Run Tools Git Window Help COSC310_A3 - C:\Users\Haneen\Desktop\COSC310_A3\Code\UnitTesting.py
C: Users Haneen Desktop COSC310_A3 Code UnitTesting.py
UnitTesting
Git:
main.py norm_punc.py Processor.py UnitTesting.py util.py corpus.txt Preprocessor.py
Run: UnitTesting
Project Commit Pull Requests Structure Favorites
Git Run TODO Problems Terminal Python Console Event Log

```
hello
Parts of Speech:  [('hello', 'NN')]
how are you today
Parts of Speech:  [('today', 'NN')]
what is your job
Parts of Speech:  [('job', 'NN')]
how do you work
Parts of Speech:  [('work', 'NN')]
nice to meet you too
Parts of Speech:  [('nice', 'JJ'), ('meet', 'NN')]
what is a star
Parts of Speech:  [('star', 'NN')]
hello
Parts of Speech:  [('hello', 'NN')]
what is your job
Parts of Speech:  [('job', 'NN')]
....
----------------------------------------------------------------------
Ran 7 tests in 105.727s

OK

Process finished with exit code 0
```

5 points: Based on your system, include a list of at least 5 features that you can extract from your code or design that can be shared with others as an API

GUI - GUI has the basic design of a conversation window that provides a clear view. It can be useful for text based interfaces that require better interactions.

Synonym recognition - code for synonym has the ability to define the synonym of the word that has been provided. It can be used in any project that involves words, like a dictionary in order to provide more flexibility of the program.

POS tagging - The pos tagging code has the ability to label each word in the sentence. It can distinguish the word sense and the syntactic role of a word if it's needed. It is useful for anything that involves NLP tasks.

Phrasal - Is useful to clean text corpus or inputs if it's needed. It is useful for NLP developments to improve the flexibility of the system.

Correct spelling - This code has the ability to read a word if it has basic spelling errors. Good for any project that involves inputs or requires identifying words.