

HLT Chatbot Project Report

System Description

ChatPoet uses verses of famous poems to generate responses that are lyrical, flowery, and infused with the words of some of the greatest poets of all time. If you're looking for an introspective conversational partner with vivid language, the Poetic Speak chatbot can talk about love, life, death, nature, and tales of the human condition. There is infinite subject matter covered by poetry which is why the topic itself is broad. As a way of adding variations to the ChatPoet responses, the URL can be changed to any poet's homepage on the site mypoeticside.com.

This chatbot was built from scratch using:

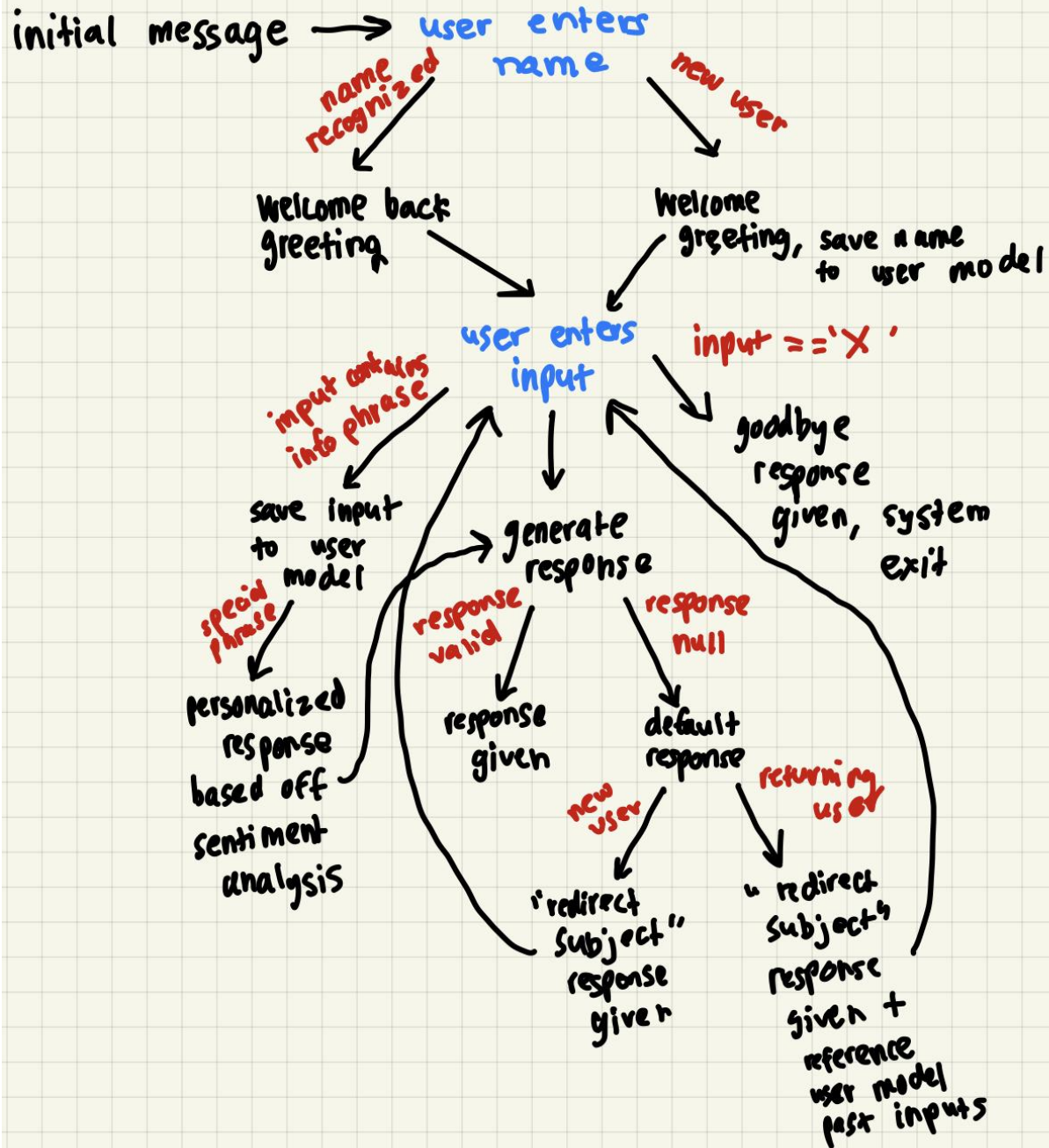
- Python
- NLTK
- Scikit-Learn
- BeautifulSoup
- ... and other libraries which play a minor role, these can be viewed in the source code

Poems were scraped off the website in order to build the Knowledge Base. For more information on the Knowledge Base construction, see the respective report section. The Knowledge Base is a dictionary of the 50 most common tokens found in the input data as well as the poem stanzas in which that word is found. To preprocess the text and create the tokens, I used NLTK `word_tokenize` and set each token to lowercase to eliminate duplicates. Then I removed English stopwords and tokens of a length greater than three. Then I used NLTK WordNetLemmatizer to create lemmas of these tokens to further restrict the set to unique tokens. Finally, I used NLTK `pos_tag` to extract only the nouns as my key terms for the knowledge base.

The code that handles the main chatbot functionality consists of 3 functions. First, the program starts off at `run_chatbot()`. Then the conversation begins between the user and the chatbot. In order for the chatbot to reply it calls `generate_reply()`. For the purpose of adding personalized remarks to the chatbot dialog, a function called `is_positive()` performs sentiment analysis on select user input patterns to change the response produced by the chatbot. Each time

Nina Rao
CS 4395.001

`run_chatbot()` is called, it represents a new instance or new conversation with ChatPoet. The function begins by unpickling the top terms, poem stanzas, and knowledge base that were all set up and pickled earlier. Then I established some phrases denoting personal opinion or information such as “I feel” or “I like” in a list. Each time the chatbot encounters these phrases in user input, the user input is saved in a user model in a dictionary key corresponding to the user’s name. For certain phrases, the chatbot modifies the input to reframe it towards the user as a personalized response. The reframing is either positive/congratulatory or negative/sympathetic and determined in `is_positive`. The function `is_positive` uses the NLTK `SentimentIntensityAnalyzer` to classify a statement as positive or negative. I did this by measuring whether the compound score in the polarity score is greater than zero. Compound scores range from -1 to 1 and give an indication of overall sentiment, which can be easier than comparing negative scores to positive scores. ChatPoet asks for the user’s name at the beginning of every conversation and is able to recognize returning users by referencing the user model keys. Apart from the personalized statements, the generative element of the responses comes from `generate_reply()`, which has two paths. First, if there is an exact match between any of the key terms and a word in the user input, the function will run the sklearn cosine similarity algorithm on the sentences or values corresponding to that key. Second, if there is no exact match then the function will run cosine similarity on all stanzas, regardless of whether they fall under a key class or not. These stanzas again are stored in a text file which is pickled on creation and unpickled in the chatbot code. In order to use cosine similarity to compare the user input sentence to other sentences in the poem data, I used the sklearn `TfidfVectorizer` which takes text data and converts it into a numerical vector based on term frequency. These vectors can be fed to the `cosine_similarity` method to produce a matrix. From this matrix, I select the second highest score since the highest will be the match between the user input sentence and itself. Then I return this sentence along with another high scoring match sentence to add length to the chatbot response. If for any reason, `generate_reply()` does not return an appropriate response, the chatbot will default to a message saying it does not understand and will refer to one of the user’s previous inputs (if the user exists in the user model) and ask to change the subject. After each conversation with ChatPoet is terminated, the code writes back the updated user model to a pickle.



Nina Rao
CS 4395.001

Sample Dialogs

Oscar Wilde:

ChatPoet: Good day! Welcome to ChatPoet, poetry is a literary art form focused on the aesthetics of language.
Feeling ruminative? Have a conversation with me! Before we begin, what is your name?

Nina

ChatPoet: It is wonderful to meet you Nina. Start talking or press X to exit.

You: I am in love!

ChatPoet: You are in love? Lovely to hear!

ChatPoet: I that I am great and strong. -
For I am drowning in a stormier sea

You: Do you think love is a weapon?

ChatPoet: And strange it was to think that he

You: What is the cure for a broken heart?

ChatPoet: To the hearts that are sick for thee
The Chaplain's heart is far too sick,

You: But how do I fix it?

ChatPoet: How sad it seems.
Some do it with a bitter look,

You: The end of a relationship is pretty sad, but sometimes it means new beginnings.

ChatPoet: Sometimes about the painted kiosk
is that time dead? lo! with a little rod

You: I want to feel alive, the blood coursing through my veins.

ChatPoet: With the spilt-out blood of the rose-red wine.
And the wild regrets, and the bloody sweats,

You: I have no regrets, I live by one credo, YOLO

ChatPoet: For he who live more lives than one
I have no store

Emily Dickinson:

You: I'm so hungry!

ChatPoet: You are so hungry? I'm sorry to hear.

ChatPoet: Nor was I hungry—so I found
Too hungry to be borne

You: What is love?

ChatPoet: 'Twas Love—not me—
Love provided Thee?

You: I don;t know

ChatPoet: Don't you know?
Don't you know—me?

Nina Rao
CS 4395.001

ChatPoet: Good day! Welcome to ChatPoet, poetry is a literary art form focused on the aesthetics of language.
Feeling ruminative? Have a conversation with me! Before we begin, what is your name?

Nina

ChatPoet: It is wonderful to meet you Nina. Start talking or press X to exit.

You: The weather is beautiful today

ChatPoet: Today.

A little weather-worn,

You: The sun is out and the breeze is cool

ChatPoet: And the upper side-is the Sun-

The Orchard, when the Sun is on-

You: What is the best season?

ChatPoet: Which for a little season made my youth

The best Vitality

You: I like Spring

ChatPoet: I have a Bird in spring

The spring decoys.

You: Yes, the birds are chirping

ChatPoet: We-are the Birds-that stay.

It was as if a chirping brook

You: The rain made all the flowers bloom

ChatPoet: With all the flowers the dead love best.

The cunning hand which made the flowering hawthorn branches bow

You: That is pretty deep

ChatPoet: Just-how-deep-

To tell the pretty secret

You: I have no secrets, do you?

ChatPoet: No more to do have I

That You never do it-

You: Ok if you say so...

ChatPoet: Not for you-to say

I say to you, said Jesus-

You: My life is full of small pleasures

ChatPoet: And then-the size of this "small" life-

You: I feel afraid that things won't last

ChatPoet: I think that Earth feels so

ChatPoet: It is wonderful to meet you Mimi. Start talking or press X to exit.

You: I like to have my cake and eat it too

ChatPoet: Men eat of it and die.

And made as He would eat me up -

You: I don't want to die from eating cake

ChatPoet: Could die-did die-

I could die-to know-

Knowledge Base

```
love : ['          Is it thy will--Love that I love so well--\n', '          My love, and all the glory of the stars.\n', 'For love i
heart : ['To the hearts that are sick for thee\n', '          Now at their sight, my heart being lit with flame \n', '
soul : ['          In straitened bonds the soul that should be free,\n', '          Which seek the essence of my soul to mar.\n', 'I
night : ['For love of it the passionate nightingale\n', 'Who watch him night and day;\n', 'For we did not meet in the holy night,\n
hand : ['          To have walked hand in hand with Love, and seen\n', 'Closed in the wasted hollow of her hand\n', '
thing : ['          Nor walked where Beauty is a thing unknown.\n', '          But surely it is something to have been\n', '
which : ['          Which sent up lilies, dove-like, to his knees.\n', 'Which the nine Muses hold in empery,\n', '
gold : ['With garnered gold in wide domain,\n', '          Barter my cloth of gold for hodden grey,\n', '
moon : ['High in the sapphire heavens hung the moon,\n', '          And many a moon and sun will see\n', 'They mocked the moon in

wind : ['A cooling wind crept from the land of snows,\n', '          For southern wind and east wind meet\
pain : ['          And at thy pleasure weave that web of pain\n', '          Poor toys to soothe such ang
life : ['Yet from the sorrows of his life\n', 'It never feels decay but gathers life\n', 'Since first Man to
face : ['          Stood face to face with Beauty, known indeed\n', 'Some kneeling girl with passionless pale
fair : ['When Love and Life are fair:\n', 'And we knew that ere one dawn grew fair\n', 'When Love and Life a
word : ['          Clutches his sword in fierce surmise\n', '          Where, girt and crowned by swor
dawn : ['He does not wake at dawn to see\n', 'And we knew that ere one dawn grew fair\n', 'God's dreadful da
hair : ['And flagrant daffodil hair,\n', 'And the flagrant daffodil hair. -\n', '
blood : ['          England with bare and bloody feet\n', 'For blood and wine are red,\n', 'And blood and
thou : ['          Ay! though the gorged asp of passion feed\n', 'Ah no! to this bright flower a thousand yea
light : ['          Not marking how the spotted hawk in flight\n', 'The purple dragon-fly had no delight\n',
Christ : ['Dear Christ! the very prison walls\n', 'Ah! wounds of Christ! they were living things,\n', 'Chris
thee : ['To the hearts that are sick for thee\n', 'Thou bringest a brighter star with thee\n', '
crimson : ['Beautiful star with the crimson lips\n', 'O beautiful stars with the crimson lips\n', '
peace : ['          Nay peace: behind my prison's blinded bars\n', '          Peace, peace! we wrong the nc
till : ['          O still ravine! O stormy deep!\n', 'Terror was lying still.\n', 'So still it lay that e
grey : ['A father grey with grief and tears,\n', '          Barter my cloth of gold for hodden grey,\n', 'I
silver : ['With silvered wing and amethystine throat.\n', 'With sails of silver by.\n', 'With sails of silve
hath : ['O well for him who ne'er hath known\n', 'But well for him whose feet hath trod\n', '
passion : ['          Sordello's passion, and the honied line \n', '          Ay! though the gorg
world : ['Dead ere the world's glad youth had touched its prime,\n', 'With the world's legions led by clamor
```

The knowledge base was constructed using a similar web-crawling technique from the previous assignment. A function `urlgen()` kicks off the code with a single url which will be the origin. In the source code, I have left other urls to try in the comments. These urls are for various other poets to generate different results. This starting url is processed in the function `webscrape()`. This function opens the url and uses BeautifulSoup to parse the html. By navigating the HTML of the origin url, I was able to capture the text for each poem belonging to that author and write the lines to a file called `chatbot_data.txt`. Preprocessing is done to identify top/key terms and these terms become the keys of the knowledge base dictionary. The corresponding values are all the stanzas that contain the term. The knowledge base has 50 entries.

User Models

The user model is another dictionary where the keys are the names of users and the values are the inputs that were flagged by the chatbot as containing personal information, opinions, or feelings.

```
Rayven : ["I don't know you...", 'I love it!!', "I'm happy", "I'm so hungry!", "I'm neutral.", "I'm hot", 'I love you']  
Nina : ['I am happy', "I feel afraid that things won't last", 'I am pretty tired actually']  
Mimi : ['I like to have my cake and eat it too', "I don't want to die from eating cake"]
```

Evaluation

I believe ChatPoet succeeds in being able to generate relevant responses to user input and hold a limited conversation using poetry as the knowledge base. The approach I used was a pattern-based approach. Unlike ML approaches, which learn and utilize the entire context of the user input, pattern-based chatbots simply identify patterns. Though let it be known, cosine similarity is used to supplement many machine learning algorithms such as neural networks. ChatPoet's capabilities depend heavily on the input data. If a large and diverse set of poems are scraped, ChatPoet has a greater chance of finding a more tailored match to the user's input. Smaller datasets, as a result, have limited capability. This is why I chose prominent poets with large poem portfolios to test ChatPoet. Despite ChatPoet being built off of purely rules and patterns, I did find that every now and then, it would respond with a phrase that was shockingly apt or witty. This of course cannot be attributed to ChatPoet's "intelligence" as the ChatPoet cannot predict nor improve from conversations. It can only match patterns and return statements with the highest vector similarity. Sometimes, ChatPoet doesn't make sense at all. It returns stanzas that are cut off (due to a poet's writing style) or ambiguous in meaning. If you press further and ask a followup question, it may repeat itself if you use the same key term in your input. But the fact that sometimes just dumb luck could result in a gem of an interaction is endearing in itself. Poetry is an art form that refuses to follow the conventions of formal language. It contains figures of speech, unconventional syntax, imaginative vocabulary – things that without a discerning eye, could be misconstrued or misunderstood as nonsense. But if you take to ChatPoet's replies the same lens that you take to the analysis of poetry, you may be able to find your own unique interpretations in its machine-driven speech.