CS 4395.01 Human Language Technologies

Portfolio: Finding or Building a Corpus

Instructor: *Dr. Karen Mazidi*

Organized by:

Ishaaq Razack - imr180000

Date: Oct 9, 2022

1. Describe how you created your knowledge base, include screen shots of the knowledge base, and indicate your top 10 terms.

The purpose of my knowledge base was to contain information regarding Roger Federer, a professional tennis player from Switzerland. In order to create the knowledge base, I first had to take the list of the 40 most frequent terms across the web scraper's results and then narrow them down to only 10 terms using my own domain knowledge. Afterwards, I created a dictionary using those 10 terms as keys. The value for each of the keys was the key's definition, which I tried to keep as relevant to Federer's tennis career as possible.

A screenshot of knowledge base is below:

```
{'federer': 'Roger Federer is a professional tennis player from Switzerland, and is
considered by many to be the best tennis player of our time.',
 'open': 'Most tennis tournaments are called opens (e.g., the Korean Open).',
 'grand': 'Four annual tennis tournaments are called grand slams, and these are the
most competitive tournaments to play in.',
 'wimbledon': 'The most famous of the four grand slams, Federer has won this
tournament 8 times, the most out of any player ever.',
 'dom': 'Dominic Theim is a younger player with the potential to replace Federer as
the world No. 1, having already beat him 5 times out of their 7 games.',
 'us': 'The US Open is another grand slam tournament, with Federer being one of only
three players to reach 5 wins in the tournament.',
 'return': "Federer's return is one of his top strengths in tennis, which allows him
to return powerful serves from players like Andy Murray.",
 'tennis': 'The tennis world has been dominated by Federer for several years, with few
players being able to compete with him.',
 'djokovic': 'Novak Djokovic is one of the few players that can compete with Federer,
and is notorious for his two-handed backhand.',
 'nadal': "Rafael Nadal is Federer's lifetime rival, having about the same amount of
achievements."}
```

2. Write up a sample dialog you would like to create with a chatbot based on your knowledge base

Hi, I'm RogerBot, a fan-bot of Roger Federer. I love to talk about Roger Federer! What is your name?

    I'm Ishaaq.

Hi Ishaaq. Do you like Roger Federer?

    Yes, I am a big fan of his and admire his tennis skills.

Me too! What is your favorite Roger Federer tournament?

    Oh, I don't know. But it has to be from 2017, after he recovered from his injuries.

If you had to choose between the Australian Open or Wimbledon, which would it be?

    I guess Australian Open, but maybe because that's the first time I watched him live.

The Australian Open is a good choice. Did you know |random fact|?

    No, I didn't know that.