

Austin Girouard and Aarya Patil

ATG180001 / AAP180014

Dr. Karen Mazidi

## Assignment 6: Finding or Building a Corpus

We created our knowledge base by selecting 10 interesting and unique terms from the top 30 most frequent terms across all 15 URLs. We found that these words would make a cool knowledge base to form the foundation of a chat bot about with general knowledge about the movie “Interstellar.” The knowledge base is fairly simple, containing definitions and facts for each of the terms.

Here are some screenshots from our knowledge base.

```
kb_dict['film'] = 'A film - Also called a movie, motion picture, moving picture, picture, photoplay or flick - is ' \
    'a work of visual art that simulates experiences and otherwise communicates ideas, stories, ' \
    'perceptions, feelings, beauty, or atmosphere through the use of moving images.'
kb_dict['christopher nolan'] = 'Christopher Edward Nolan CBE is a British-American filmmaker. Known for his ' \
    'Hollywood blockbusters with complex storytelling, Nolan is considered a leading ' \
    'filmmaker of the 21st century. His films have grossed $5 billion worldwide.'
kb_dict['space'] = 'Space is a three-dimensional continuum containing positions and directions. In classical ' \
    'physics, physical space is often conceived in three linear dimensions. Modern physicists ' \
    'usually consider it, with time, to be part of a boundless four-dimensional continuum known as ' \
    'spacetime.'
kb_dict['role'] = 'A role is one of the characters that an actor or singer can play in a film, play, or opera.'
kb_dict['matt damon'] = 'Matthew Paige Damon is an American actor, film producer, and screenwriter. Ranked among ' \
    'Forbes\' most bankable stars, the films in which he has appeared have collectively ' \
    'earned over $3.88 billion at the North American box office, making him one of the ' \
    'highest-grossing actors of all time.'
kb_dict['time'] = 'Time is the continued sequence of existence and events that occurs in an apparently ' \
    'irreversible succession from the past, through the present, into the future.'
kb_dict['drama'] = 'Drama is the specific mode of fiction represented in performance: a play, opera, mime, ' \
    'ballet, etc., performed in a theatre, or on radio or television.'
```

```
kb_dict['actress'] = 'A woman whose profession is acting on the stage, in movies, or on television.'
kb_dict['world'] = 'The world can be used to reference Earth. In its most general sense, the term "world" refers ' \
    'to the totality of entities, to the whole of reality or to everything that is. The nature of ' \
    'the world has been conceptualized differently in different fields. Some conceptions see the ' \
    'world as unique while others talk of a "plurality of worlds".'
kb_dict['character'] = 'A character (or speaker, in poetry) is a human or other entity in a narrative (such as a ' \
    'novel, play, radio or television series, music, film, or video game).'
```

Here is the list of the 10 terms we chose, followed by its calculated term frequency (tf):

film : 0.017829402626747634

nolan : 0.006037282869651179

space : 0.00536647366191216

role : 0.004625052958621664

damon : 0.004413218471967236

time : 0.0030009885609377208

drama : 0.0025067080920773903

actress : 0.0024360965965259146

world : 0.0023654851009744385

character : 0.0023654851009744385

A sample dialog for the chatbot based on the knowledge base could be implemented as shown below.

Jerry: Hello! I recently watched the movie Interstellar, and I would love to learn more about Matt Damon.

Chatbot: Matthew Paige Damon is an American actor, film producer, and screenwriter. Ranked among Forbes' most bankable stars, the films in which he has appeared have collectively earned over \$3.88 billion at the North American box office, making him one of the highest-grossing actors of all time.

Jerry: That's so cool! Is there a difference between a film and a movie?

Chatbot: The terms are interchangeable! A film - Also called a movie, motion picture, moving picture, picture, photoplay or flick - is a work of visual art that simulates experiences and otherwise communicates ideas, stories, perceptions, feelings, beauty, or atmosphere through the use of moving images.

Jerry: Good to know. Thank you!

Chatbot: You're welcome! Goodbye.