

ATAI Chatbot Report

Rafael Dubach	18-75-140	rafael.dubach@uzh.ch	rafael.dubach_bot
---------------	-----------	----------------------	-------------------

1. Introduction

The Chatbot is implemented in Python and contained in a Jupyter Notebook. The knowledge graph, embeddings, and other files are downloaded at startup to ensure the latest version is used and to avoid copyright infringement. Using a Jupyter Notebook allows for the use of Google Collab and ensures compatibility with the ARM64 architecture of my machine.

2. Requirements

My Chatbot is capable of answering all five types of questions as required by the teaching team. The provided example questions have been answered and match the provided answers. The outputs can be found at the end of this document.

2.1 Factual questions

Factual questions get answered with the provided knowledge graph. After the question gets processed and an intent gets set, the question gets answered with the provided knowledge graph. This is done with this list comprehension snippet `"graph_ret = [o for s,p,o in graph.triples((entity.uri, question.predicates.uri, None))]"`.

As example, the provided entity.uri would be the link to Good Will Hunting (Q193835) and the question.predicate.uri would be the link to Director (P57). `Graph_ret` would therefore result in the link to Gus Van Sant (Q25186). If a question can't be answered with the provided knowledge graph, the chatbot tries to answer the question with embeddings. But this does not happen with factual questions

2.2 Embedding questions

Embedding questions get answered with the provided embeddings. The embeddings which are provided are not perfectly correct, that's why often the first three answers are provided which was discussed in the Olat forum. The calculation is done as discussed in the lecture by calculating the distance with `lhs = head + pred`. The rating questions received their own intent to answer the questions in a more precise way. They still get answered with embeddings, but the bot provides not only "PG-13" but actually returns the description of what PG-13 means.

2.3 Multimedia questions

Multimedia questions are answered by reading in the json file and searching the knowledge graph for the relevant ID and picture. If found, the picture is returned directly in the user interface without a link. This allows for a seamless and user-friendly experience.

2.4 Recommendation questions

The process of providing recommendations is complex and requires careful consideration. The project description states that the knowledge graph can be used to answer recommendation questions, but after testing both methods, it was determined that using embeddings produces superior results. As a result, the Chatbot uses embeddings to provide recommendations. To handle different entities, the Chatbot creates a list for each entity containing movie suggestions and their corresponding scores as a dictionary. These lists are then processed and ranked based on similarity and score to produce the final recommendation. This process takes into account the likelihood of a suggestion being a good recommendation if it appears multiple times in the processed lists. After preprocessing and ranking the suggestions, the answer is returned to the user. This feature is highly effective and is highlighted for its usefulness.

Q: Recommend movies similar to Hamlet and Othello.

A: My recommendations are Henry V, Richard III, Brief Encounter, Becket

Type	Movie	Based on	Author	Genre	Year
Q	Hamlet	Hamlet	William Shakespeare	Drama	1948
Q	Othello	Othello	William Shakespeare	Drama	1965
A	Henry V	Henry V	William Shakespeare	Drama	1944
A	Richard III	Richard III	William Shakespeare	Drama	1995
A	Brief Encounter	Still Life	Noël Coward	Drama	1945
A	Becket	Becket ou l'Honneur de Dieu	Jean Anouilh	Drama	1964

2.5 Crowdsourcing questions

For the crowdsourcing questions, the dataset is preprocessed to remove malicious workers. Malicious workers are defined as those who take less than 10 seconds to answer a question and have a lifetime approval rate lower than 50 percent. These values can be adjusted as needed to fit the specific requirements of the task. Fleiss Kappa is then calculated for each batch, using a custom function as prebuilt options do not support the necessary number of raters.

2.5.1 User Engagement Statistics

I cleaned the crowd data by filtering out workers, which have a lifetime approval rate of < 50%. Also I filtered out tasks which were solved in under 10 seconds.

Number of tasks accomplished by the crowd workers per batch	
Batch	Number
7QT	63 Questions answered
8QT	60 Questions answered
9QT	64 Questions answered

Throughput as the number of tasks completed per minute	
Batch	Number
7QT	0.40623 Questions per Minute
8QT	0.31729 Questions per Minute
9QT	0.34154 Questions per Minute

Number of tasks skipped by the crowd workers per batch (if it applies)	
Batch	Number
7QT	0
8QT	0
9QT	16

3. Picking the right movie entity

When searching for a specific movie in the data, the Chatbot must determine which movie the user is likely looking for. To differentiate between movies with similar titles, the Chatbot checks the number of awards each movie has received. As awards are a good measure of popularity, the movie with the most awards is preferred. If this is not sufficient to identify the desired movie, the oldest movie found is returned to the user. This ensures that the user receives the most relevant information possible.

4. Last words

There is so much more to this chatbot. But since I'm restricted to two pages, this is it. On the next pages you will find all answered questions which the teaching team gave us. They are all answered correctly.

5. Attachement

5.0 Fulfilment of requirements

In this section I'd like to demonstrate that all example questions are answered exactly as required.

5.1 Factual questions

Question	Answer
Who is the director of Good Will Hunting?	My answer is Gus Van Sant
Who directed The Bridge on the River Kwai?	My answer is David Lean
Who is the director of Star Wars: Episode VI - Return of the Jedi?	My answer is Richard Marquand

5.2 Embedding questions

Question	Answer
Who is the screenwriter of The Masked Gang: Cyprus?	My answer is Cengiz Küçükayvaz, Murat Aslan, Melih Ekener
What is the MPAA film rating of Weathering with You?	The rating of Weathering with You? is PG-13 – Parents Strongly Cautioned
What is the genre of Good Neighbors?	The genre of Good Neighbors is drama, comedy-drama, comedy film

5.3 Multimedia questions

Question	Answer
Show me a picture of Halle Berry.	<p>Here is an image of Halle Berry image:0353/rm3257480192</p> 
What does Julia Roberts look like?	<p>Here is an image of Julia Roberts image:3739/rm3651656960</p> 
Let me know what Sandra Bullock looks like.	<p>Here is an image of Sandra Bullock image:0453/rm2393611008</p> 

5.4 Recommendation questions

Question	Answer
Recommend movies similar to Hamlet and Othello.	Oh you should watch the following: Henry V, Richard III, Brief Encounter, Becket
Given that I like The Lion King, Pocahontas, and The Beauty and the Beast, can you recommend some movies?	My recommendations would include: The Hunchback of Notre Dame, Tarzan, Treasure Planet, Aladdin
Recommend movies like Nightmare on Elm Street, Friday the 13th, and Halloween.	Good choices! You could watch the following: What's the Matter with Helen?, Wes Craven's New Nightmare, Halloween II, Friday the 13th Part 3

5.5 Crowdsourcing questions

Question	Answer
What is the box office of The Princess and the Frog?	My answer is 267000000 - according to the crowd, who had an inter-rater agreement of 0.24 in this batch. The answer distribution for this specific task was 2 support votes and 1 reject vote (If it has more reject votes than support votes, then the answer is the fix value from the crowd)
Can you tell me the publication date of Tom Meets Zizou?	My answer is 2011-01-01 - according to the crowd, who had an inter-rater agreement of 0.04 in this batch. The answer distribution for this specific task was 0 support votes and 3 reject vote (If it has more reject votes than support votes, then the answer is the fix value from the crowd).
Who is the executive producer of X-Men: First Class?	My answer is Sheryl Lee Ralph - according to the crowd, who had an inter-rater agreement of 0.23 in this batch. The answer distribution for this specific task was 2 support votes and 1 reject vote (If it has more reject votes than support votes, then the answer is the fix value from the crowd).