FCIM.FIA Autumn 2024

Lab: Natural Language Processing and Chat Bots

Handed out: Tuesday, November 19, 2024

Please, I need some help!

Our dear city... Our beloved Luna-City, the first human settlement on the Moon, has grown into a radiant (not directly) community known for its diversity, attracting not only tourists eager to experience life beyond Earth but also a growing population of residents who call the city their home. As the city expands, navigating its complex infrastructure and accessing essential information has become increasingly challenging for visitors and locals.

The government has always prided itself on innovation and accessibility for all people. The local organization, traditionally responsible for guiding citizens and visitors, the Lunar Information Institute, is now overwhelmed by the huge volume of inquiries. Being in a quite tickling position, the government decides it needs help with people's assistance tasks!

Remembering previous encounters, the city government considered handling this task to "HeinleinAI", being the best idea! The required development of a system capable of providing directions, offering basic knowledge about Luna-City's services, and attractions, and assisting with daily life in Luna-City. It should also serve as the first point of contact for both tourists and residents, helping without needing to visit the Lunar Information Institute in person.

However, more challenges are yet to be discovered! It turned out the government does not have any data to provide to fulfil their demand, so the company will need to cover this part by itself! Also, there is only one platform, Telegram, available for all residents and tourists visiting Luna-City, so providing the solution to a large mass is limited to it!

However, knowing your previous works, everyone is sure the system you will develop, will be the best possible solution!

General Guidelines

- Submit your solution as a .zip archive, containing .ipynb or .py files and the report (if applicable).
- Do NOT host your solution in public repositories (e.g. Github etc). You can use private repositories if you need to.
- **Plagiarism is NOT tolerated!**

Grading Policy

Task 1 Set up the Telegram Bot. Interact with BotFather on Telegram to obtain an API token. Create your Telegram Bot (its name should follow the pattern *FIA_Surname_Name_FAF_21x*). Make sure you are able to receive and send requests to it. **(1p.)**

Task 2 Create a dataset that will serve as a training set for your model. It should follow the rules:

- an entry consists of two parts: the question and the answer;
- there are at least 75 entries written by you in your dataset;
- questions should be something tourists or locals can ask about a new city.

You can increase your dataset by adding open-source data. However, you MUST clearly show the questions written by you.
Split your dataset into train and validation.
Hint: it is recommended to split it into 80% and 20%, but you can adjust it according to your needs **(1p.)**

Task 3 Use Tensorflow or Pytorch to implement the architecture of the Neural Network you are planning to use. It is highly recommended to use a Seq2Seq model (implement an LSTM or GRU architecture). You are NOT allowed to use pre-built or existing solutions (yep, connecting to GPT will not work). **(3p.)**

Task 4 Train your model and fine-tune it based on the chosen performance metrics. **(1p.)**

Task 5 Integrate your model into your Telegram ChatBot, so that the sent messages are taken as input by the model and its output is sent back as a reply. **(2p.)**

Task 6 Handle potential errors that may occur, such as model errors or invalid inputs. **(1p.)**

Report & Presentation Clear explanations, report formatting, code quality, comments in the code, docstrings, visualisations if relevant etc. **(1p.)**

Good Luck!