

Programming assignment 3

Rituraj Kulshresth (B18CSE046)

Saurabh Shashikant Burewar (B18CSE050)

Data

The dataset used is the KnowledgeNet dataset (<https://github.com/diffbot/knowledge-net>). This dataset contains passages of texts and a list of facts along with the properties like DATE_FOUNDED, HEADQUARTERS, SUBSIDIARY_OF, FOUNDED_BY, CEO, etc.

Processing Data

We parse through the data and make a subset of the 150 facts containing the relations “CEO” in the file “ceosubset.json” and use that as our dataset. Our dataset is a list of facts where each fact contains an id, text, subject, relation and object.

We also have a list of all companies in “compnaynames.txt” and a list of CEOs in “ceonames.txt”.

These data files can be created by running the following python files in the given order -

python data.py

python datatxt.py

```
{
  "company": [
    {
      "id": "7183:58:80:18:34:4",
      "Text": "John F. Lundgren is CEO and Chairman of Stanley Black & Decker (SWK), a position he has held since 2005.",
      "Subject": "Stanley Black & Decker",
      "Relation": "CEO",
      "Object": "John F. Lundgren"
    },
    {
      "id": "7183:82:85:18:34:4",
      "Text": "John F. Lundgren is CEO and Chairman of Stanley Black & Decker (SWK), a position he has held since 2005.",
      "Subject": "SWK",
      "Relation": "CEO",
      "Object": "John F. Lundgren"
    },
    {
      "id": "7089:69:96:18:34:4",
      "Text": "Stephen Haddrill is Chief Executive Officer of the Financial Reporting Council (FRC), a position he has held since 2005.",
      "Subject": "Financial Reporting Council",
      "Relation": "CEO",
      "Object": "Stephen Haddrill"
    },
    {
      "id": "7089:98:101:18:34:4",
      "Text": "Stephen Haddrill is Chief Executive Officer of the Financial Reporting Council (FRC), a position he has held since 2005.",
      "Subject": "FRC",
      "Relation": "CEO",
      "Object": "Stephen Haddrill"
    }
  ]
}
```

ceosubset.json

```
Stanley Black & Decker
SWK
Financial Reporting Council
FRC
National League for Nursing
Real News Network
Sorgente Group
St Kilda Youth Service
WestPoint Home
Rolls-Royce
Manhattan Associates
MANH
Blue Jeans Music BMI
James Allen Promotions
IB Group
APS
Pinnacle West
Pinnacle West
Hackensack Trust Co.
Garden State National Bank
Disneyland of Dairy Stores'
```

compnaynames.txt

```
John F. Lundgren
John F. Lundgren
Stephen Haddrill
Stephen Haddrill
Beverly Malone
Paul Jay
Valter Mainetti
Sillers
Normand Savaria
John Frederick Rishton
Dabbieri
Dabbieri
Stephen A. Love
Stephen A. Love
Thomas Peterffy
Brandt
```

ceonames.txt

Experiments

We use RASA open source to create our chatbot.

RASA has YAML files that contain information about entities, intents and actions. In our case, we have a chatbot that has information about companies and their CEOs.

We put our JSON and text files created above in the /data directory.

In “nlu.yml”, we define our intents. This includes classic intents like greet, goodbye, tell a joke, etc. which we took from the RASA demo application. We added intents for -

- FAQs - These include questions like what a company is and what is a ceo?
- Confirm exists - These include questions regarding whether a company or a ceo exists like is __ a company or is __ a ceo?
- Query knowledge base - These include general queries like mentioning a few companies.

In “responses.yml”, we define the responses for our FAQs.

In “stories.yml”, we define certain example stories for our intents and actions.

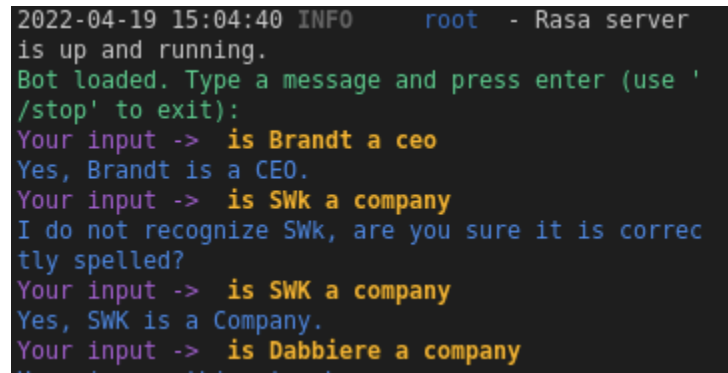
In “config.yml”, we define the policies and number of epochs for training.

In “domain.yml”, we connect all the other files by mentioning all the entities, intents, actions and responses.

In “actions.py”, we have one class to check existence. This uses the text files created above as knowledge and if asked if a company or a CEO exists, it gives the appropriate response depending on whether that name is in our knowledge.

We also have a memory knowledge base class which queries the JSON knowledge base.

Below is the screenshot of the chatbot -



```
2022-04-19 15:04:40 INFO      root - Rasa server
is up and running.
Bot loaded. Type a message and press enter (use '
/stop' to exit):
Your input -> is Brandt a ceo
Yes, Brandt is a CEO.
Your input -> is SWk a company
I do not recognize SWk, are you sure it is correc
tly spelled?
Your input -> is SWK a company
Yes, SWK is a Company.
Your input -> is Dabbieri a company
```

Future work

- We can make the bot learn from the passages and answer the questions smartly.
- Currently the bot only answers questions related to ceos. We can make it answer questions related to other details in the facts
- We can increase our database to contain more information about companies and CEOs.

Contribution

Rituraj Kulshresth (B18CSE046) - 50% of everything

Saurabh Shashikant Burewar (B18CSE050) - 50% of everything

Coding was done in Pair programming and the report was written in google docs together.

References

- <https://rasa.com/docs/rasa/>
- <https://www.youtube.com/watch?v=sazsWmP2d3o>
- <https://github.com/RasaHQ/pokedex-demo>
- <https://github.com/RasaHQ/rasa>