Google
Summer of Code

Donkeybot

Vasilis Mageirakos, GSoC ‘20

Donkeybot

Premise, resources and desired outcome.

Concept :

Due to the vast amount of support requests, we are looking into methods to assist the support team in answering these requests. Ideally, the support would be provided by an intelligent bot able to process and understand the user's requests and finally trigger appropriate action.

Gather and analyse the relevant data

Create the bot able to identify questions and answers

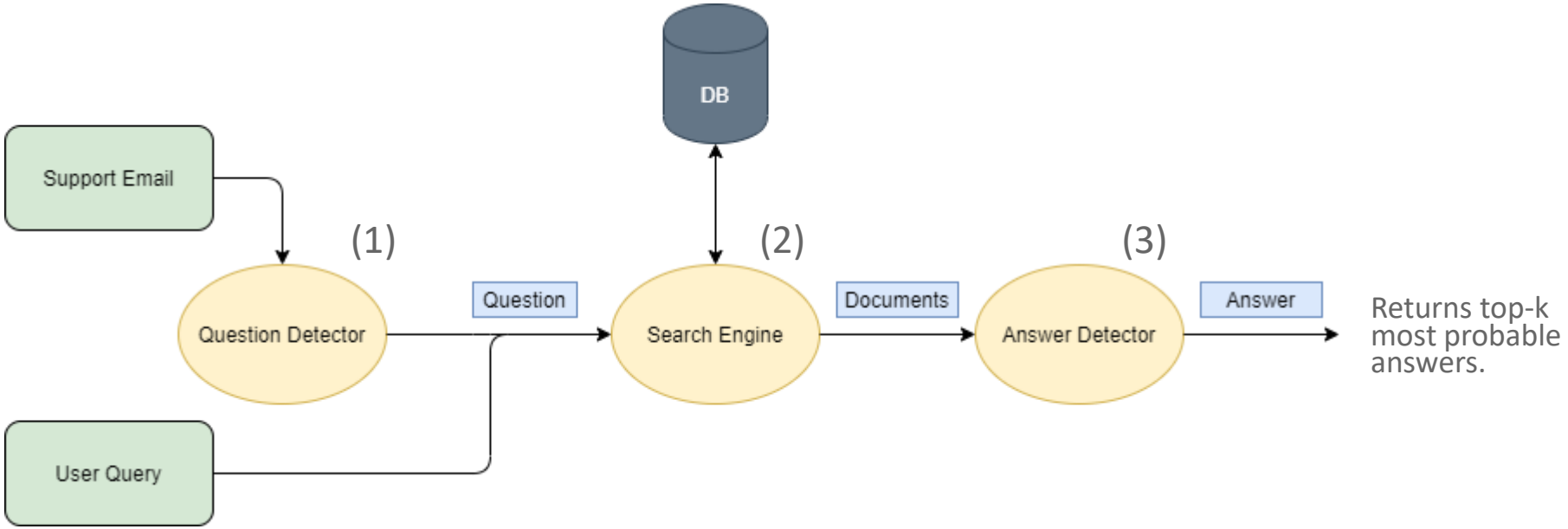
Generate appropriate answers on new user queries

Deliverables :

1. Fetching and parsing of emails, GitHub issues and Rucio documentation.
2. Prototype Bot creation able to utilize the above data.
3. Question-answering prototype pipeline and demo.

Question-Answering Pipeline

Question Answering system's architecture.



(1) Uses regex patterns to extract Questions from support emails.

(2) Uses BM25 algorithm to retrieve top-k most relevant documents.

(3) Uses BERT to find answers in each document retrieved with corresponding confidence scores.

Examples 1/2

Let’s look at some Rucio specific questions.

Source : Rucio support emails

```
Question :  When is a dataset considered touched by the system?

number 1 asnwer (by confidence)
[ { 'answer': 'physically downloaded',
    'confidence': 0.5526032861544721,
    'extended_answer': 'uch these files, they must be physically downloaded to '
                       'be considered touched. And'}]
```

```
Question :  When does a touch happen in the system?

number 1 asnwer (by confidence)
[ { 'answer': 'when the dataset is used as input for a panda task or when '
              'rucio download is used to access the data.',
    'confidence': 0.5806299379923985,
    'extended_answer': 'Hi fac8a3, A "touch" occurs when the dataset is used '
                       'as input for a panda task or when rucio download is '
                       "used to access the data. I don't see any tasks "
                       'defined'}]
```

Examples 2/2

Let’s look at some more Rucio specific questions.

Source : Rucio documentation

```
Question :  What are the supported databases for Rucio?

number 1 asnwer (by confidence)
[ { 'answer': 'MySQL, PostgreSQL, Oracle, and SQLite',
    'confidence': 0.9758034113866643,
    'extended_answer': 'parate container. It supports MySQL, PostgreSQL, '
                       'Oracle, and SQLite as database backends.\n'
                       '\n'
                       'This i',
    'metadata': { 'bm25_score': 2.2687560321466513,
                  'doc_id': 17,
                  'doc_type': 'general',
                  'name': 'installing_daemons.rst',
                  'url': 'https://github.com/rucio/rucio/blob/master/doc/source/installing_daemons.rst'}}]
```

```
Question :  How are rucio users authenticated?

number 1 asnwer (by confidence)
[ { 'answer': 'by credentials,',
    'confidence': 0.8449646327554738,
    'extended_answer': 'A Rucio user is authenticated by credentials, such as '
                       'X509 certificates,\n'
                       'us',
    'metadata': { 'bm25_score': 4.5579457311873846,
                  'doc_id': 55,
                  'doc_type': 'general',
                  'name': 'overview_Rucio_account.rst',
                  'url': 'https://github.com/rucio/rucio/blob/master/doc/source/overview_Rucio_account.rst'}}]
```

More information

Useful links on sources mentioned.

Donkeybot repository : https://github.com/rucio/donkeybot

Contact : b.mageirakos@gmail.com (or Slack)

GitHub (@mageirakos) : https://github.com/mageirakos

Google Summer of Code : https://summerofcode.withgoogle.com/

GSoC project : https://bit.ly/3i9we8H