



Boom AI Assessment

Verlon Salaysay

#3 - Deploy a question-answering model using FastAPI and Docker by following this documentation. Below are the specifications:

- Demonstrate the use of HuggingFace question-answering model
- Use FastAPI to expose an endpoint
- Use Docker to serve and deploy the endpoint
- Bonus points: Host it on a free AWS EC2/GCP VM/DigitalOcean Droplet

Solution Reference:

<https://github.com/verlonsalaysay/qa-model>

• Demonstrate the use of HuggingFace question-answering model

```
In [18]: new_questions = [
    'How many teams compete in the Premier League ?',
    'When does the Premier League starts and finishes ?',
    'Who has the highest number of goals in the Premier League ?',
]

json_data = {
    'questions':new_questions,
}

response = requests.post(
    'http://0.0.0.0:8000/get_answer',
    json=json_data
)

for d in response.json():
    print('\n'.join(["{} : {}".format(k, v) for k,v in d.items()])+'\n')

orig_q : How many teams compete in the Premier League ?
best_q : How many clubs are currently in the Premier League?
best_a : 20

orig_q : When does the Premier League starts and finishes ?
best_q : When does the Premier League have its playing season?
best_a : During the course of a season (from August to May)

orig_q : Who has the highest number of goals in the Premier League ?
best_q : Who has the record for most goals in the Premier League?
best_a : Newcastle United striker Alan Shearer holds the record for most Premier League goals wi
```

```
In [15]: new_questions = [
    'How much money per year does the League generate in worldwide television?',
    'When did they publish Taylor Report?',
    'When did they sign the Founder Members Agreement',
]

json_data = {
    'questions':new_questions,
}

response = requests.post(
    'http://0.0.0.0:8000/get_answer',
    json=json_data
)

for d in response.json():
    print('\n'.join(["{} : {}".format(k, v) for k,v in d.items()])+'\n')

orig_q : How much money per year does the League generate in worldwide television?
best_q : How much revenue does the league generate in domestic and international television rights?
best_a : The league generates €2.2 billion per year in domestic and international television rights

orig_q : When did they publish Taylor Report?
best_q : In which year was the Taylor Report published?
best_a : 1990

orig_q : When did they sign the Founder Members Agreement
best_q : When was the Founder Members Agreement signed?
best_a : The Founder Members Agreement, signed on 17 July 1991
```

- Use FastAPI to expose an endpoint

```
47 lines (36 sloc) | 1.14 KB

1  import uvicorn
2  from fastapi import FastAPI, Request
3  from utils import QASearcher
4
5  app = FastAPI()
6
7  @app.post("/set_context")
8  async def set_context(data: Request):
9      """
10     Fastapi POST method that sets the QA context for search.
11
12     Args:
13         data(`dict`): Two fields required 'questions' (`list` of `str`)
14            and 'answers' (`list` of `str`)
15     """
16     data = await data.json()
17
18     qa_search.set_context_qa(
19         data['questions'],
20         data['answers']
21     )
```

- Use Docker to serve and deploy the endpoint

```
dev@dev-Standard-PC-i440FX-PIIX-1996:~/boomAI-qa$ sudo docker run -p 8080:8080 qamodel2
INFO:      Started server process [1]
INFO:      Waiting for application startup.
INFO:      Application startup complete.
INFO:      Uvicorn running on http://0.0.0.0:8000 (Press CTRL+C to quit)
```