

Обработка языка. Практика 5

Генерация текста

Михаил Пожидаев

24 октября 2023 г.

Подготовка

```
from sys import argv
import numpy as np
import torch
from transformers import GPT2LMHeadModel, GPT2Tokenizer
np.random.seed(42)
torch.manual_seed(42)
```

Загрузка

```
def load_tokenizer_and_model\  
    (model_name_or_path):  
    return GPT2Tokenizer.from_pretrained( \  
        model_name_or_path \  
    ), GPT2LMHeadModel.from_pretrained(model_name_or_path)
```

Запуск

```
tok, model = load_tokenizer_and_model \  
# Альтернативно "ai-forever/ruGPT-3.5-13B" \  
  ("sberbank-ai/rugpt3large_based_on_gpt2") \  
generated = generate \  
  (model, tok, argv[1], num_beams=10) \  
print(generated[0])
```

Пример

Запрос

Если пойти по Томску,...

Результат

Если пойти по Томску, то можно попасть в музей деревянного зодчества под открытым небом «Усадьба Рукавишниковых», который находится на территории музея-заповедника «Томская писаница» и является одним из старейших деревянных усадеб России.

Спасибо за внимание!

Всё о курсе: <https://marigostra.ru/materials/nlp.html>

E-mail: msp@luwrain.org

Канал в Телеграм: <https://t.me/MarigostraRu>