

Импортируем библиотеки

```
import re
from collections import defaultdict
import random
from bs4 import BeautifulSoup
import requests
```

Создаём биграмную модель через класс

```
class Bigram():

    def __init__(self):
        self.corpus = ''
        self.dict_bigrams = defaultdict(list)

    def fit(self, corpus):
        assert isinstance(corpus, str), 'в fit должна подаваться строка'
        assert '.' in corpus, 'в тексте обязательно должна быть хотя бы одна точка'

        self.corpus = re.compile('[^a-zA-Z0-9a-яA-Я. ]').sub('', corpus.lower())
        self.corpus = self.corpus.replace('.', ' [END] [START]')
        self.corpus = '[START] ' + self.corpus
        self.corpus = self.corpus.split()
        for first_word, second_word in zip(self.corpus, self.corpus[1:]):
            self.dict_bigrams[first_word].append(second_word)

    def generate(self):
        assert self.corpus != '', 'Нельзя применять методы generate не обучив модель с пом

        first_word = "[START]"
        result = []
        while True:
            next_word_candidates = self.dict_bigrams[first_word]
            first_word = random.choice(next_word_candidates)
            result.append(first_word)
            if first_word == "[END]": return " ".join(result[:-1]) + '.'
```

Возьмём в качестве корпуса текста обзор Supreme Commander на сайте Stopgame.ru

```
url = "https://stopgame.ru/show/20664/supreme_commander_review"
html = requests.get(url).text
soup = BeautifulSoup(html, 'html5lib')
corpus = soup.find('section', attrs={"class": "article article-show"}).text
```

Создаём модель и обучим её

```
bigram = Bigram()
```

```
bigram.fit(corpus)
```

Сгенерируем текст

```
bigram.generate()
```

```
↳ 'в богатом выборе тактических возможностей.'
```

Работает!