

Итоговый тест по главе

1. Каков порядок работы конвейера языкового моделирования?

- ☐ Сначала модель, которая обрабатывает текст и возвращает необработанные прогнозы. Затем токенизатор осмысливает эти предсказания и при необходимости преобразует их обратно в текст.
- ☐ Сначала работает токенизатор, который обрабатывает текст и возвращает идентификаторы. Модель обрабатывает эти идентификаторы и выдает прогноз, который может представлять собой некоторый текст.
- ☒ Токенизатор обрабатывает текст и возвращает идентификаторы. Модель обрабатывает эти идентификаторы и выдает прогноз. Затем токенизатор можно снова использовать для преобразования этих прогнозов обратно в текст.

Correct! Правильно! Токенизатор может использоваться как для токенизации, так и для де-токенизации.

Submit

You got all the answers!

2. Сколько измерений имеет тензор, выводимый базовой моделью Transformer, и каковы они?

- ☐ 2: Длина последовательности и размер батча
- ☐ 2: Длина последовательности и скрытый размер
- ☒ 3: Длина последовательности, размер батча и скрытый размер

Correct! Верно!

Submit

You got all the answers!

3. Что из перечисленного ниже является примером токенизации по подсловам?

- ☒ WordPiece

Correct! Да, это один из примеров токенизации по подсловам!

- ☐ Посимвольная токенизация
- ☐ Разделение по пробелам и пунктуации
- ☒ BPE

Correct! Да, это один из примеров токенизации по подсловам!

- ☒ Unigram

Correct! Да, это один из примеров токенизации по подсловам!

- ☐ Ничего из вышеперечисленного

Submit

You got all the answers!

4. Что такое голова модели?

- ☐ Компонент базовой сети Transformer, который перенаправляет тензоры на корректные слои
- ☐ Также известный как механизм самовнимания, он адаптирует представление токена в соответствии с другими токенами последовательности
- ☒ Дополнительный компонент, обычно состоящий из одного или нескольких слоев, для преобразования прогнозов трансформера в специфический для конкретной задачи выход

Correct! Именно так. Адапционные головы, также известные просто как головы, встречаются в разных формах: головы языкового моделирования, головы ответов на вопросы, головы классификации последовательностей...

Submit

You got all the answers!

5. Что такое AutoModel?

- ☐ Модель, которая автоматически обучается на ваших данных
- ☒ Объект, возвращающий правильную архитектуру на основе контрольной точки

Correct! Именно: в AutoModel для возврата правильной архитектуры достаточно знать контрольную точку, с которой нужно инициализироваться.

- ☐ Модель, которая автоматически определяет язык, используемый для входов, чтобы загрузить правильные веса

Submit

You got all the answers!

6. На какие техники следует обратить внимание при объединении в батч последовательностей разной длины?

- ☒ Усечение (truncating)

Correct! Да, усечение - это правильный способ выровнять последовательности так, чтобы они вписывались в прямоугольную форму. Но разве это единственный способ?

- ☐ Возврат тензоров
- ☒ Дополнение (padding)

Correct! Да, дополнение - это правильный способ выровнять последовательности так, чтобы они вписывались в прямоугольную форму. Но разве это единственный способ?

- ☒ Маскирование внимания (attention masking)

Correct! Обязательно! Маски внимания имеют первостепенное значение при работе с последовательностями разной длины. Однако это не единственная техника, о которой следует помнить.

Submit

You got all the answers!

7. В чем смысл применения функции SoftMax к логитам, выводимым моделью классификации последовательностей?

- ☐ Она смягчает логиты, чтобы они были более надежными.
- ☒ Она применяет нижнюю и верхнюю границы, чтобы они были понятны.

Correct! Правильно! Результирующие значения ограничены между 0 и 1. Однако это не единственная причина, по которой мы используем функцию SoftMax.

- ☒ Общая сумма вывода равна 1, что позволяет получить вероятностную интерпретацию.

Correct! Правильно! Однако это не единственная причина, по которой мы используем функцию SoftMax.

Submit

You got all the answers!

8. Какой метод является основным в API токенизатора?

- ☐ encode, поскольку он может кодировать текст в идентификаторы и идентификаторы в прогнозы.
- ☒ Вызов объекта токенизатора напрямую.

Correct! Точно! Метод `__call__` токенизатора - это очень мощный метод, который может обрабатывать практически все. Это также метод, используемый для получения прогнозов из модели.

- ☐ pad
- ☐ tokenize

Submit

You got all the answers!

9. Что содержит переменная result в этом примере кода?

```
from transformers import AutoTokenizer

tokenizer = AutoTokenizer.from_pretrained("bert-base-cased")
result = tokenizer.tokenize("Hello!")
```

- ☒ Список строк, каждая из которых является токеном

Correct! Конечно! Преобразуйте это в идентификаторы и отправьте их в модель!

- ☐ Список идентификаторов
- ☐ Строка, содержащая все токены

Submit

You got all the answers!

10. Есть ли что-то неправильное в следующем коде?

```
from transformers import AutoTokenizer, AutoModel

tokenizer = AutoTokenizer.from_pretrained("bert-base-cased")
model = AutoModel.from_pretrained("gpt2")

encoded = tokenizer("Hey!", return_tensors="pt")
result = model(**encoded)
```

- ☐ Нет, кажется, все правильно.
- ☒ Токенизатор и модель всегда должны быть из одной контрольной точки.

Correct! Правильно!

- ☐ Хорошей практикой является дополнении и усечении с помощью токенизатора, поскольку каждый вход - это батч.

Submit

You got all the answers!