

Токенизаторы, проверка!



Отличная работа по завершению этой главы!

После этого глубокого погружения в токенизаторы вы должны:

- Уметь обучать новый токенизатор, используя старый в качестве шаблона.
- Понимать, как использовать смещения для сопоставления позиций токенов с их исходным положением в тексте
- Знать различия между BPE, WordPiece и Unigram.
- Уметь комбинировать блоки, предоставляемые библиотекой 🧐 Tokenizers, для создания собственного токенизатора
- Уметь использовать собственный токенизатор в библиотеке 🧐 Transformers

<> [Update](#) on GitHub