Text Mining 2

文本表示

- [TF-IDF(Term frequency - Inverse document frequency)](#)

  ```
  # 阅读官方文档了解3个类的用法，主要参数；
  from sklearn.feature_extraction.text import TfidfTransformer
  from sklearn.feature_extraction.text import CountVectorizer
  from sklearn.feature_extraction.text import TfidfVectorizer
  ```

- [词袋模型应用](#)

- [什么是Representation Learning?](#)

- [word2vec](#)

- [嵌入表示](#)

- [Gensim: Word2vec实战](#)

- [Doc2Vec](#)

- [2016年前的文本表示回顾](#)

- [EMLO : Embeddings from Language Models](#)

- [Transformer](#)

- [从word2vec, ELMo到BERT](#)

练习

采用[Gensim](#)

- 任务：https://zhuanlan.zhihu.com/p/52705512
- 数据：https://github.com/wolfkin-hth/novels/tree/master