



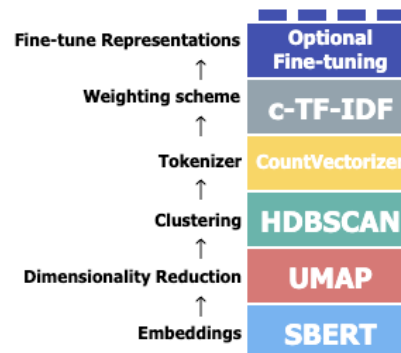
Processamento de Linguagem Natural com Transformers

Processamento de Linguagem Natural com Transformers

O Que é o BERTopic?

BERTopic é uma ferramenta de modelagem de tópicos que utiliza transformadores e estrutura c-TF-IDF (class-based-TF-IDF) para criar clusters densos, permitindo tópicos facilmente interpretáveis, mantendo palavras importantes nas descrições dos tópicos.

O BERTopic é uma junção de várias técnicas que se complementam, incluindo representações usadas nos modelos BERT, que estudamos neste capítulo. A imagem abaixo ilustra as principais técnicas do BERTopic:



O BERTopic é realmente incrível com uma performance surpreendente. E gratuito. Pode ser instalado como um pacote Python:

pip install bertopic

E o BERTopic tem ainda uma vasta documentação, repleta de exemplos e dicas de como usar a ferramenta para resolver problemas de PLN, como o agrupamento de texto por similaridade (a modelagem de tópicos). Acesse a documentação no link abaixo:

<https://maartengr.github.io/BERTopic/index.html>

Detalhes sobre o BERTopic serão apresentados durante as aulas.