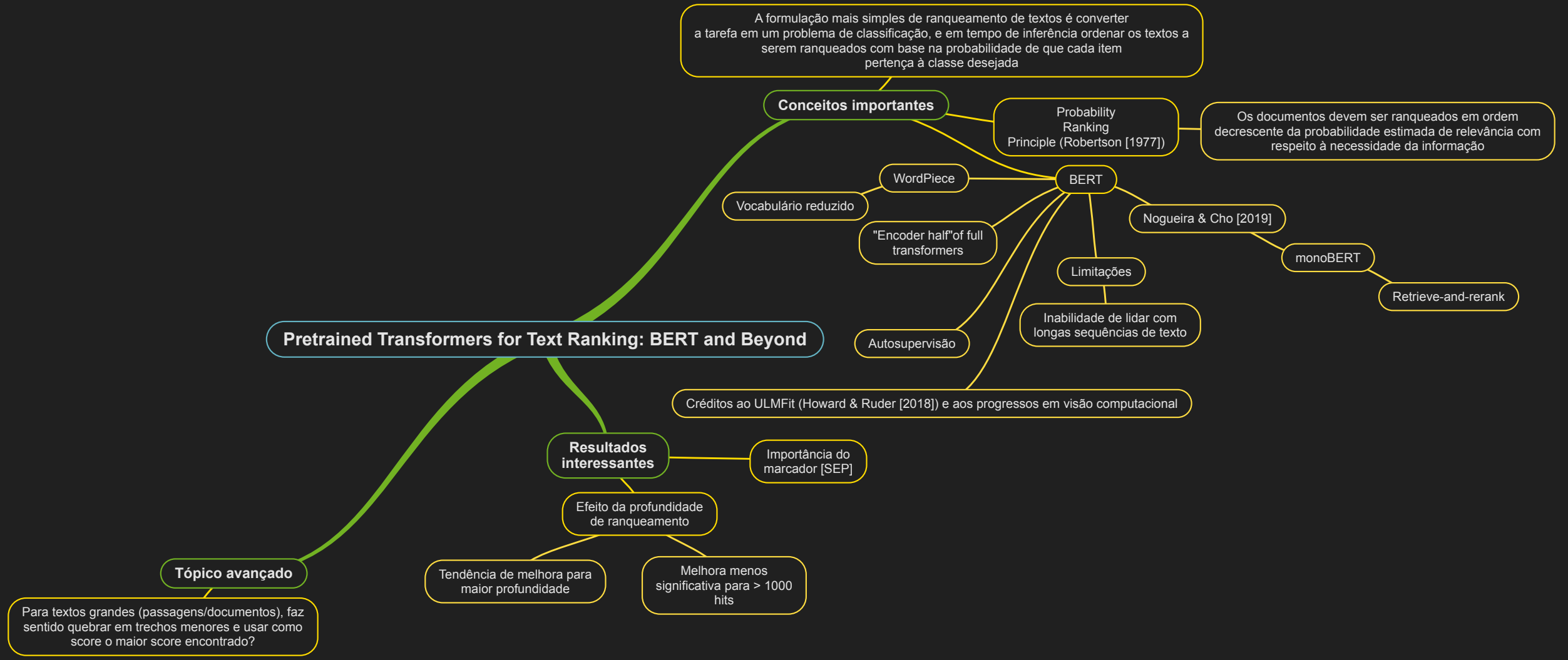
Pretrained Transformers for Text Ranking: BERT and Beyond
A formulação mais simples de ranqueamento de textos é converter a tarefa em um problema de classificação, e em tempo de inferência ordenar os textos a serem ranqueados com base na probabilidade de que cada item pertença à classe desejada
Conceitos importantes
Probability Ranking Principle (Robertson [1977])
Os documentos devem ser ranqueados em ordem decrescente da probabilidade estimada de relevância com respeito à necessidade da informação
BERT
WordPiece
Vocabulário reduzido
"Encoder half"of full transformers
Nogueira & Cho [2019]
monoBERT
Retrieve-and-rerank
Limitações
Inabilidade de lidar com longas sequências de texto
Autosupervisão
Créditos ao ULMFit (Howard & Ruder [2018]) e aos progressos em visão computacional
Resultados interessantes
Importância do marcador [SEP]
Efeito da profundidade de ranqueamento
Tendência de melhora para maior profundidade
Melhora menos significativa para > 1000 hits
Tópico avançado
Para textos grandes (passagens/documentos), faz sentido quebrar em trechos menores e usar como score o maior score encontrado?