

結果比較與質化分析:

BM25 的結果受到關鍵字的出現頻率顯著影響，如果查詢中的關鍵字在文件中頻繁出現，則會得到較高的相似性分數。

例如在 BM25 的結果中，所有檢索結果都因為 AIDS，似乎只要內容包含多次“AIDS”關鍵字就能因此得到高分。

BM25 高度仰賴關鍵字出現次數的特性，可能因此導致一些具有高度相關資訊的文件，但不常直接提及關鍵字而因此被低估。

Sentence-Transformer 能夠認知到查詢關鍵字本身的意義，即使關鍵字直接出現的頻率沒那麼高，但只要上下文意義相似也能被視為相關。

另外，Sentence-Transformer 給出的相似性分數較不會受到文件長度的影響，短文與長文能得到合理的分數。

最後，其中一些文件儘管幾乎不直接提及“AIDS”也被視為相關，可能是因為它們討論了相關的性疾病、政策或教育等資訊，可以見得 Sentence-Transformer 對於文意有一定程度的理解。

例如 index: 10209 中，得到極高的相似度分數: 0.9502177238464355

但該文僅有少次使用“AIDS”這個關鍵詞，但有提到 HIV 以及相關的教育計劃、預防措施等等因此被視為高度相關。