# Analysis of different approaches

BM25 scoring function

The BM25 scoring function does not work well when the keyword searched for is contained in more than half of the corpus documents. In this case, the inverse document frequency function will return a negative value, and a document that does not contain the phrase queried for might be scored higher than a document that contains the query.

However, this behavior can be treated in a number of ways. We implemented one of these fixes, and we fixed it by giving each term in the function a floor of 0.005, to avoid counting very common words.

N-gram

Skip bi-gram

Passage term matching

**Passage term has two major strengths since it checks how often a candidate answer appears in the same passage as the question terms REWRITE.** One of the two major strengths that come from this is that the order of the words and syntactic structure does not matter. The other one is that passage term matching will take into account passages that do not contain the correct answer, but will help extract information from that passage to support candidate answers extracted from other passages.

However, this second strength mentioned can also be a weakness. By using passages that do not contain the correct answer to support other candidate answers, passages that are closely related to the clue but are not actually correct will be assigned too much credit.