

# Sources: What is the relevance algorithm used in Muse?

Muse calculates document relevance using a modified "weighted term frequency" algorithm to calculate a relevance "score" for each document. "Document" here refers to the metadata document returned by the search, not the full text of the original document. The documents are then sorted on the basis of that score. The higher the score, the more relevant the document to the query. While each score is a number, there are no absolute values, so each score only has meaning within its own results set. Such ideas as 'threshold' scores have no meaning.

This algorithm makes no attempt to normalize the frequency of terms according to the total size of the document (thus more 'wordy' documents will be scored higher), or on the occurrence or absence of particular fields. The algorithm generally provides a good ranking order and is optimized for speed, as it runs in real time.

Note that the reliance on query terms in this, as in virtually all other relevance algorithms, means that there will often be documents which score 0.0, because the query terms did not occur in the returned (metadata) portion of the record, even though it may occur in the full document and thus was indexed by the search engine.

Unique solution ID: #1052

Author: Administrator

Last update: 2013-09-02 11:35