How is relevance ranking determined in EBSCOhost?

EBSCOhost employs a comprehensive relevance ranking strategy that utilizes numerous criteria, including term frequency, field weighting, exact field matching, and content attribute boosting, to provide the user with the most relevant results for their search queries. EBSCO's goal is to display the most relevant results on the first page.

The major contributing factor in relevance scoring is the frequency of the user's search terms in matching EBSCOhost metadata and full-text records. Like all search engines, EBSCOhost begins by finding records that contain the words that match the user's search query. Some matching fields are considered more important than others for relevance scoring purposes and are weighted to take advantage of their relative importance.

Maximizing Accuracy with Field Ranking

The fields below are the most influential fields used in relevance ranking calculations, and are listed in order of influence.

1. Subject heading
2. Title
3. Author-supplied keywords
4. Abstract
5. Authors
6. Full-text

Additional metadata fields, beyond those listed above, also contribute to relevance scoring.

Minimizing Full Text Influence

To minimize the influence of high-frequency hits on matching words in full-text documents, such as long scholarly articles, eBooks, and government reports, the full-text field has the lowest field weight. Additionally, the relevance ranking algorithm uses a normalization scoring model so that very high hit counts in full-text documents do not artificially inflate relevance scores for these documents.

Value Ranking

Specific content attributes of matching records may also contribute to relevance scoring. These content attributes include:

- Publication date
- Publication type
Points are added to the relevance score based on criteria not specific to the user's search terms. Examples include:

- EBSCOhost will prioritize ranking of newly published records and peer-reviewed records over older and non-peer-reviewed equivalents.
- Records identified as a publication type of "book review" or "news" receive a lower relevance weighting than other publication types.

Additionally, specific fields may be configured to support a 'Field Match' boost. For example, a relevance scoring boost is applied when the user's query exactly matches the title field of a library catalog record. The boost is configured so that each non-matching term causes the boost points to be decremented by a specified amount down to a minimum boost for partial matches.

Adjacency Bias

Search terms often include adjacent words that form a phrase. A document containing the phrase is usually more relevant than one that contains the words in isolation. Documents with the phrase receive a scoring adjustment in their favor. This adjustment is an implicit phrase bias, (or adjacency bias), for words in exact order adjacent to one another. It applies to search terms that are implicit phrases, meaning the words are not explicitly identified as a phrase by quotation marks. All things being equal, documents with implicit phrases found in citation fields, which match the search terms, will rank higher than documents that contain only the individual words. Adjacency bias does not guarantee that documents with implicit phrases will rank higher than those without the phrases. Also, adjacency bias does not apply to words found in the full text of documents.

Delivering Relevant Results

The various factors described above combine to produce a composite EBSCOhost relevance score for all matching items for a given EBSCOhost search. These scores determine the order of records in the EBSCOhost results list. There is no simple formula for relevance ranking, but a multitude of factors that vary based on the user's query, the content searched, and EBSCOhost profile settings. As an area of active development, we continually tune and improve relevance ranking to give users the results they want for every search, in every context.