

# Lucene Relevance Scoring And Boosting

## Scoring

The search engine calculates a relevance 'score' and ranks search results in that order.

The score is a calculation of how relevant a product is to the search term. Note that if two products both contained "a4 paper" in their description, they might receive different scores. A product would receive a higher score if the words "a4" or "paper" appeared more than once in the indexed text. Also a product with shorter indexed text would be scored higher, as it is considered more relevant if all the words are matched in a shorter description than in a longer one.

The score is calculated by considering the following:

- The **Frequency of search terms in products** - for example, if the word "paper" is included in 8000 of 10000 products, but the word "green" is in a much smaller number of products, the term "green" is considered more relevant than the term "paper"
- The **number of times a search term appears in a product** - for example, if a product contains the word "paper" five times in its description, that product is considered more relevant than another product that might only contain the term "paper" once.
- The **length of the product text** - if a search term is found in a product that contains a short amount of text, that product is considered more relevant than a product with a larger amount of text.
- **Artificial "boost" factors** - administrators can "boost" the relevance of products based on a number of factors. The boost value can increase (when >1) or decrease (when <1) the standard score of a product.

The "Score" and "Boost" values are included in the search output when testing searches through the "Tools" Lucene Administration screen. The Boost Value is calculated at index time, so any changes to boost values require a reindex. Note that the Score displayed includes the boost value in its calculation (i.e. it is the score after boosting).

You can also include an explanation of the score by checking the "Explain" checkbox and clicking on the '>' at the beginning of the row. The Score is basically the addition and/or multiplication of several factors that impact the relevance of the term. These are:

- idf - Document Frequency. Score based on the number of documents that contain the search term - i.e. the smaller the number of documents, the higher the score.
- tf - Term Frequency. Score based on the number of times the term appears in the field.
- fieldNorm - a score calculated by the length of the field (number of words). The smaller the number of words, the higher the fieldNorm. The Boost Value is also included in this value.
- queryNorm - normalisation value of a query. This will be multiplied into the weight of each query term.

From these base values 'fieldWeight' and 'queryWeight' are calculated. These are then multiplied to determine a 'weight' for the search phrase. Multiple search phrases are then added together to get the final score.

## Boosting

The score can be boosted by several "artificial" factors:

### Product Boost

Individual products can be boosted based on:

- Number of times a product has been sold - therefore products that are sold more regularly will be considered more relevant
- "Click-throughs", or the number of times the product detail page has been viewed - products viewed more regularly are considered more relevant
- A Custom Implementation - the Product.BoostSourceValue can be calculated by a custom stored procedure for a site

The boost factor applied is between 1 and a maximum number defined in the Lucene settings. Products with Boost Source Values below a threshold can also be ignored.

The boost values can also be allocated by rank (i.e. evenly from the lowest to the highest product) or by ratio, where the Boost Source Value is included in the boost calculation, rather than the rank of the value.

### Boost Words

Products can be boosted based on whether their text contains certain values. As an example this feature can be used to boost certain brands. If we sell batteries, we can apply a boost factor of 2 to any battery containing the word "duracell", which means that brand of product would be considered more relevant than products without the word "duracell".

Similarly, you can use boost words to decrease the relevance of products. We can apply a boost factor of 0.2 to products containing the word "eveready", which would make those products less relevant.

### Field Boost

The site administrator can also define multiple fields to be included in the search index, and can assign a boost factor to a particular field. For example, a product search index could contain a field for the product description and a field for the product category description. The category description could be given a boost factor > 1 to make search term matches in that field more relevant than the product description

## Related help

- [Product Search with Lucene Search Text Engine](#)
- [Lucene Introduction](#)
- [Lucene Query Translation Maintenance](#)
- [Build or Rebuild the Lucene Search Index](#)
- [Lucene Relevance Based Search - Set Up](#)
- [Lucene Advanced Search Queries](#)
- [Lucene Trim Word Maintenance](#)
- [Set Up Lucene Index Fields](#)
- [Resources And Template Tokens For Lucene Search](#)
- [Lucene Search Tools](#)
- [Configure Lucene - Fast Start](#)
- [Did You Mean - Lucene Search Suggestions](#)