

Stemmed Lucene Search

Contents of this Page

- [Stemmed Lucene Implementation Details](#)
 - [Algorithm:](#)
 - [Example of use:](#)
 - [Associated JUnits:](#)

Stemmed Lucene Implementation Details

Search with the Lucene query syntax, using stemmed terms. A search for 'trees' will get a hit on 'tree' This requires an extra indexed field when it is enabled in the load.

Algorithm:

The Stemmed Lucene search has the following characteristics:

- This search is case in-sensitive.
- It searches on the stem property value.
- Parsing is done with our custom stemming analyzer. This has the following filters:
 - LowerCaseFilter - for setting to lowercase
 - StopFilter - to remove stop words (the, a, etc.) from the search
 - SnowballFilter - for stemming

Example of use:

The following examples are based on the Automobiles coding scheme.

Example 1:

Search string: Automobiles

Lucene query: stem_propertyValue:automobil

Result: 1 result

- entity code: A0001
- entity description: Automobile

Example 2:

Search string: Automobiled

Lucene query: stem_propertyValue:automobil

Result: 1 result

- entity code: A0001
- entity description: Automobile

Associated JUnits:

Junits can be found here: <https://github.com/lexevs/lexevs/blob/master/lbTest/src/test/java/org/LexGrid/LexBIG/Impl/function/query/lucene/searchAlgorithms/TestStemming.java>