SurroundQueryParser

<!> Solr4.0

- Overview
- Limitations

Overview

The Surround query parser plugin provides support within Solr for the Surround query language.

From surround QueryParser javadocs:

This parser generates queries that make use of position information (Span queries). It provides positional operators (w and n) that accept a numeric distance, as well as boolean operators (and, or, and not, wildcards (* and ?), quoting (with "), and boosting (via $^{\circ}$).

The operators (W, N, AND, OR, NOT) can be expressed lower-cased or upper-cased, and the non-unary operators (everything but NOT) support both infix (a AND b AND c) and prefix AND(a, b, c) notation.

The W and N operators express a positional relationship among their operands. W is ordered, and N is unordered. The distance is 1 by default, meaning the operands are adjacent, or may be provided as a prefix from 2-99. So, for example, 3W(a, b) means that terms a and b must appear within three positions of each other, or in other words, up to two terms may appear between a and b.

See also chapter 9 of Lucene In Action (2nd ed).

This was added to Solr through SOLR-2703

Limitations

Queries are not analyzed in any way, so to get good results, applications must be aware of how a text field is indexed. A lightweight analysis chain such as lowercasing using the LowerCaseFilterFactory followed by the {{EnglishMinimalStemFilterFactory}}to remove plurals is recommended.

By default, the behavior when too many queries are generated is to return an error and partial results. You can override this behavior by adding maxBasicQueries as a local parameter, e.g. {!surround maxBasicQueries=10000}

There is no room for default operator in surround query parser. For example, the query dog cat throws ParseException.

Highlighting support for the surround query parser was fixed in SOLR-3060] (see the [https://issues.apache.org/jira/browse/SOLR-3060)