package com.twitter.search.common.schema;

import org.apache.lucene.analysis.Analyzer;

import org.apache.lucene.analysis.core.WhitespaceTokenizer;

/\*\*

\* The majority of the code is copied from Lucene 3.1 analysis.core.WhitespaceAnalyzer. The only

\* new code is the getPositionIncrementGap()

\*/

public final class SearchWhitespaceAnalyzer extends Analyzer {

@Override

protected TokenStreamComponents createComponents(final String fieldName) {

return new TokenStreamComponents(new WhitespaceTokenizer());

}

/\*\*

\* Make sure that phrase queries do not match across 2 instances of the text field.

\*

\* See the Javadoc for Analyzer.getPositionIncrementGap() for a good explanation of how this

\* method works.

\*/

@Override

public int getPositionIncrementGap(String fieldName) {

// Hard-code "text" here, because we can't depend on EarlybirdFieldConstants.

return "text".equals(fieldName) ? 1 : super.getPositionIncrementGap(fieldName);

}

}