Lucene Introduction

What is Lucene?

Lucene is a high-performance, full-featured, open source text search engine. It is used for many commercially available search solutions.

Originally developed in Java, Lucene has been ported to Microsoft's .NET platform through a project called Lucene.NET.

Lucene's features include:

- Powerful, Accurate and Efficient Search Algorithms
- ranked searching – best results returned first (otherwise known as ranking by relevance)
- many powerful query types: phrase queries, wildcard queries, proximity queries, range queries and more
- fielded searching (e.g., code, description, brand)
- Spelling correction
- relevance "boosting" allowing enhancement of relevance rank based on external factors (i.e. product popularity from web site history, product brand, etc)

Learn more: www.lucene.net

How Lucene is used on the CV ecommerce platform

Many search solutions based on Lucene are run externally from the source web site. In other words, website administrators need to provide a data feed to the external service so that catalogue data can be indexed. Then integration is required to allow switching between the main website and the third party search website.

On the CV ecommerce platform, Lucene.NET has been implemented directly into the application codebase. This provides the following advantages:

- Fast indexing and an up-to-date search index due to the source data being local to the search engine.
- Search queries executed locally on the web site server, providing faster and more reliable searching
- No third party Saas or License fees
- Ability to apply additional business logic, such as filtering search results based on the current user's stock security (i.e. applying item restrictions)

Features

- Ability to define search index fields based on one or more fields from the product or category tables
- Ability to define index field "boost" values, giving higher weight to search terms in particular fields if required.
- Ability to boost product relevance based on product values, such as the number of times an item has been clicked-through, or the number of times an item has been sold
- Define "Boost Words" allowing products containing certain words to be given higher relevance (e.g. promote your brand above others)
- Define Search Query Translations allowing automatic or "did you mean?" (manual) replacement of user's search queries (e.g. user searched for "rubber", system responds with "did you mean 'eraser'?")
- Spelling correction database generated from the product text which can be used for "did you mean" search suggestions.
- Optional use of word stemming for indexing and search queries - (i.e. laminate/laminating/laminator/laminated are all equivalent when stemmed)
- Define Search Index Translations allowing items to be found based on alternate text.
- Define "Trim Words" allowing consistent treatment of suffixes (e.g allowing "100mm" and "100 mm" to be used interchangeably)
- Search Tools allowing direct query on the Search Index including display of boost values, relevance scores and score explanations to assist with debugging search query results
- Configurable settings allowing features of the search engine to be turned on or off.
- Catalogue Browse based on relevance - listing products based on their score (including boost values)

Related Resources

- Product Search with Lucene Search Text Engine
- Lucene Introduction
- Lucene Query Translation Maintenance
- Predictive Search
- Build or Rebuild the Lucene Search Index