

Report 1: Lucene

Description Lucene

In this section we will discuss the functionality of Lucene. Most of the information below is based on the documentation of Lucene. ¹

Document Analysis and Indexing

Lucene has several different analyses for indexing content in different ways and languages. Some examples that were seen in class are **MinHash** and **Shingle**. It is also possible to create your own Analyzer and Stemmer.

Indexing in Lucene is handled by an `IndexWriter` that stores terms and document ids in an inverted index using Skip-Lists instead of B-Trees. ²

Query Processing

Document Search and Retrieval

Lucene vs. Solr

Solr is a web-based ³

1. https://lucene.apache.org/core/8_9_0/core/index.html ↩
2. <https://stackoverflow.com/questions/2602253/how-does-lucene-index-documents> ↩
3. <http://www.lucenetutorial.com/lucene-vs-solr.html> ↩