**Full Text Search**

* **Definition:** **Full-text search refers to techniques for searching a single computer-stored document or a collection in a full-text database**. Full text search is a technique, which allows conducting search through documents and databases not only by a title, but also by content. Unlike metadata search methods, which analyze only the description of the document, full text search goes through all the words in the document, showing information that is more relevant or the exact information that was requested.

For example, with a Full-Text query you can search for the word “werewolf” in the title of any of our movies or you want to search Dent in any PDF or any big storage date then this will ignore the case sensitive and give you result of like DENT, dent, dEnt like vise.

* **Why do we need it?**

Full text search may be useful when one needs to search for:

* a name of the person in a list or a database.
* a word or a phrase in a document.
* a web page on the internet.
* products in an online store, etc.
* a regular expression.
* **how Full text work:** There are different ways of realization of full text search. To make it easier, let’s divide methods into two groups:
  1. **String searching algorithms:** To find a substring matching of a pattern (needed expression) in a text, we’ll go through the document(s) until the match is found or the text is finished. In fact, most of these methods are rather slow.
* simple text searching.
* Rabin-Karp algorithm.
* Knuth-Morris-Pratt algorithm.
* Boyer-Moore (-Horspool) algorithm.
* approximate matching.
* a regular expression.
  1. **Indexed search**. Create index of every word and then when ever we search any word it finds 1st index of that word and then using that index it will search the word and give the result
* **Disadvantage of the FULL text:**

synonym problem

use of abbreviations

The existence of dialects also complicates the search. For example, users might not meet results “colour”, querying “color”, or searching for “trainer” find shoes instead of a mentor.

Another problem is homonyms. These are words, that being spelled in the same way mean completely different things.

There are many different platforms which are using full text search technique are like Elastic search, solr search, Lucene search.

* **Elastic Search**: Elasticsearch is a highly scalable open-source full-text search and analytics engine. It allows you to store, search, and analyze big volumes of data quickly and in near real time.

Elasticsearch uses a data structure called an inverted index, which is designed to allow very fast full-text searches. An inverted index lists every unique word that appears in any document and identifies all the documents each word occurs in.

* **Solr search:** Solr is a leading open-source enterprise search platform from the Apache Software Foundation's Lucene project. With its flexibility, scalability, and cost effectiveness, it’s also often used as a document-based NoSQL database with transactional support that can be used for storage purposes and even a key-value store.

Solr works by gathering, storing, and indexing documents from different sources and making them searchable in near real-time. It follows a 3-step process that involves indexing, querying, and finally, ranking the results – all in near real-time, even though it can work with huge volumes of data.

* **Lucene Search:** Apache Lucene is a high-performance and full-featured text search engine library written entirely in Java. Lucene creates a big index. The index contains word id, number of docs where the word is present, and the position of the word in those documents. So when you give a single word query it just searches the index (O (1) time complexity)

**Links:**

* <https://en.wikipedia.org/>
* <https://www.sciencedirect.com/topics/computer-science/full-text-search>
* <https://blog.issart.com/full-text-search-how-it-works/>
* <https://www.elastic.co/guide/en/elasticsearch/reference/current/documents-indices.html>
* https://sematext.com/guides/solr/