

UI for Qualitative Analysis with Elasticsearch

Student

Nicolas Karrer

Introduction: Elasticsearch is an open-source search engine based on the Lucene library. It is used for full-text search and data analysis and can handle a large volume of data. However, when working with Elasticsearch, it is sometimes necessary to do a qualitative analysis of the search and its results to improve the delivery of found content to the user. Even though Elasticsearch does a good job to find and rank documents based on text search, it is not always transparent why certain documents are matched or ranked the way they are. Understanding the inner workings of the Elasticsearch can be challenging, even for technical users. One way to gain insight into the search results is by using the Elasticsearch's "Explain API", which provides an in-depth analysis of how the content of a document is matched and scored based on the configuration and the executed query. However, using this API requires knowledge of sending HTTP Requests to a server as well as writing, reading, and interpreting JSON content, which can be challenging for non-technical users. For developers, configuring indexes and supporting content managers with analytics and comparing documents, without having a tool to support them, can be a tedious job.

Objective: The objective of this project is to develop a user interface to provide access to the explain API for two target groups: content managers and developers. For content managers, it should be a help to understand the ranking of documents in a search result, so they can improve the content if necessary. The user interface should save time for developers to debug and interpret index definitions, search queries as well as interpreting the search results.

Result: The resulting web application is called ElastiQA. It can connect to multiple Elasticsearch installations and allows users to create and save analytics for search queries. It will make using of the Explain API more accessible to non-technical users by providing a user-friendly interface to access the information provided by the API. The interface provides additional information for technical users. In addition, the user interface can store and re-run analyses of queries, making it useful for regression testing of key search results. The functionality of the Postman software served as an inspiration for ElastiQA, but it is tailored to a specific use case. Overall, the ElastiQA interface will help content managers to understand the ranking of documents in a search result and improve the content where necessary. It will also save developers time in debugging and interpreting search queries and the search results.

Advisor
Prof. Dr. Markus Stolze

Subject Area
Software and Systems

The result of an Explain API call as displayed in ElastiQA Own presentation

Field	Term	Score
search_boost_simple_de	fine	512.6046
search_boost_simple_de	ans	327.5348
search_boost_simple_de	fi	99.35726
search_boost_simple_de	art	42.75883
search_boost_simple_de	ans	44.688919
search_boost_simple_de	fi	24.34072
search_boost_simple_de	art	16.75726
search_boost_simple_de	fine ans	49.68825

The result of an Explain API call as displayed in ElastiQA Own presentation

Field	Term	Score
search_boost_simple_de	fine	512.6046
search_boost_simple_de	ans	327.5348
search_boost_simple_de	fi	99.35726
search_boost_simple_de	art	42.75883

Comparison of two results of an Explain API call as displayed in ElastiQA Own presentation

Field	Term	Score
search_boost_simple_de	fine	512.6046
search_boost_simple_de	ans	327.5348
search_boost_simple_de	fi	99.35726
search_boost_simple_de	art	42.75883