A Study of Revolutionizing Real Estate: Search Engine using Machine Learning (ML), Artificial Intelligence (AI) and Data Analytics in ElasticSearch Algorithm

Nisha Sudhir Patil[1], Kalyani K. Shinde[2], Abhishek R. Kulkarni[3], Amol M. Jagtap[4], Ajit Mali[5]

[1]Dept. of Computer Science & Engineering, Rajarambapu Institute of Technology, Islampur, India.
[2]Dept. of Computer Science & Engineering,Rajarambapu Institute of Technology, Islampur, India.
[3]Dept. of Computer Science & Engineering, Rajarambapu Institute of Technology, Islampur, India.
[4]Professor, Dept. of Computer Science & Engineering, Rajarambapu Institute of Technology, Islampur, India.
[5]Professor, Dept. of Computer Science & Engineering, Rajarambapu Institute of Technology, Islampur, India.

---***---

Abstract - *This paper deals with the study of revolutionizing real estate search engines using machine learning (ML), Artificial Intelligence (AI) and Data analytics. The efficiency of searching a property can be enhanced using the best algorithm. The aim is to decrease time complexity and space complexity. Since decrease in the time and space complexity is inversely proportional to accuracy and speed of searching . Fast searching and accuracy of results obtained will be considered while designing the algorithm.*

Key Words: **Real Estate, Machine Learning, Artificial Intelligence, Data learning algorithms, Optimization, SEO, etc**.

1. INTRODUCTION

Real Estate is a combination of land and structure built upon it. Online real estate market is completely obsessed with authentication, searching, sorting and data learning algorithms. The fluctuations occur while searching properties are due to poor data learning and data mining algorithms. Machine learning and Artificial Intelligence are used to increase modularity, legitimacy, and accuracy to searching for data. Machine Learning is technology which automatically learns from data and is ready to create predictions and supported information. There are many algorithms for searching and sorting to improve the efficiency of searching properties using ML and AI from which we have discussed the *'ElasticSearch algorithm'* which gives us the best results. The working of elastic Search algorithm is discussed that acquire customer's needs and deliver the best results they are looking for.

Elasticsearch algorithm helps to detect duplicate contents and has intelligence to detect patterns. This algorithm is a data learning algorithm delivering personalized experience to authenticated and reliable users of the system. The problems for searching property can be solved with the solution by transforming the start state to goal state and decreasing time & space complexity using Machine learning. Elasticsearch may be distributed. It is an

analytics engine that is able to solve a growing number of use cases while searching .Since it is a centralized part of elastic stack, it centrally stores your data. Hence, we can find the expected and unexpected data stored from the database.

Here, we have done study on various aspects of elasticSearch Algorithm which is considered one of the best algorithms for searching data stored.

2. LITERATURE REVIEW

Subhani shaik, Nallamothu Naga Malleswara Rao explained development in searching and analysis of the document using elasticsearch algorithm. Elasticsearch has the ability to divide the index into multiple pieces that are shards. When new data is stored , firstly it is indexed and then elasticsearch server predicts which of the shards are responsible for the data. Explained creation of shards in nodes, creation of index , creation of directory, retrieving a document, searching a document, Search with query DSL, full text search, phrase search, using stopword list , updating stopwords and performance of stopword.Analyzing a data using elasticsearch. Explained process of tokenization and normalization with details.[4]

Darshita Kalyani and Dr. Devarshi Mehta explained working of searching techniques and how data is indexed using the elasticsearch big data technology algorithm. Represented concept of inverted index for searching used in elasticsearch algorithm. Explanation of elasticsearch, its working using stopword list concept. And it can improve the search time complexity. The process of generating tokens and their normalization is called analysis. Analysis makes the search process faster. To obtain greater efficient relevancy of data documents, scoring function should be improved [5]

R. Vidhya and G. Vadivu explained the implementation of elasticsearch servers to create search engines. Also datasets that should be prepared to prepare the database. He explained implementation of search engine with search

page, search result and gives idea of future scope and future enhancement that can be done over a model.[6]

Paresh Kamble explained an elasticsearch database with a horizontal and more scalable, distributed database which is built on Apache's Lucene that helps to deliver a good search experience. He explained data and document maintenance using appropriate data structure and concept & advantages of Clustering and Indexing during the data storage, querying and retrieval. Working of JSON over HTTP and HTTPS protocols. All fields are indexed automatically while storing, and the indices are used in each single query, to display results with highest speed.[7]

C. Bhadane, H. A. Mody, D. U. Shah, P. R. Sheth explained elasticSearch is a schema less indexing technique. Hence hashing or indexing is similar to graphs in it. ElasticSearch achieve fast search responses because, instead of searching the text directly, it searches an index and direct to text .[8]

Pragya Gupta, Sreeja Nair explained how elasticsearch manages data and makes it easily accessible. It is released as open source with terms of the Apache License. Elasticsearch is the most popular search engine. Elasticsearch is having many advantages like speed, authentication, scalability, and hardware efficiency. Elasticsearch works as a tool to write a query in human words and display results in humanreadable format. Elasticsearch is a standalone database server. It is written in Java and using HTTPS/JSON protocol. It optimizes the data according to language based searches. The data is stored in a sophisticated format. Elasticsearch is very convenient, supporting clustering and leader selection out of the box.[9]

3. METHODOLOGY

The basic steps to search properties and deliver more robust & accurate results with greater efficiency is to differentiate the properties into Residential Real Estate (RRE) and commercial Real Estate units (CRE). These differentiated estates/ properties are results of filtered and searched keywords. Algorithms understand keywords that are given as input to search engines and then display best results. Search algorithms help determine the ranking of a web page also. Each search engine uses a specific set of rules that determines if a web page is real or spam. The results determine a site's ranking on the search engine results page. There are many search engines that optimize the searching efforts: amazon, google, yahoo, ask, bing, etc.

The best searching algorithm in 2019-2020 is an elasticsearch algorithm. The pro's of elasticsearch are:

1)   Gives scalable search solution

2)   Easy to retrieve indices after server has been crashed
3)   Uses JSON and Java API's
4)   Auto Indexing of JSON documents

The algorithm is totally based in Apache Lucene which is an open and free source library and originally written completely in Java by Doug Cutting in 1999 . It is a cross platform operating system provided for indexing and search. Basic model of elasticsearch:
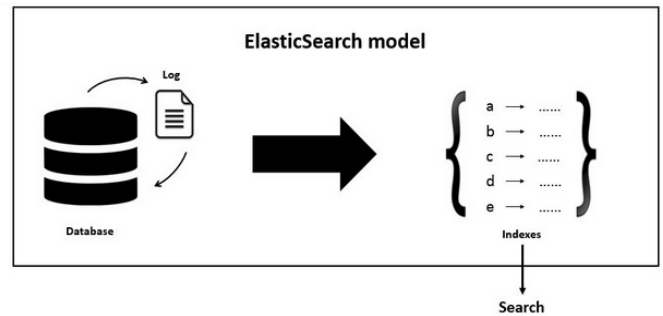
Fig -1: Elasticsearch Model [5]

Machine learning, artificial intelligence and data analytics are technologies that help to improve searching efficiency of an elasticSearch algorithm.

We will study the workings of the elasticSearch algorithm.

We can use Rank plugin to elasticsearch algorithms to increase ranking. To implement Machine Learning in elasticsearch for ranking we need to:

1.   **Measure what users deem relevant** through analytics to build a judgment list grading documents as exactly relevant, moderately relevant, or not relevant for queries.[1]

2.   **Hypothesize might help us to predict relevance,** such as the TF*IDF of specific field matches, recency, personalization of account for user for better results , etc.[1]

3.   **Train a model** that can accurately map features to a relevance score.[1]

4.   **Deploy the model** to your search infrastructure, using it to rank search results in production.[1]

After we have done training the data, we can issue the searches. Following is simplest query that searching for 'rambo'

```
{ "query": { "match": { "_all": "rambo" } }, "rescore": {
"window_size": 20, "query": { "rescore_query": { "ltr":
{ "model": { "stored": "test" }, "features": [{ "match": {
"title": "rambo" } }, { "multi_match": { "query":
"rambo", "type": "cross_fields", "tie_breaker": 1.0,
"fields": ["overview", "genres.name", "title", "tagline",
"belongs_to_collection.name",           "cast.name",
"directors.name"] } }] } } } } }
```

Fig -3: Arising query for Searching

The following flowchart help us to understand how elasticsearch algorithm fetches the accurate or approximate result and display it:

Chart -1: Searching Flowchart for ElasticSearch[7]

Elasticsearch is just a document-oriented approach while manipulating data stored in JSON format.

Data can be organized and stored based on the index and type. There can be multiple indexes and types. You can think of the index as a database in a regular relational database and type as tables. So, we can sort the data using indexes and find more relevant data searching techniques. But it can be done after the study between relational databases and elasticsearch.

Moreover we should know how data can be imported to the databases and indexing can be done while importing.

Following graph gives a straight forward graphical representation between the relational database and elasticsearch for how the data being imported to the databases.

Chart -1: Comparison between importing data to Mysql and Elasticsearch

While implementing, there can be many ways for implementation. But while highlighting two important implementation techniques, we can list out:

1.   Run it in on localhost

2.   Services that run only elasticsearch separately from your instance, for example on Amazon using AWS elasticsearch.

For implementing elasticsearch algorithms one can use Laravel.

Importing data depends on each machine, but just for comparison, this was done on a local machine with medium grade hardware. We can see less time is required for importing the data to elasticSearch than MySQL.

MATCHING RESULTS	TIME ELAPSED (MS)	AMOUNT OF USERS
28468	50	1M
13889	46	500K
7093	35	250K
2653	33	100K
2114	26	75K
1546	25	50K
721	25	25K
278	23	10K
186	21	5K
36	19	1K
13	11	500
2	10	100

**Table -1:** Behaviour of ElasticSearch depending on amount of users

Above table gives the behaviour of elasticsearch depending on the amount of users. It gives the stability in time elapsed while performing a simple query.

MATCHING RESULTS	TIME ELAPSED (MS)	AMOUNT OF USERS
13211	2696	1M
7469	1153	500K
4258	521	250K
1739	323	100K
1382	286	75K
710	199	50K
194	151	25K
87	92	10K
42	78	5K
17	59	1K
8	52	500
5	45	100

Table -2: Behaviour of Laravel depending on amount of users

Now from Table -2, we can better see time comparison between elasticsearch and Laravel Eloquent. ElasticSearch is having constancy in time elapsed while performing simple queries.
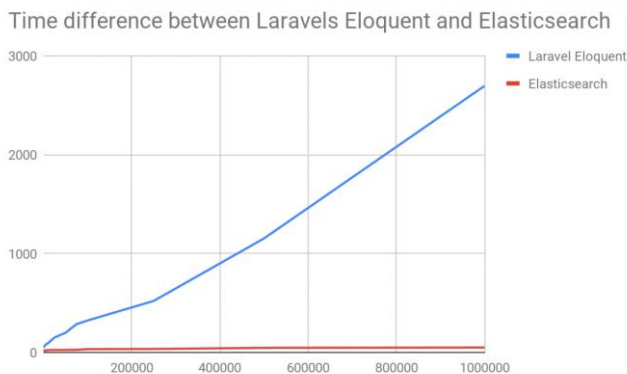
Chart -3: Difference between Laravel Eloquent and ElasticSearch

As we can see from the performance chart -3, elasticsearch shines the best at handling big data where on the lower quantity of data.

Next comes the elasticsearch as service, it needs space, depending on the size of your data, handling big data will cause higher usage of memory and CPU power.

4. RESULT AND DISCUSSION

From the tables above we can see that Elasticsearch has no problem with searching through big data and it's included fuzziness retrieves more similar data with no significant impact on elapsed time. Hence, Elasticsearch can be used as a property search engine in the real estate market.

Conventional MySQL and Laravel Eloquent don't promptly give you the way to do that. At last, Elasticsearch can be utilized as a Data distribution center, where you have archives with a wide range of qualities and non-unsurprising outlines. Since Elasticsearch search is blueprint less, it won't make any difference that you store different reports there, you will in any case have the option to look through them effectively and rapidly.

5. CONCLUSION

After studying the way elasticSearch works and the different Optimization API it provides, we can conclude that it is the reliable and the fastest to work with. We also now have clear advantages over Apache Lucene. Thus, for this project, elasticSearch is the most suitable search technique in real estate for searching properties.

REFERENCES

[1]   Doung Turnbull, "Machine Learning for Smarter Search With Elasticsearch" (https://dzone.com/articles/learning-to-rank-in-elasticsearch-machine-learning) Last Edited | June 2020

[2]   Stanislas Chaillou, Daniel Fink and Pamella Gonçalves "Urban Tech on the Rise: Machine Learning Disrupts the Real Estate Industry" Field Actions Science Reports The journal of field actions. Special Issue 17 | Dec 2017 Artificial Intelligence and Robotics in the City

[3]   Subhani shaik, NalamothuNaga Malleswara Rao "A Conceptual Review of Elastic ElasticSearch – Survey Paper" International Journal for Research in Applied Science & Engineering Technology (IJRASET) SSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887Volume 5 Issue XI November 2017

[4]   Subhani shaik, Nallamothu Naga Malleswara Rao "Enhancement of Searching and analyzing the document using ElasticSearch" International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 04 Issue: 11 | Nov -2017 p-ISSN: 2395-0072

[5]   Darshita Kalyani, Dr. Devarshi Mehta "Paper on Searching and Indexing Using Elasticsearch" International Journal Of Engineering And Computer Science ISSN:2319-7242 Volume 6 Issue 6 June 2017, Page No. 21824-21829 Index Copernicus value (2015): 58.10 DOI: 10.18535/ijecs/v6i6.45

[6]   R. Vidhya and G. Vadivu "Research Document Search using ElasticSearch" Indian Journal of Science and Technology,Vol 9(37), September 2016, ISSN: 0974-6846

[7]   Paresh Kamble "Study on ElasticSearch for database" International Multidisciplinary e-journal ISSN: 2277-4262, SJIF: 4.081 June 2015

[8]   C. Bhadane, H. A. Mody, D. U. Shah, P. R. Sheth "Use of ElasticSearch for Intelligent Algorithms to Ease the Healthcare Industry" International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-3 Issue-6, January 2014

[9]   Pragya Gupta, Sreeja Nair "Survey Paper on ElasticSearch" International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Index Copernicus Value (2013): 6.14 | Impact Factor (2014): 5.611