Code Explanation, Outputs And Screenshots :

1. Web Crawler :

Code Explanation :

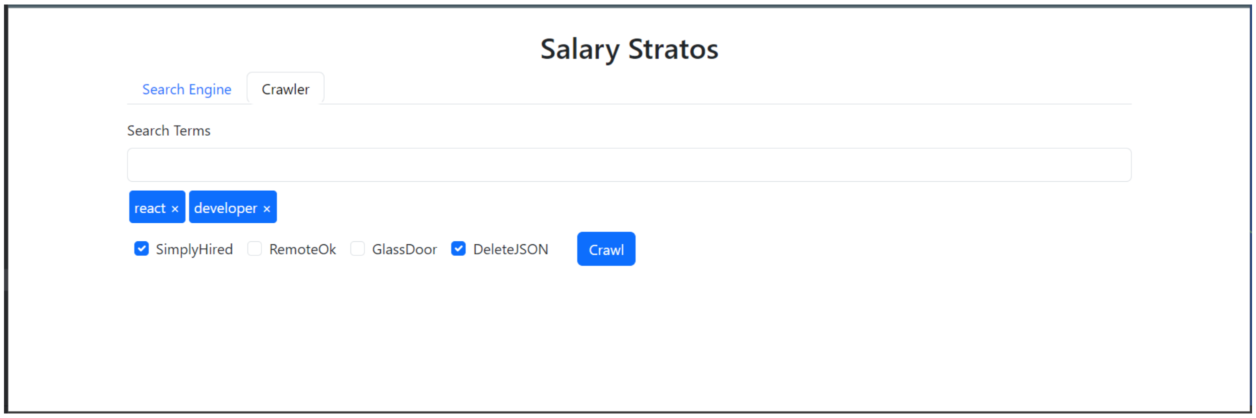
Scrapping in Simply hired and GlassDoor. The crawlWebPage method is part of a web scraping application that extracts job-related information from a website based on specified search terms. It utilizes Selenium WebDriver with a browser to navigate through the website, search for jobs corresponding to each term, and scrape relevant details from individual job pages. The method iterates through the provided search terms, accesses the search results page for each term, extracts job links, and continues to subsequent result pages until at least min job links are collected. It then proceeds to visit each job page, extracts job data using the scrapeJobData method, performs data validation, and stores the validated job information in a collection. The scrapJobLinks method uses Jsoup to parse the HTML source of the search results page, extracts job links, and adds them to a queue. The uniqueJobs set ensures that duplicate job entries are not included in the final collection. The extracted job data is then saved and appended to a JSON file using the saveAndAppendToJson method of the ScraperBot class. The WebDriver is closed at the end of the process.
The only difference with scrapping Remote Ok is that the data we need is already existing on the search page and we don't need to open each and every job to get all the data.

Output:

Users can enter the search terms, which website they want to crawl and whether they want to delete the json or append to the same database.json. The driver opens the chrome browser and starts crawling the website with the search terms to get the page source.

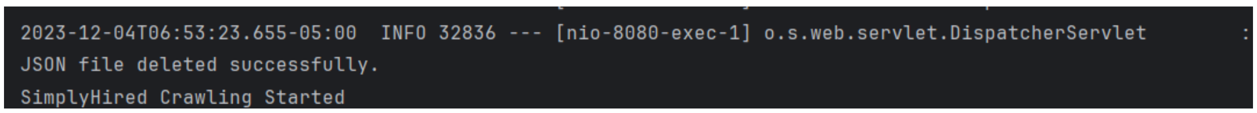
Output 1: User wants to delete the database.json and crawl only Simply Hired for search teams "react", "developer"

User clicks on crawl



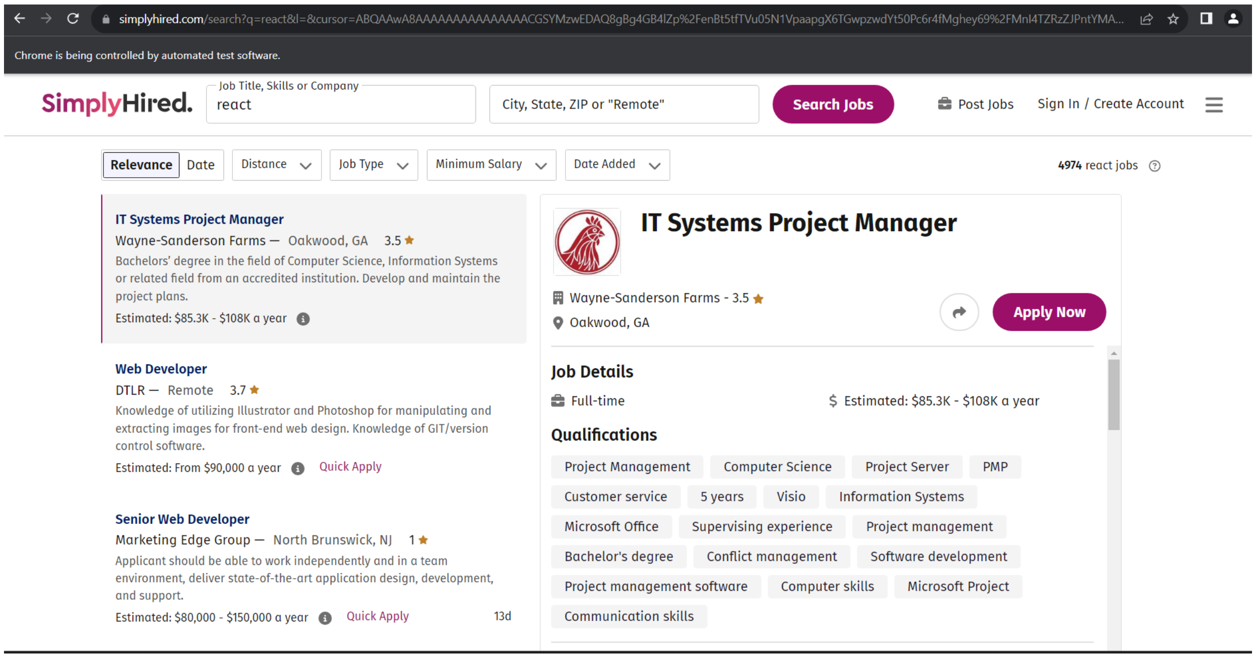
[✓] SimplyHired [ ] RemoteOk [ ] GlassDoor [✓] DeleteJSON

File deleted successfully



```
2023-12-04T06:53:23.655-05:00  INFO 32836 --- [nio-8080-exec-1] o.s.web.servlet.DispatcherServlet        :
JSON file deleted successfully.
SimplyHired Crawling Started
```

It searches for react and gets all the links on that page



simplyhired.com/search?q=react&l=&cursor=ABQAAwA8AAAAAAAAAAAAAAAACGSYMzwEDAQ8gBg4GB4lZp%2FenBt5tfTVu05N1VpaapgX6TGwpzwdYt50Pc6r4fMghey69%2FMnl4TZRzZJPntYMA...

Chrome is being controlled by automated test software.

SimplyHired. | Job Title, Skills or Company: react | City, State, ZIP or "Remote" | Search Jobs | Post Jobs | Sign In / Create Account

Relevance | Date | Distance | Job Type | Minimum Salary | Date Added — 4974 react jobs

IT Systems Project Manager
Wayne-Sanderson Farms — Oakwood, GA 3.5 ★
Bachelors' degree in the field of Computer Science, Information Systems or related field from an accredited institution. Develop and maintain the project plans.
Estimated: $85.3K - $108K a year

Web Developer
DTLR — Remote 3.7 ★
Knowledge of utilizing Illustrator and Photoshop for manipulating and extracting images for front-end web design. Knowledge of GIT/version control software.
Estimated: From $90,000 a year Quick Apply

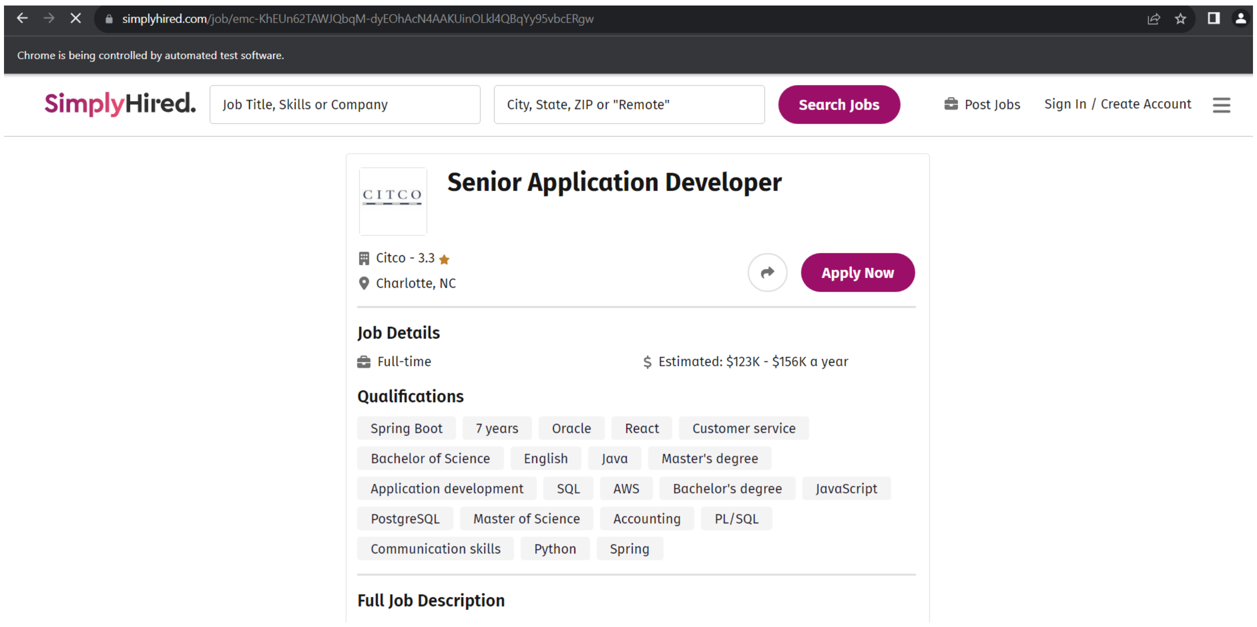
Senior Web Developer
Marketing Edge Group — North Brunswick, NJ 1 ★
Applicant should be able to work independently and in a team environment, deliver state-of-the-art application design, development, and support.
Estimated: $80,000 - $150,000 a year Quick Apply 13d

IT Systems Project Manager
Wayne-Sanderson Farms - 3.5 ★
Oakwood, GA
Apply Now

Job Details
Full-time — Estimated: $85.3K - $108K a year

Qualifications
Project Management | Computer Science | Project Server | PMP | Customer service | 5 years | Visio | Information Systems | Microsoft Office | Supervising experience | Project management | Bachelor's degree | Conflict management | Software development | Project management software | Computer skills | Microsoft Project | Communication skills

It opens each link and gets all the job data from that link



simplyhired.com/job/emc-KhEUn62TAWJQbqM-dyEOhAcN4AAKUinOLkl4QBqYy95vbcERgw

Chrome is being controlled by automated test software.

SimplyHired. | Job Title, Skills or Company | City, State, ZIP or "Remote" | Search Jobs | Post Jobs | Sign In / Create Account

Senior Application Developer
Citco - 3.3 ★
Charlotte, NC
Apply Now

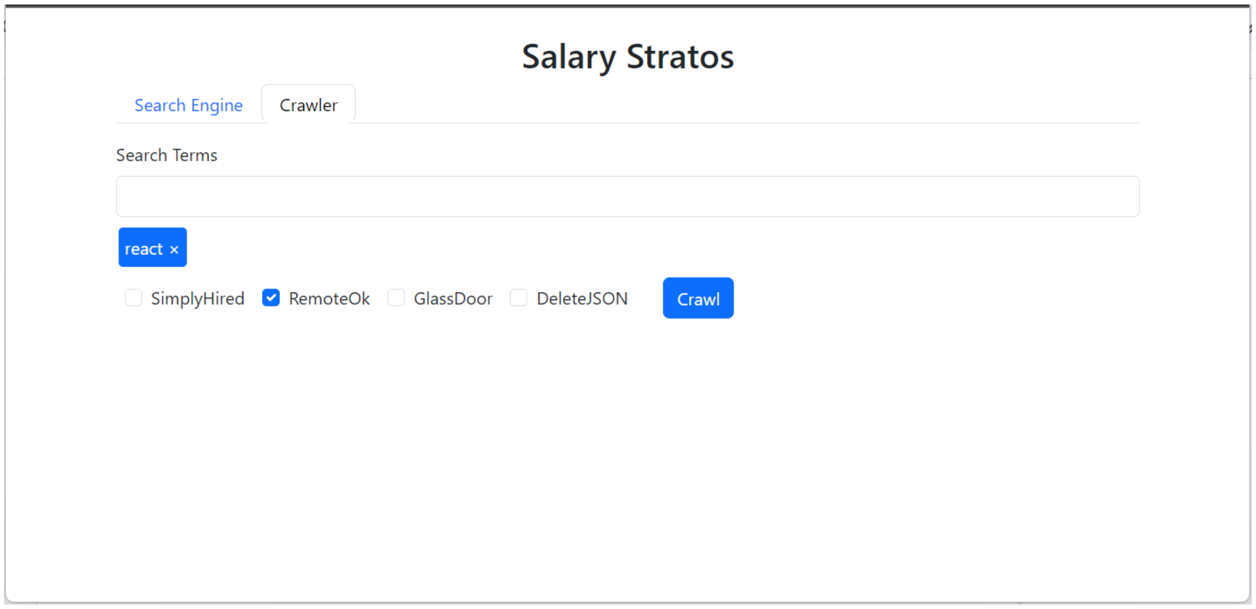
Job Details
Full-time — Estimated: $123K - $156K a year

Qualifications
Spring Boot | 7 years | Oracle | React | Customer service | Bachelor of Science | English | Java | Master's degree | Application development | SQL | AWS | Bachelor's degree | JavaScript | PostgreSQL | Master of Science | Accounting | PL/SQL | Communication skills | Python | Spring

Full Job Description

<u>Output 2: User wants to crawl only Remote Ok for search teams “react”, “developer” and wants to append to the database.json</u>

User clicks on crawl



Salary Stratos

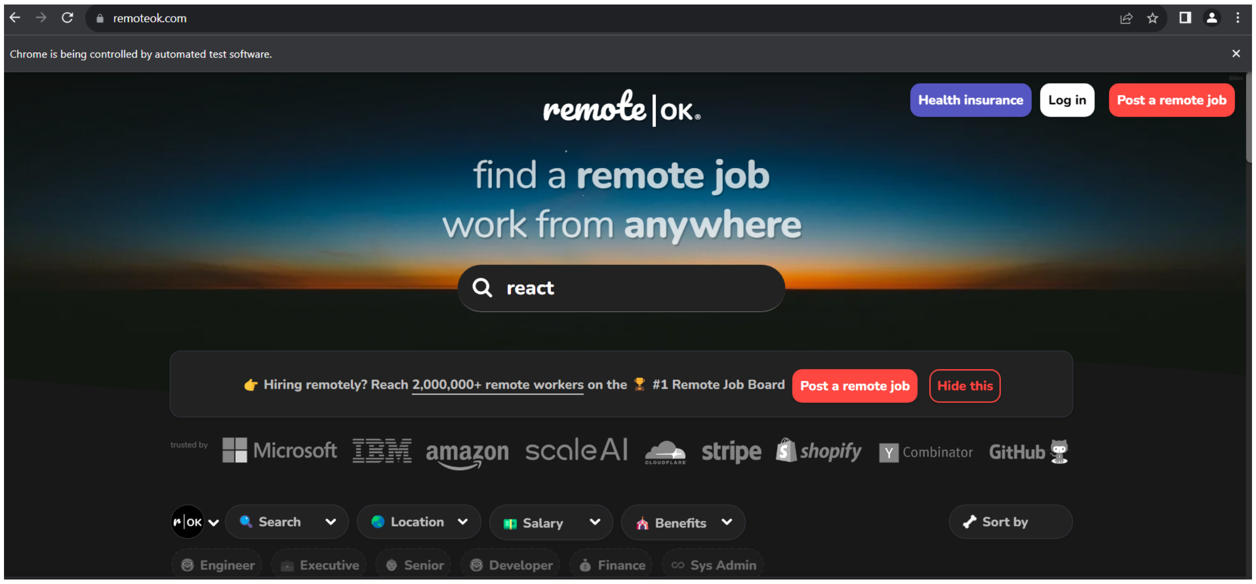
Search Engine | Crawler

Search Terms

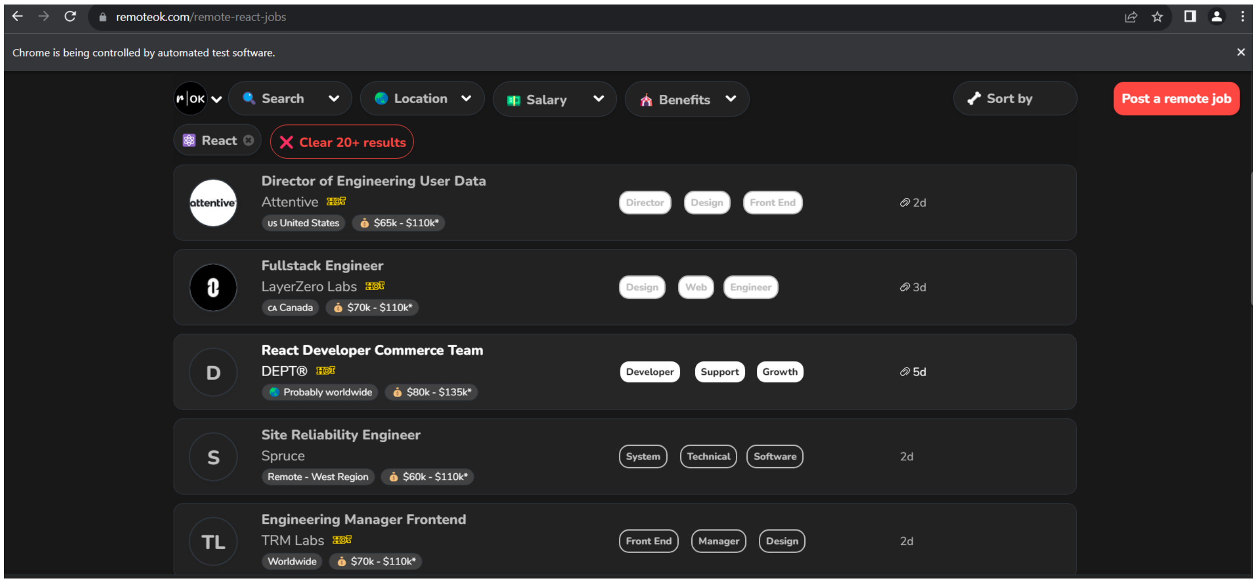
react ×

SimplyHired [✓ RemoteOk] GlassDoor DeleteJSON

Search for “react”



It gets all the jobs on this page



remoteok.com/remote-react-jobs

Chrome is being controlled by automated test software.

Search | Location | Salary | Benefits | Sort by | Post a remote job

React | Clear 20+ results

Director of Engineering User Data
Attentive
us United States | $65k - $110k*
Director | Design | Front End
2d

Fullstack Engineer
LayerZero Labs
ca Canada | $70k - $110k*
Design | Web | Engineer
3d

React Developer Commerce Team
DEPT®
Probably worldwide | $80k - $135k*
Developer | Support | Growth
5d

Site Reliability Engineer
Spruce
Remote - West Region | $60k - $110k*
System | Technical | Software
2d

Engineering Manager Frontend
TRM Labs
Worldwide | $70k - $110k*
Front End | Manager | Design
2d

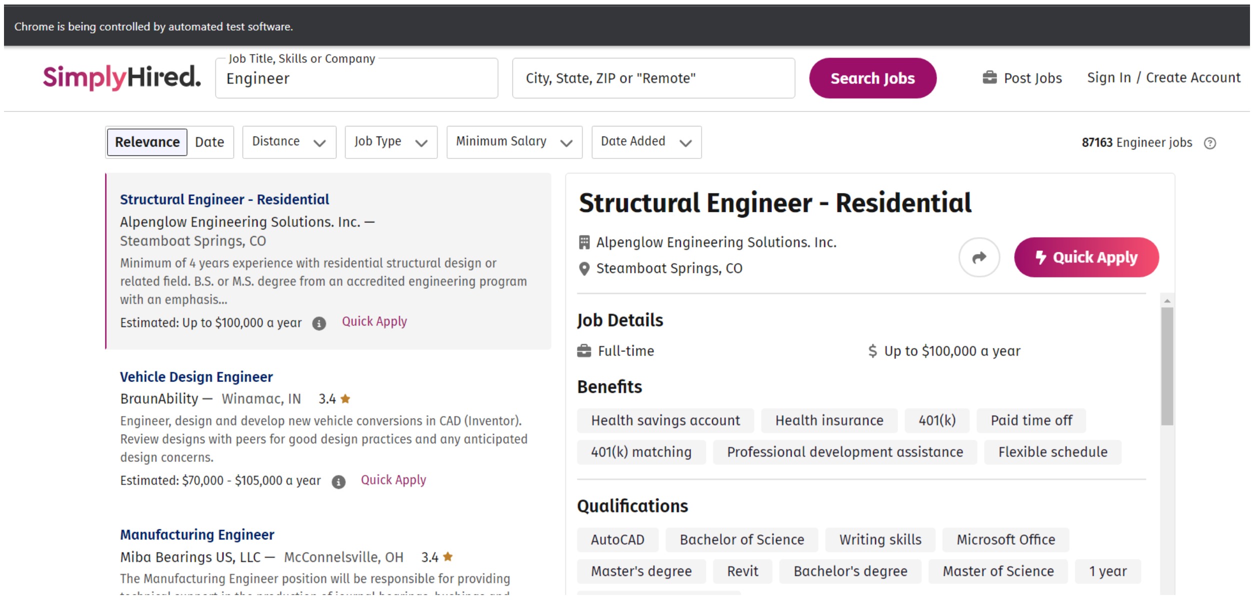
<u>Output 3 : If file does not exist when starting the application it will crawl the data for default search terms and create the database.json file.</u>

```
String[] searchTermsList = new String[]{
          "Engineer", "Exec", "Senior", "Developer", "Finance", "Sys
Admin", "JavaScript", "Backend", "Golang", "Cloud", "Front End"
};
```

File was not found

```
2023-12-04T07:04:59.747-05:00  INFO 41512 --- [  restartedMain] c.a.S.S
File does not exist re-crawl data for default search terms
Bot loaded for first time
```

Re-crawl data for the default search terms



Chrome is being controlled by automated test software.

SimplyHired. | Job Title, Skills or Company: Engineer | City, State, ZIP or "Remote" | Search Jobs | Post Jobs | Sign In / Create Account

Relevance | Date | Distance | Job Type | Minimum Salary | Date Added | 87163 Engineer jobs

Structural Engineer - Residential
Alpenglow Engineering Solutions. Inc. — Steamboat Springs, CO
Minimum of 4 years experience with residential structural design or related field. B.S. or M.S. degree from an accredited engineering program with an emphasis...
Estimated: Up to $100,000 a year | Quick Apply

Vehicle Design Engineer
BraunAbility — Winamac, IN 3.4 ★
Engineer, design and develop new vehicle conversions in CAD (Inventor). Review designs with peers for good design practices and any anticipated design concerns.
Estimated: $70,000 - $105,000 a year | Quick Apply

Manufacturing Engineer
Miba Bearings US, LLC — McConnelsville, OH 3.4 ★
The Manufacturing Engineer position will be responsible for providing

Structural Engineer - Residential
Alpenglow Engineering Solutions. Inc.
Steamboat Springs, CO
Quick Apply

Job Details
Full-time | Up to $100,000 a year

Benefits
Health savings account | Health insurance | 401(k) | Paid time off | 401(k) matching | Professional development assistance | Flexible schedule

Qualifications
AutoCAD | Bachelor of Science | Writing skills | Microsoft Office | Master's degree | Revit | Bachelor's degree | Master of Science | 1 year

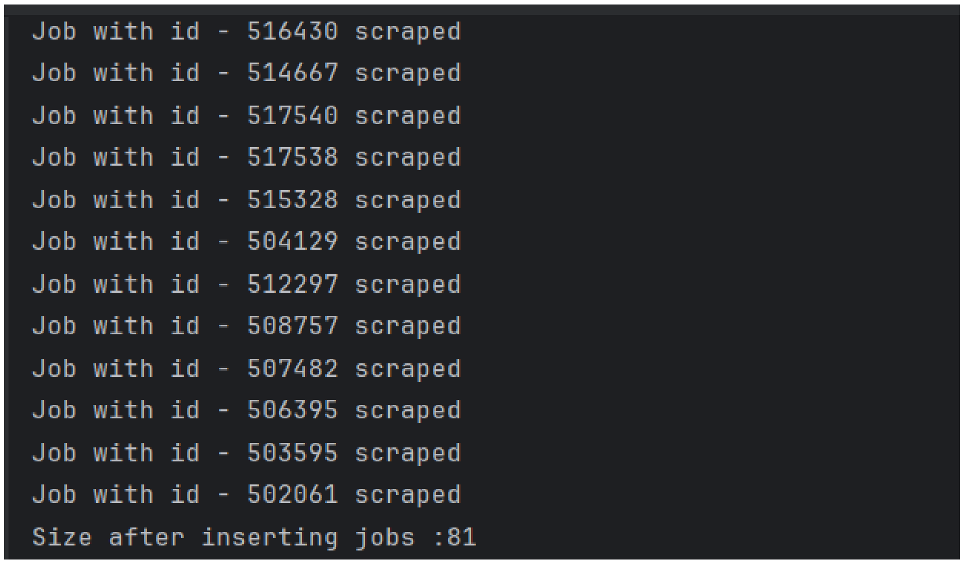
2. HTML Parser :

Code Explanation :

The scrapeJobData method is part of a web scraping application designed to extract job-related information from a webpage. It uses the Jsoup library to parse the HTML source of the page and retrieve details such as job ID, title, company name, location, website name, salary, and job description. The method applies regular expressions to extract numerical salary information and converts it into minimum and maximum salary values. Various regular expression patterns are used to handle different salary formats, including daily, weekly, monthly, yearly, and hourly rates. The extracted data is then used to create a Job object, which encapsulates the job-related information, and this object is returned by the method. Overall, the method provides a structured way to extract and organize job data from the HTML source of a job posting webpage.

Output 1 :

The jobs with the below jobIds got scrapped



```
Job with id - 516430 scraped
Job with id - 514667 scraped
Job with id - 517540 scraped
Job with id - 517538 scraped
Job with id - 515328 scraped
Job with id - 504129 scraped
Job with id - 512297 scraped
Job with id - 508757 scraped
Job with id - 507482 scraped
Job with id - 506395 scraped
Job with id - 503595 scraped
Job with id - 502061 scraped
Size after inserting jobs :81
```

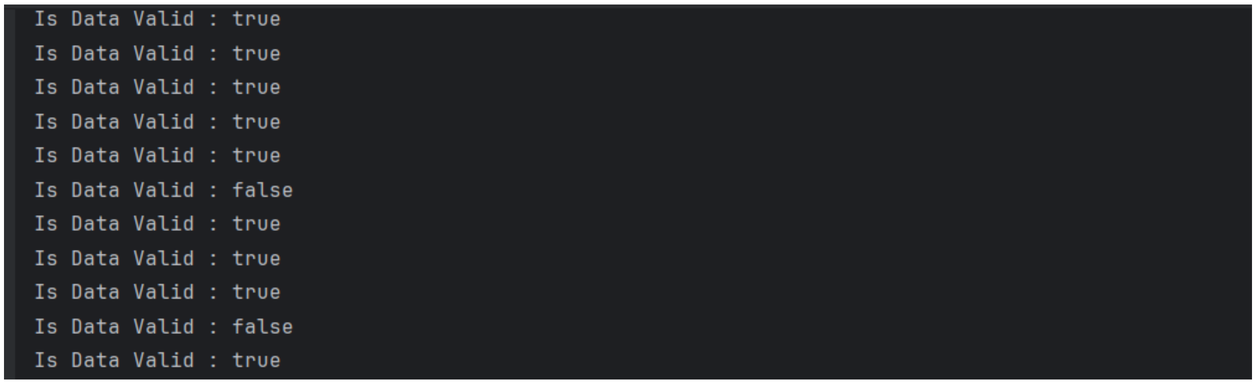
3. Data validation using regular expressions :

Code Explanation :

The code implements data validation for job-related information extracted during web scraping. It defines a set of regular expressions for validating different fields, such as job title, company name, website link, salary, location, and job description. The validateField method is a generic function that checks whether a given field matches the specified regular expression. The validateDataForOneObject method utilizes this generic validator for each field of a Job object, setting validation results for individual fields and an overall validation status. If any field fails validation, the allValid flag is set to false. During job data processing, only objects with valid information, as determined by the data validation process, are added to the jobsCollection. This ensures that only well-formed and compliant job data is included in the final collection, enhancing the overall data quality.

Output 1 :

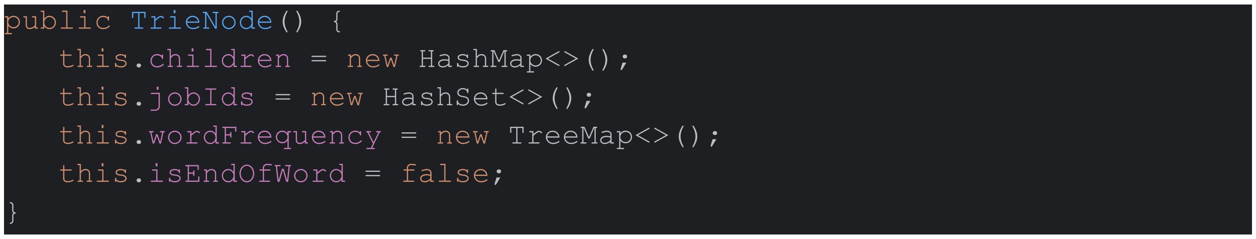
Validate data and only add to the database.json if the data is valid



```
Is Data Valid : true
Is Data Valid : true
Is Data Valid : true
Is Data Valid : true
Is Data Valid : true
Is Data Valid : false
Is Data Valid : true
Is Data Valid : true
Is Data Valid : true
Is Data Valid : false
Is Data Valid : true
```

4. Inverted Indexing and Frequency count :

Code Explanation :



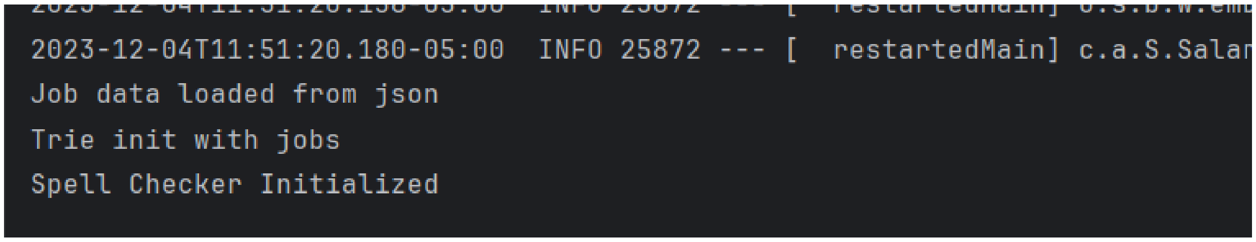
```
public TrieNode() {
   this.children = new HashMap<>();
   this.jobIds = new HashSet<>();
   this.wordFrequency = new TreeMap<>();
   this.isEndOfWord = false;
}
```

The code implements inverted indexing for efficient search operations in a large dataset of job descriptions. In the initialization of the trie (TrieDS), each job's description is processed to extract relevant tokens, excluding common English stop words. These processed tokens are then inserted into the trie data structure using the insertIntoTrie method. The trie enables quick retrieval of job IDs associated with specific words, forming an inverted index. This facilitates the search process by providing immediate access to job entries containing the queried words. The TrieNode class represents each node in the trie and includes a set of job IDs, establishing a direct link between words and their occurrences in job descriptions.

For frequency counting, the code utilizes the TrieNode's wordFrequency TreeMap within the insertIntoTrie method. This TreeMap stores the frequencies of each word within the job descriptions, organized by job ID. As words are inserted into the trie, their frequencies are updated in the corresponding TrieNode. This frequency count mechanism allows for efficient analysis of word occurrences across the dataset. During subsequent searches or data analysis, the Trie structure provides immediate access to both the job IDs containing specific words and the frequencies of those words within each job description. Overall, the combination of inverted indexing and frequency counting in the trie enhances the speed and efficiency of search operations in large-scale job datasets.

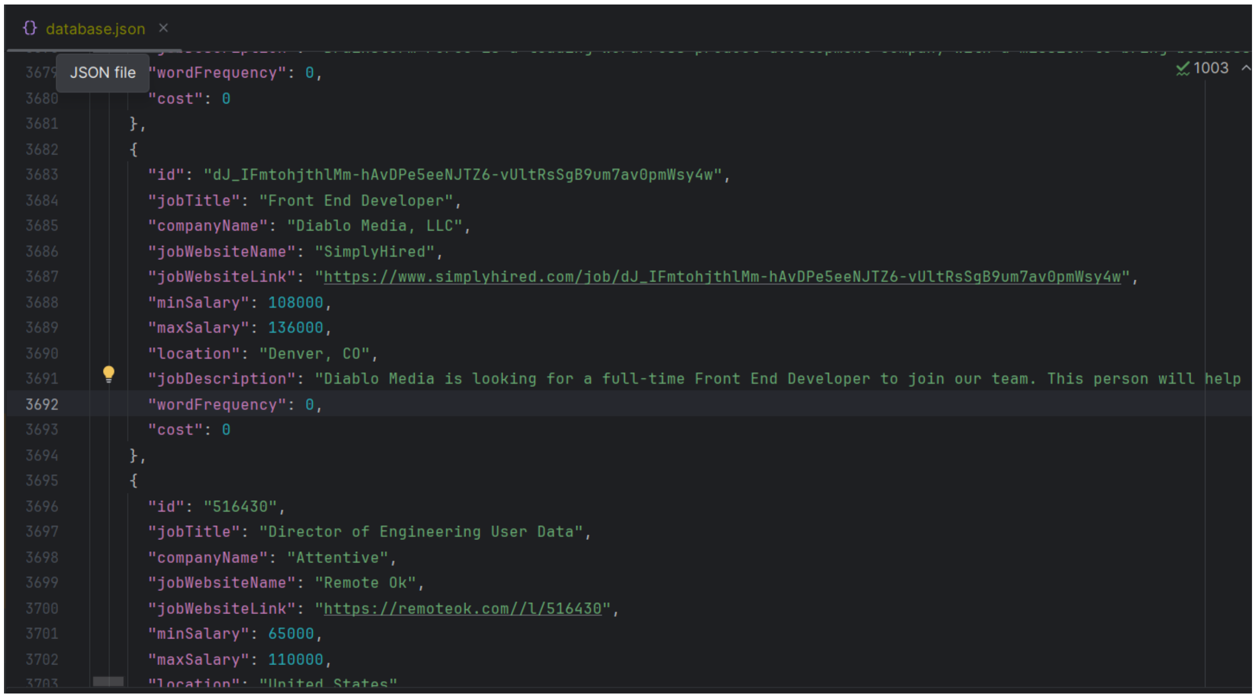
Output 1 :

The trie gets initialized with database.json every time the application starts



```
2023-12-04T11:51:20.158-05:00  INFO 25872 --- [  restartedMain] o.s.b.w.emb
2023-12-04T11:51:20.180-05:00  INFO 25872 --- [  restartedMain] c.a.S.Salar
Job data loaded from json
Trie init with jobs
Spell Checker Initialized
```

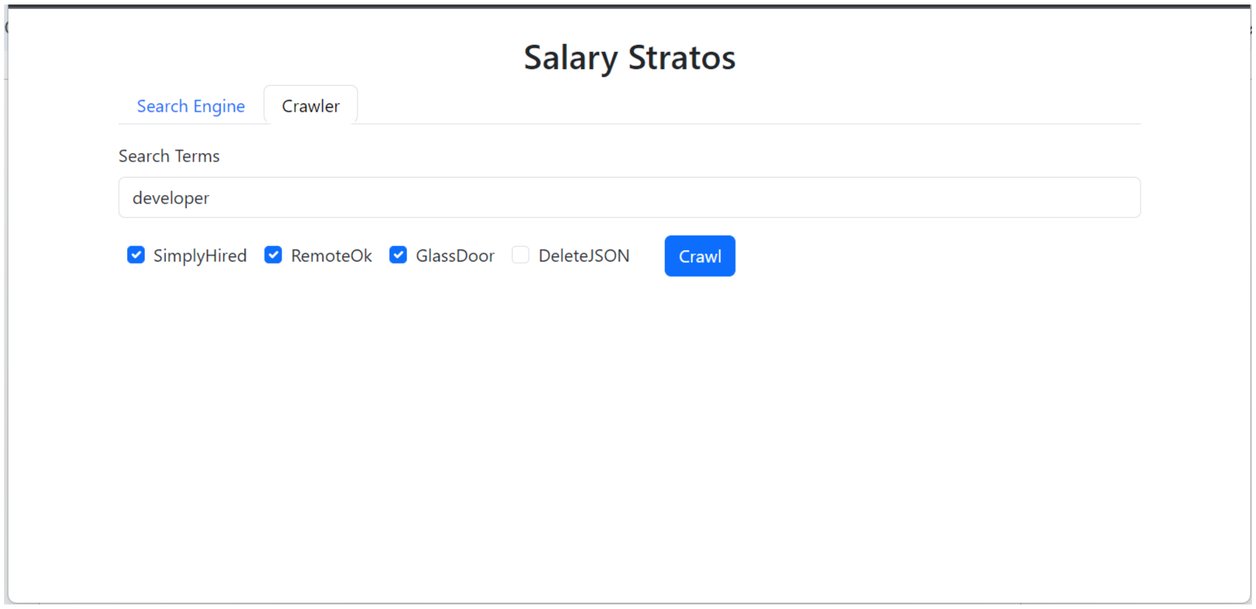
The words from database.json is used to initialize the trie

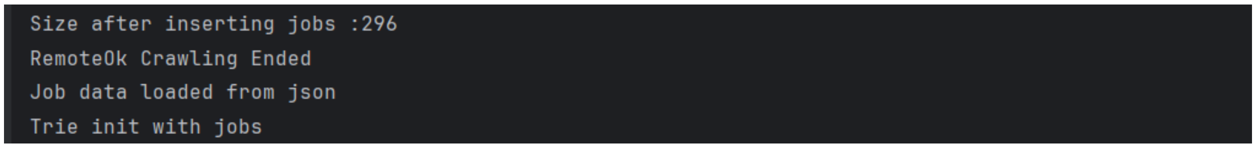


```
    "wordFrequency": 0,
    "cost": 0
  },
  {
    "id": "dJ_IFmtohjthlMm-hAvDPe5eeNJTZ6-vUltRsSgB9um7av0pmWsy4w",
    "jobTitle": "Front End Developer",
    "companyName": "Diablo Media, LLC",
    "jobWebsiteName": "SimplyHired",
    "jobWebsiteLink": "https://www.simplyhired.com/job/dJ_IFmtohjthlMm-hAvDPe5eeNJTZ6-vUltRsSgB9um7av0pmWsy4w",
    "minSalary": 108000,
    "maxSalary": 136000,
    "location": "Denver, CO",
    "jobDescription": "Diablo Media is looking for a full-time Front End Developer to join our team. This person will help
    "wordFrequency": 0,
    "cost": 0
  },
  {
    "id": "516430",
    "jobTitle": "Director of Engineering User Data",
    "companyName": "Attentive",
    "jobWebsiteName": "Remote Ok",
    "jobWebsiteLink": "https://remoteok.com//l/516430",
    "minSalary": 65000,
    "maxSalary": 110000,
    "location": "United States",
```

Output 2 :

The trie gets reinitialized whenever we scrape the data from the ui





```
Size after inserting jobs :296
RemoteOk Crawling Ended
Job data loaded from json
Trie init with jobs
```

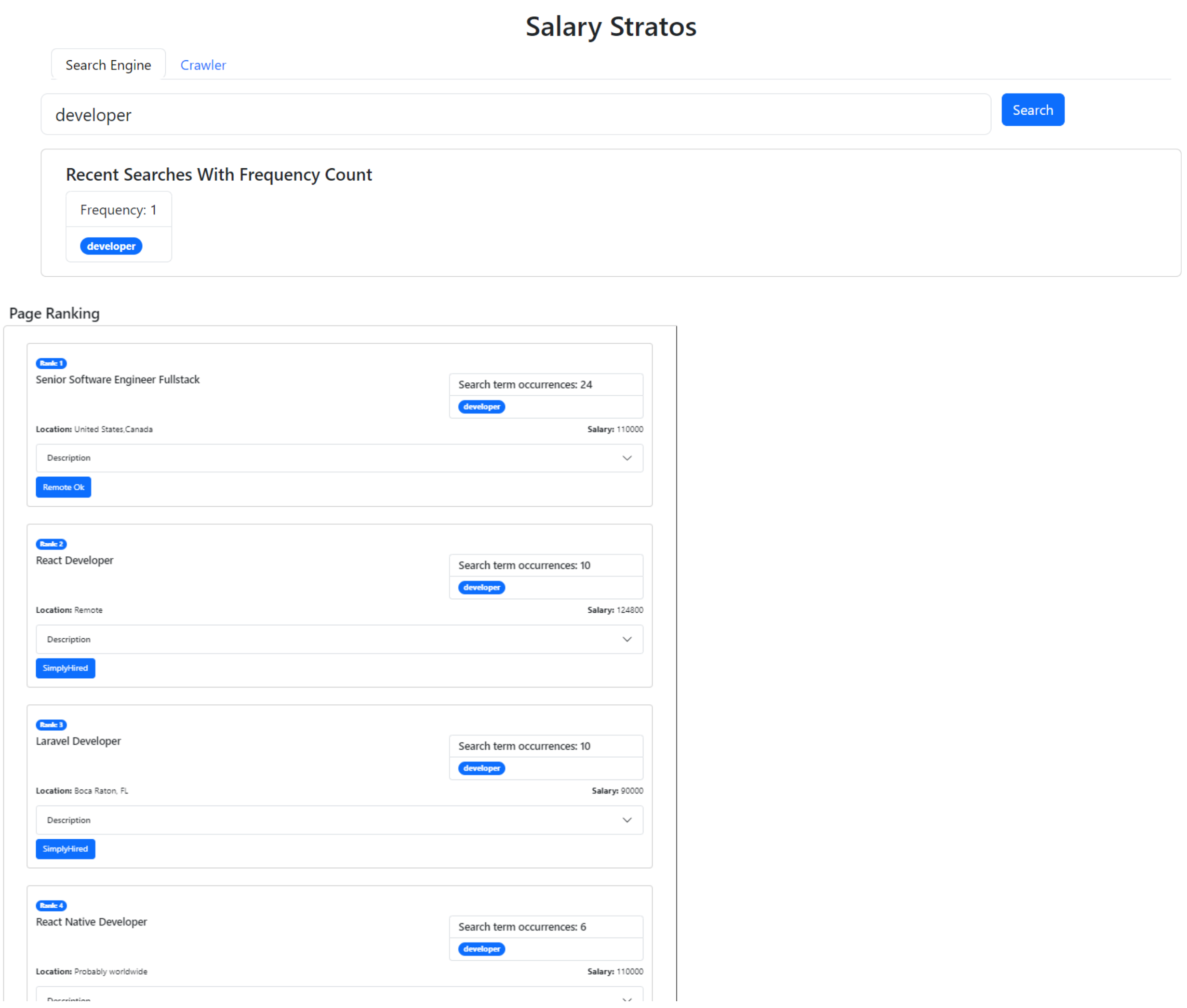
5. Page Ranking :

Code Explanation :

The PageRanking class implements page ranking by searching inverted indexed data stored in a Trie data structure. It takes an array of search terms, retrieves their frequencies from the Trie, and ranks jobs based on the cumulative word frequencies of the search terms within each job's description. For each term, it iterates over the corresponding Trie node to obtain job IDs and their word frequencies. Then, for each job ID, it retrieves the corresponding job details from the stored job data, updating the job's accumulated word frequency and word field. The jobs are then inserted into a SortedArray, sorted based on their word frequencies. If a job already has a non-zero frequency, it is temporarily removed from the sorted structure for updating before being reinserted. This is done because if we don't remove the word it will create duplication of jobs in the response array if we search for multiple search terms. The resulting SortedArray contains jobs ranked according to the cumulative frequency of the search terms within their descriptions, providing a page ranking reflecting the relevance of jobs to the search criteria.

Output 1 :

Result on the ui for one search term



Salary Stratos

Search Engine | Crawler

developer | Search

Recent Searches With Frequency Count

Frequency: 1
developer

Page Ranking

Rank: 1
Senior Software Engineer Fullstack
Search term occurrences: 24
developer
Location: United States,Canada
Salary: 110000
Description
Remote Ok

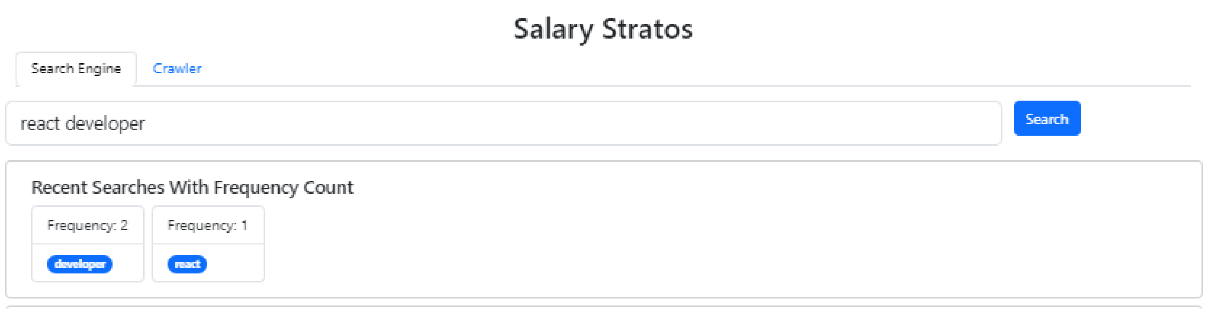
Rank: 2
React Developer
Search term occurrences: 10
developer
Location: Remote
Salary: 124800
Description
SimplyHired

Rank: 3
Laravel Developer
Search term occurrences: 10
developer
Location: Boca Raton, FL
Salary: 90000
Description
SimplyHired

Rank: 4
React Native Developer
Search term occurrences: 6
developer
Location: Probably worldwide
Salary: 110000

<u>Output 2 :</u>

Result on the ui for two search terms



Salary Stratos

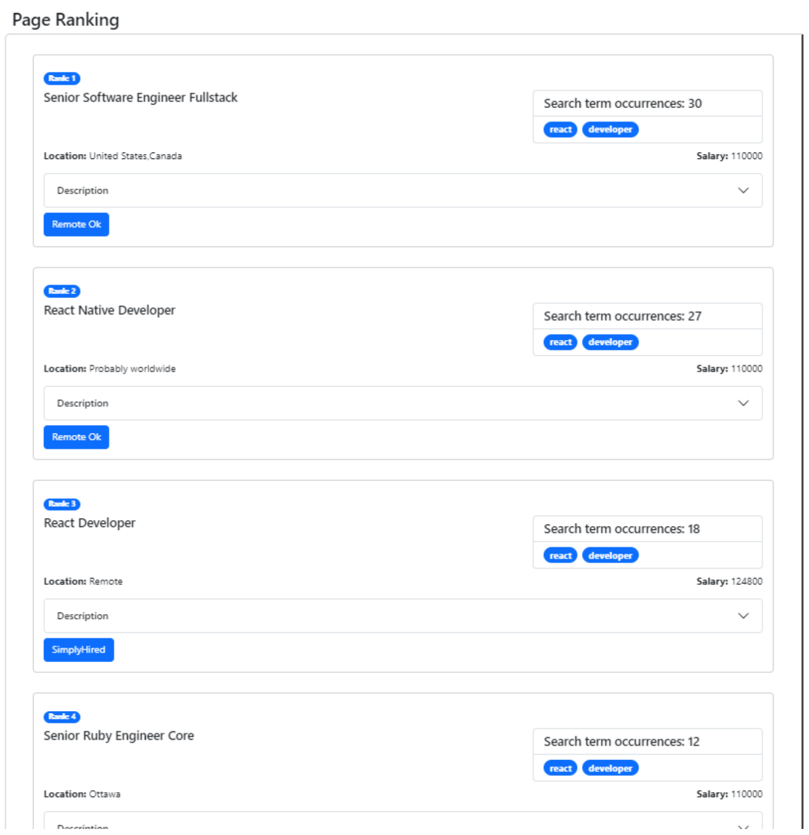
Search Engine | Crawler

react developer | Search

Recent Searches With Frequency Count

Frequency: 2
developer

Frequency: 1
react



Page Ranking

Rank: 1
Senior Software Engineer Fullstack
Search term occurrences: 30
react developer
Location: United States,Canada
Salary: 110000
Description
Remote Ok

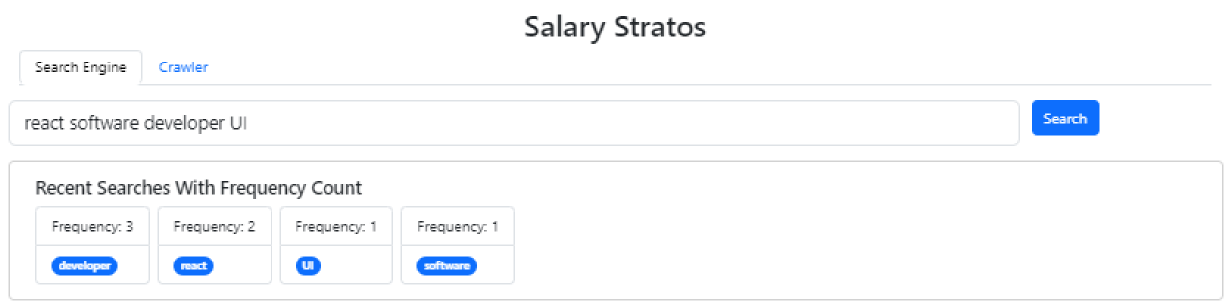
Rank: 2
React Native Developer
Search term occurrences: 27
react developer
Location: Probably worldwide
Salary: 110000
Description
Remote Ok

Rank: 3
React Developer
Search term occurrences: 18
react developer
Location: Remote
Salary: 124800
Description
SimplyHired

Rank: 4
Senior Ruby Engineer Core
Search term occurrences: 12
react developer
Location: Ottawa
Salary: 110000
Description

<u>Output 3 :</u>

Result on the ui for multiple search terms



Salary Stratos

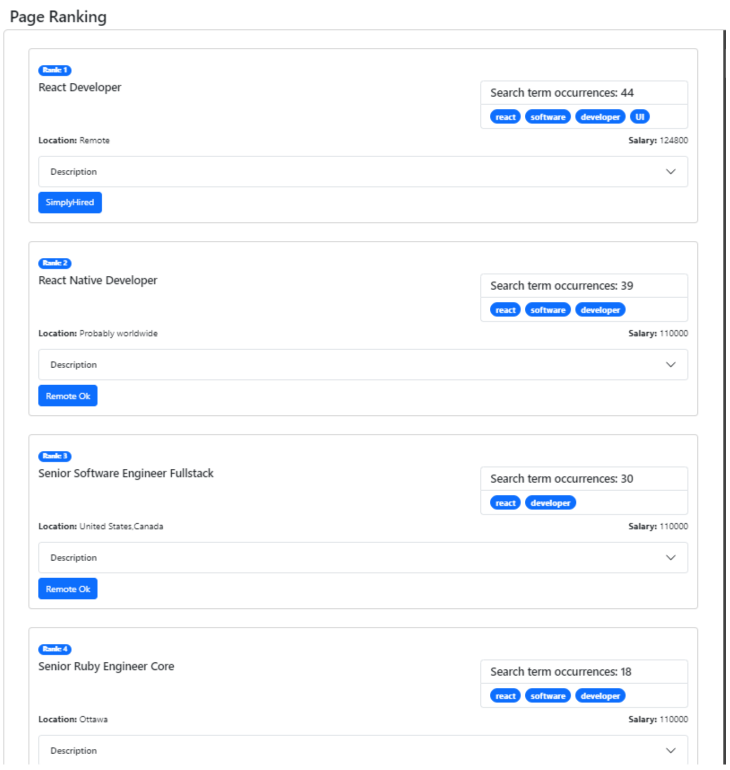
Search Engine Crawler

react software developer UI

Search

Recent Searches With Frequency Count

Frequency: 3	Frequency: 2	Frequency: 1	Frequency: 1
developer	react	UI	software


Page Ranking

Rank: 1
React Developer
Search term occurrences: 44
react software developer UI
Location: Remote
Salary: 124800
Description
SimplyHired

Rank: 2
React Native Developer
Search term occurrences: 39
react software developer
Location: Probably worldwide
Salary: 110000
Description
Remote Ok

Rank: 3
Senior Software Engineer Fullstack
Search term occurrences: 30
react developer
Location: United States,Canada
Salary: 110000
Description
Remote Ok

Rank: 4
Senior Ruby Engineer Core
Search term occurrences: 18
react software developer
Location: Ottawa
Salary: 110000
Description

6. Spell checking :

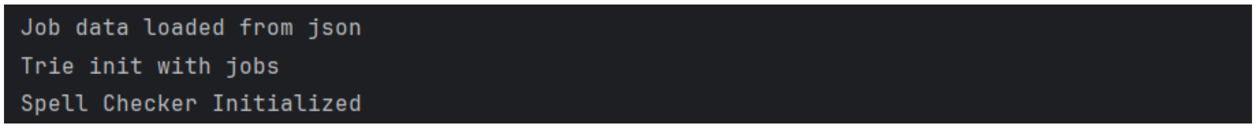
Code Explanation :

The SpellChecker class is designed to perform spell checking by maintaining a dictionary of valid words and employing the EditDistanceAlgo for suggesting similar words in case of misspelled input. The initialization process, handled by the initializeSpellChecker method, reads words from an external dictionary file (dictionaryOfWords.txt) and populates both a Trie data structure (TrieDS) and a frequency map (dict). The dict also contains all the words and the frequencies of the word from the database.json. The TrieDS efficiently stores words for quick retrieval and supports autocomplete features.

In the suggestSimilarWord method, given an input word, the code first checks if it's empty or belongs to a predefined list of invalid words. If the word is valid, it searches for an exact match in the Trie. If no match is found, it utilizes the EditDistanceAlgo to calculate the edit distance between the input word and all words in the dictionary. The results are organized in a nested TreeMap structure, where the outer TreeMap sorts suggestions based on edit distance, and the inner TreeMap further organizes them by word frequency. This prioritization ensures that suggestions are not only linguistically similar but also ranked by their prevalence in the dictionary, contributing to more accurate and contextually relevant spell-checking suggestions.

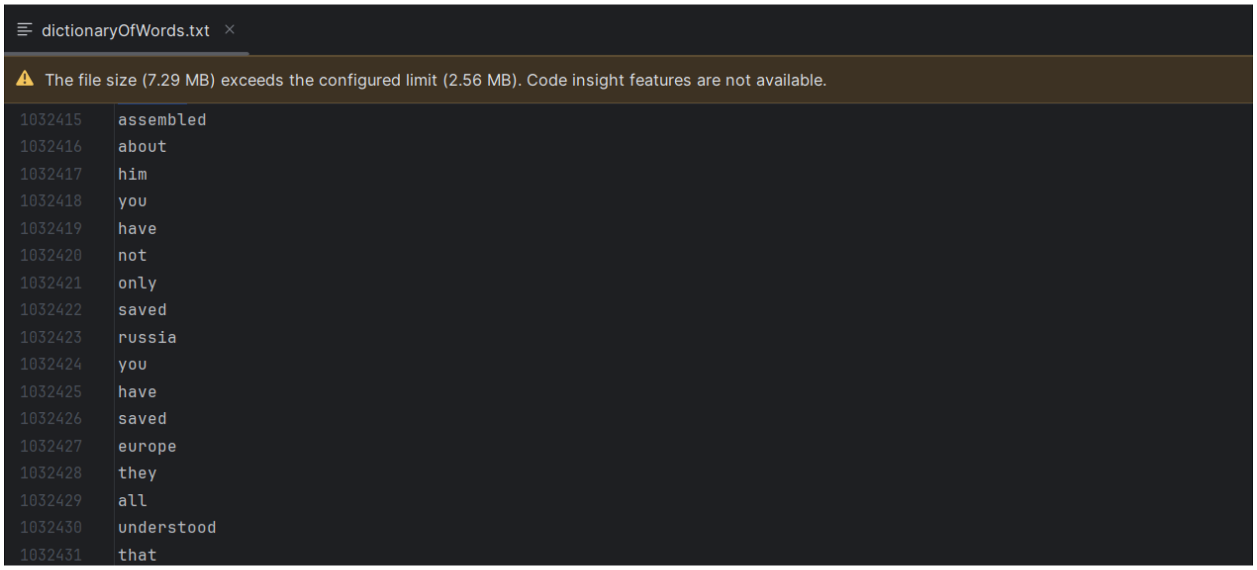
In summary, the spell-checking functionality leverages a Trie data structure, an external dictionary, and the EditDistanceAlgo to offer efficient and accurate suggestions for potential corrections based on word similarity and frequency. The data structures and algorithms employed contribute to a robust spell-checking mechanism within the provided code.

Output 1 :
When the program starts the spell checker dictionary is initialized.



```
Job data loaded from json
Trie init with jobs
Spell Checker Initialized
```

Data gets initialized from the dictionaryOfWords.txt

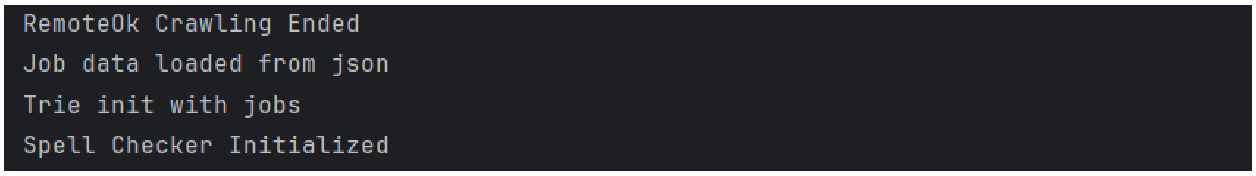


dictionaryOfWords.txt

The file size (7.29 MB) exceeds the configured limit (2.56 MB). Code insight features are not available.

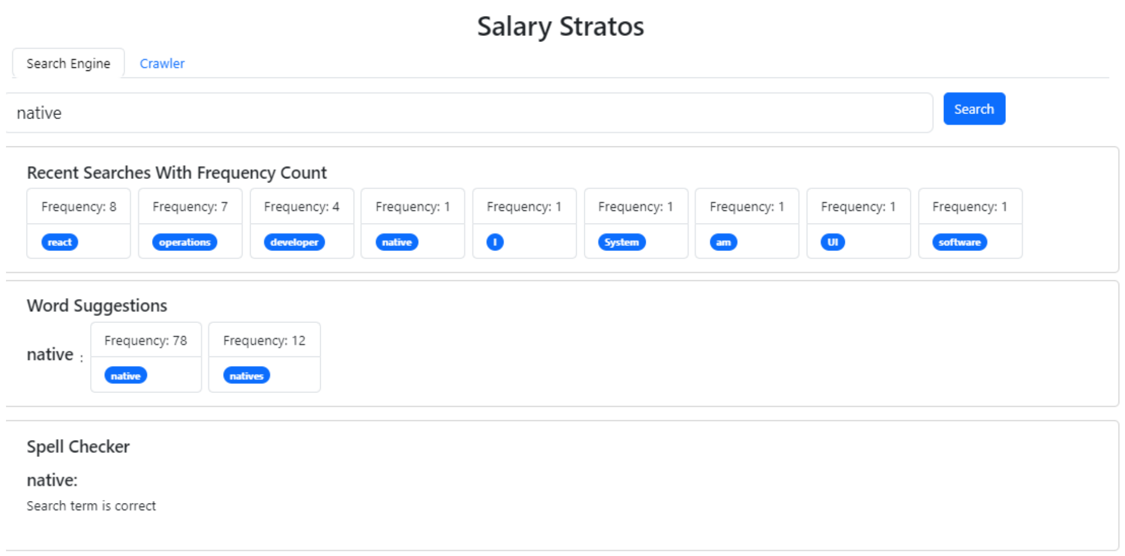
```
assembled
about
him
you
have
not
only
saved
russia
you
have
saved
europe
they
all
understood
that
```

Output 2 :
When the Scrapper is called, the spell checker dictionary is initialized.



```
RemoteOk Crawling Ended
Job data loaded from json
Trie init with jobs
Spell Checker Initialized
```

Output 3 :
Spell checking for one search term when the user hits enter. The search term native is correct.


Salary Stratos

Search Engine Crawler

native

Search

Recent Searches With Frequency Count

Frequency: 8	Frequency: 7	Frequency: 4	Frequency: 1	Frequency: 1	Frequency: 1	Frequency: 1	Frequency: 1	Frequency: 1
react	operations	developer	native	I	System	am	UI	software

Word Suggestions

native :

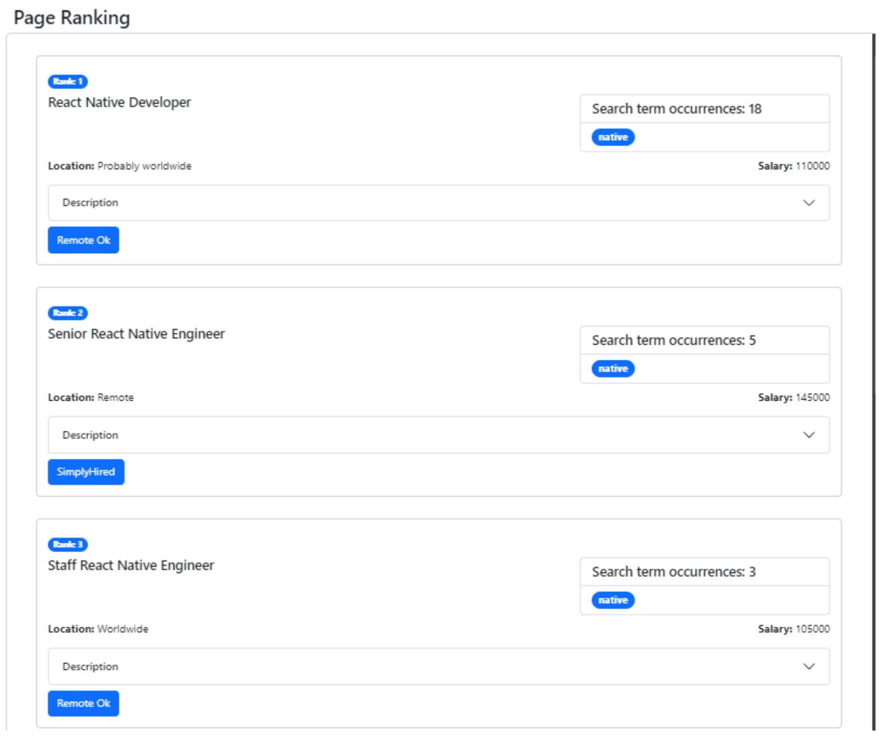
Frequency: 78	Frequency: 12
native	natives

Spell Checker

native:

Search term is correct

It will search for native


Page Ranking

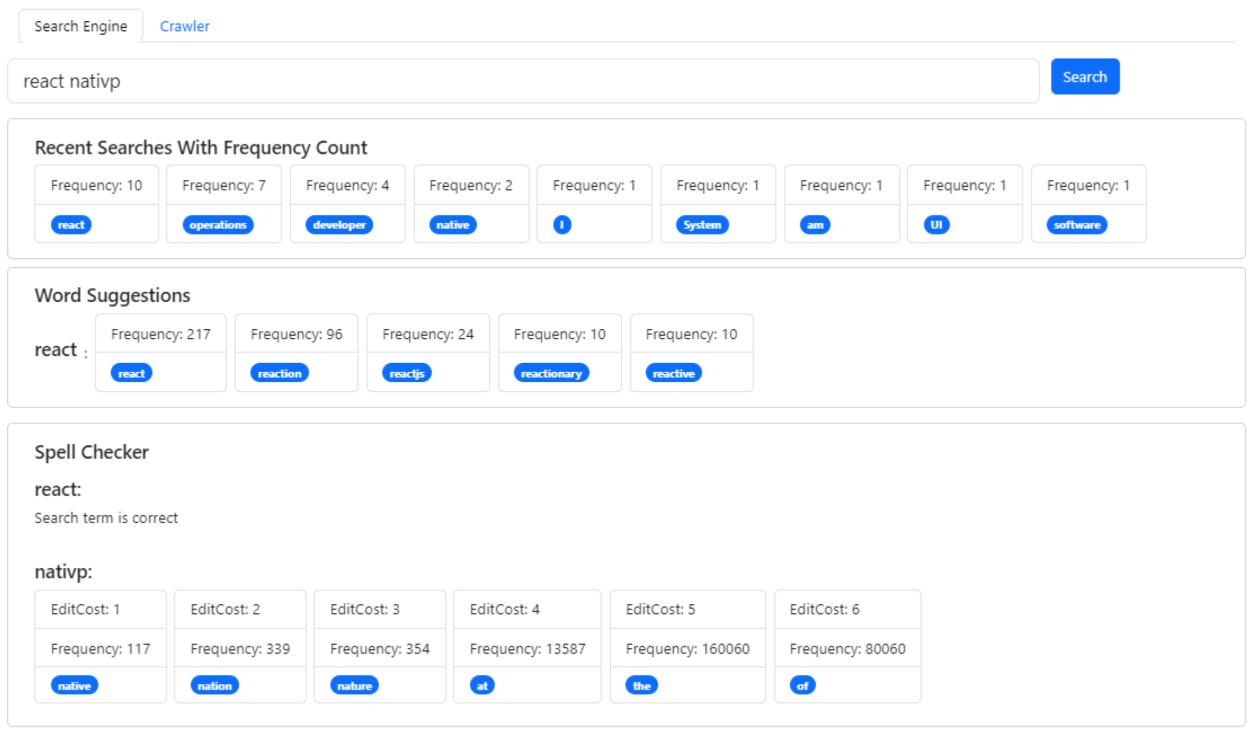
Rank: 1
React Native Developer
Search term occurrences: 18
native
Location: Probably worldwide
Salary: 110000
Description
Remote Ok

Rank: 2
Senior React Native Engineer
Search term occurrences: 5
native
Location: Remote
Salary: 145000
Description
SimplyHired

Rank: 3
Staff React Native Engineer
Search term occurrences: 3
native
Location: Worldwide
Salary: 105000
Description
Remote Ok

<u>Output 4 :</u>

Spell checking for two search terms when the user hits enter. The second search term nativp is not correct but the closest word to this is native so it suggested that based on edit distance


Search Engine
Crawler

react nativp

Search

Recent Searches With Frequency Count

Frequency: 10 react
Frequency: 7 operations
Frequency: 4 developer
Frequency: 2 native
Frequency: 1 I
Frequency: 1 System
Frequency: 1 am
Frequency: 1 UI
Frequency: 1 software

Word Suggestions

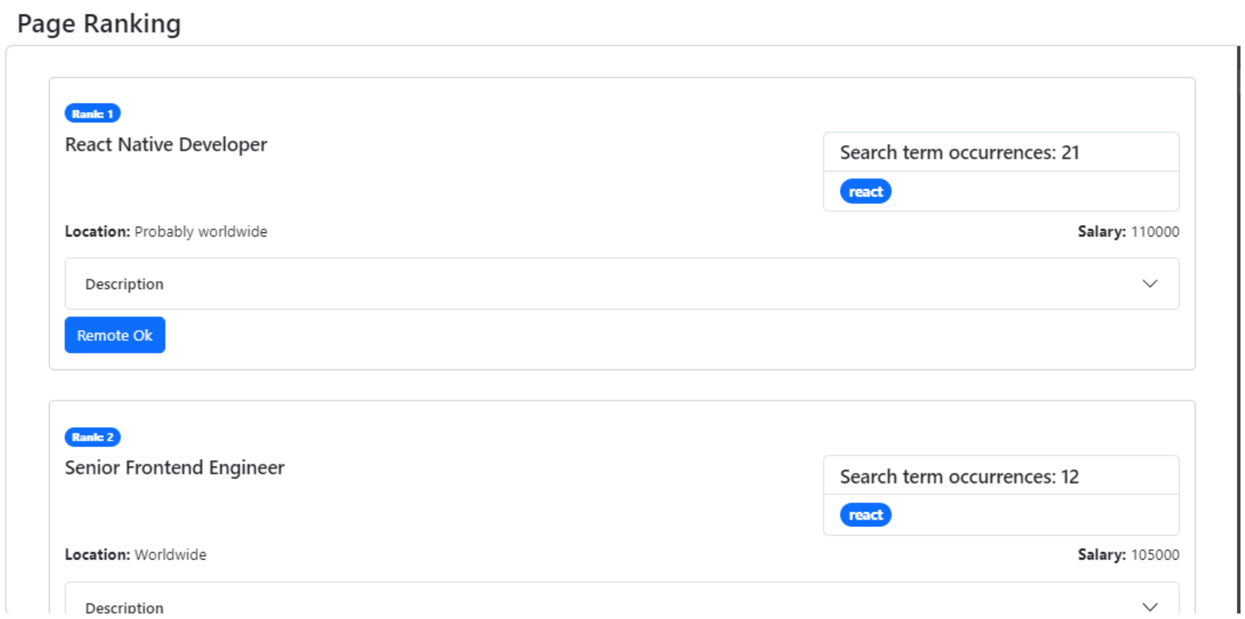
react :
Frequency: 217 react
Frequency: 96 reaction
Frequency: 24 reactjs
Frequency: 10 reactionary
Frequency: 10 reactive

Spell Checker

react:
Search term is correct

nativp:
EditCost: 1 Frequency: 117 native
EditCost: 2 Frequency: 339 nation
EditCost: 3 Frequency: 354 nature
EditCost: 4 Frequency: 13587 at
EditCost: 5 Frequency: 160060 the
EditCost: 6 Frequency: 80060 of

It will only search for react in this case as the second term is wrong.


Page Ranking

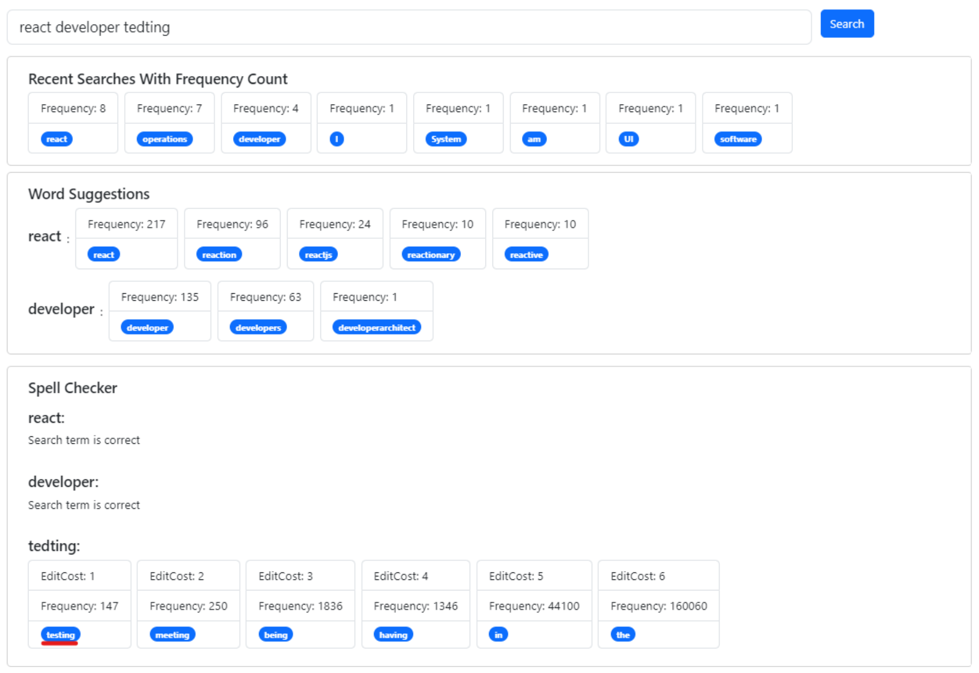
Rank: 1
React Native Developer
Search term occurrences: 21
react
Location: Probably worldwide
Salary: 110000
Description
Remote Ok

Rank: 2
Senior Frontend Engineer
Search term occurrences: 12
react
Location: Worldwide
Salary: 105000
Description

<u>Output 5 :</u>

Spell checking for multiple search terms when the user hits enter.

If we observe below we entered 3 words “react developer tedting”. We can see that react and operations are correct words but tesdting is a wrong word. It also gave the closest word to tesdtion i.e. testing based on Edit Distance.


react developer tedting

Search

Recent Searches With Frequency Count

Frequency: 8	Frequency: 7	Frequency: 4	Frequency: 1	Frequency: 1	Frequency: 1	Frequency: 1	Frequency: 1
react	operations	developer	I	System	am	UI	software

Word Suggestions

react :

Frequency: 217	Frequency: 96	Frequency: 24	Frequency: 10	Frequency: 10
react	reaction	reactjs	reactionary	reactive

developer :

Frequency: 135	Frequency: 63	Frequency: 1
developer	developers	developerarchitect

Spell Checker

react:

Search term is correct

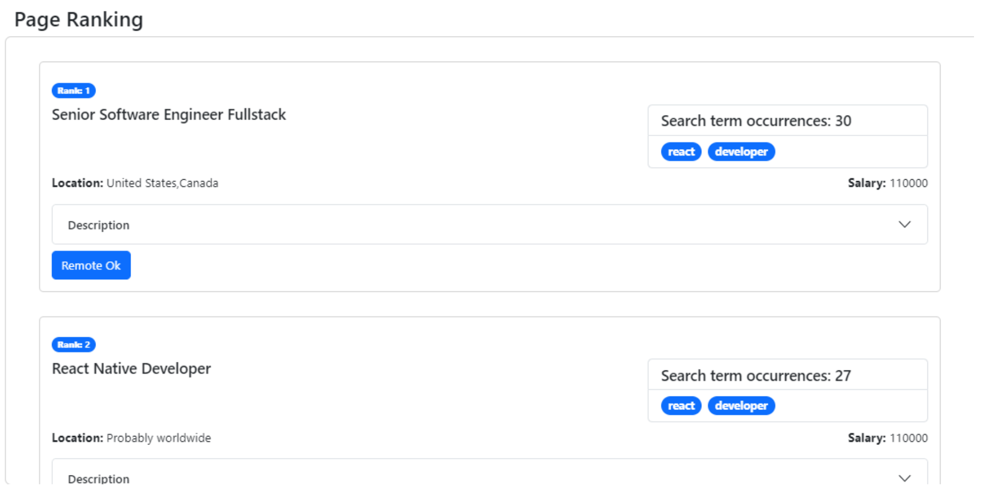
developer:

Search term is correct

tedting:

EditCost: 1	EditCost: 2	EditCost: 3	EditCost: 4	EditCost: 5	EditCost: 6
Frequency: 147	Frequency: 250	Frequency: 1836	Frequency: 1346	Frequency: 44100	Frequency: 160060
testing	meeting	being	having	in	the

The search page ranking is done only for two words in this case "react developer"


Page Ranking

Rank: 1

Senior Software Engineer Fullstack

Search term occurrences: 30

react developer

Location: United States,Canada

Salary: 110000

Description

Remote Ok

Rank: 2

React Native Developer

Search term occurrences: 27

react developer

Location: Probably worldwide

Salary: 110000

Description

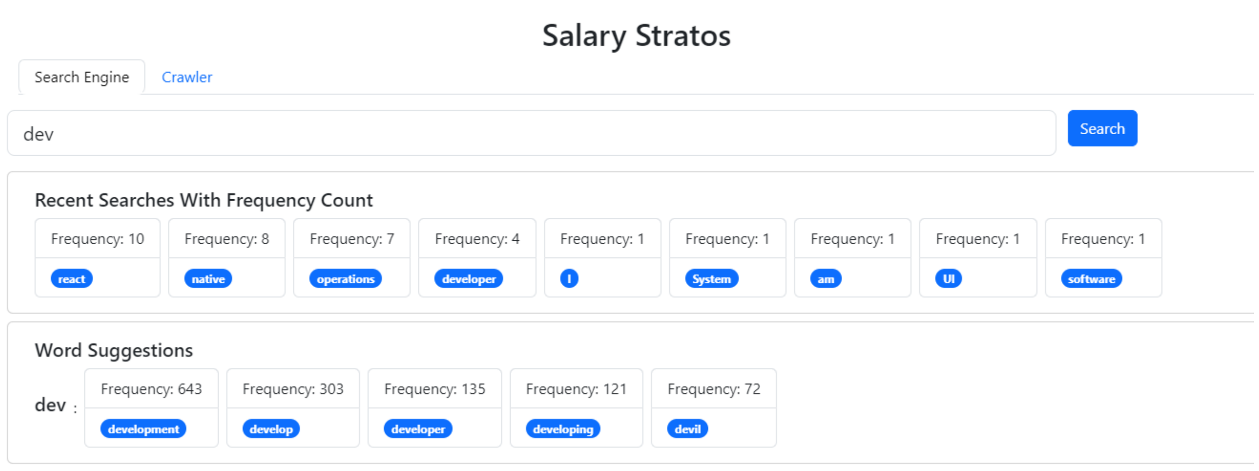
7. Word Completion :

Code Explanation :

The WordCompletion class provides functionality to generate word suggestions based on validated search terms using a Trie data structure. The getWordSuggestions method takes a list of validated search terms, a Trie structure (JobDataTrie), and the desired count of suggestions. It iterates through each search term, retrieves word suggestions from the Trie using the searchInTrieWithPrefix method, and constructs a response containing the original search term and a SortedArray of suggested words along with their frequencies. The searchInTrieWithPrefix method traverses the Trie to find words with the specified prefix, and the collectWords recursive function gathers word frequencies from the Trie nodes, constructing a SortedArray of WordFrequency objects sorted by frequency. The resulting WordSuggestionResponse encapsulates the original search terms, whether the response is valid, and a list of suggested words with their frequencies.

Output 1 :

Word Suggestion for one search term when the user enters a partial word. It's not necessary click on enter to see the word suggestions



Salary Stratos

Search Engine Crawler

dev

Search

Recent Searches With Frequency Count

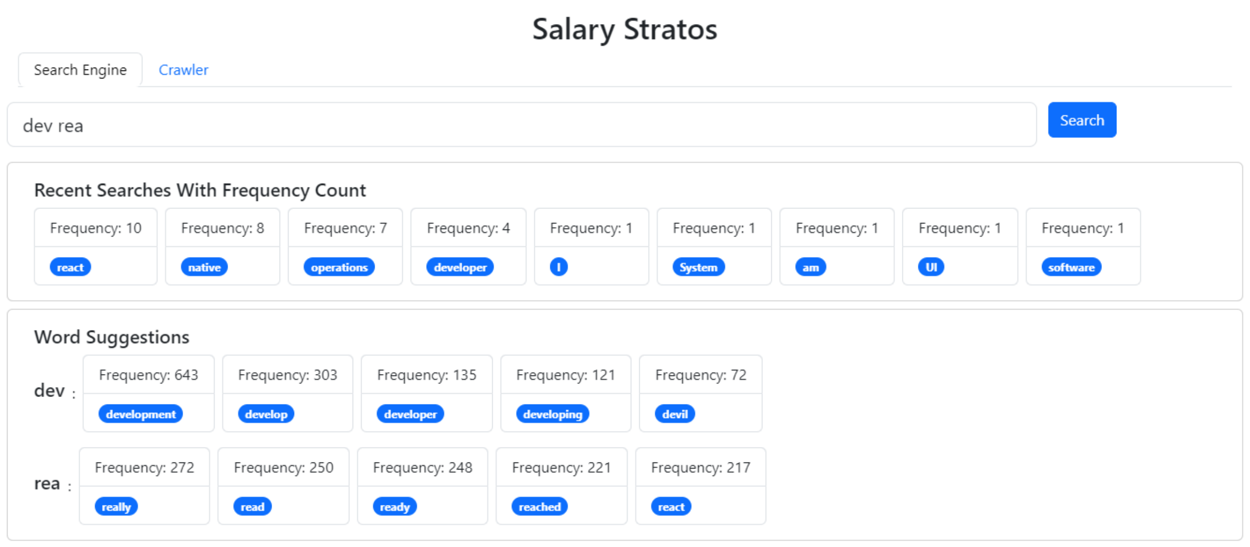
Frequency: 10	Frequency: 8	Frequency: 7	Frequency: 4	Frequency: 1	Frequency: 1	Frequency: 1	Frequency: 1	Frequency: 1
react	native	operations	developer	I	System	am	UI	software

Word Suggestions

dev :	Frequency: 643	Frequency: 303	Frequency: 135	Frequency: 121	Frequency: 72
	development	develop	developer	developing	devil

Output 2 :

Word Suggestion for two search terms when the user enters a partial word



Salary Stratos

Search Engine Crawler

dev rea

Search

Recent Searches With Frequency Count

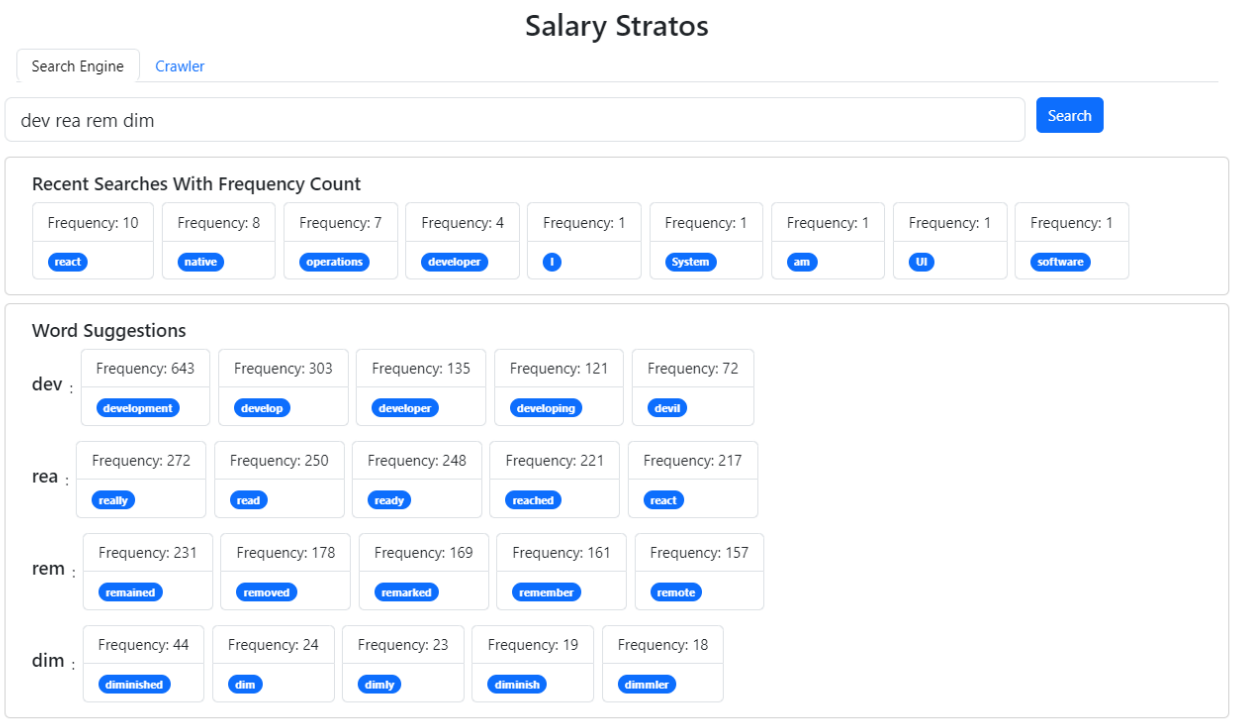
Frequency: 10	Frequency: 8	Frequency: 7	Frequency: 4	Frequency: 1	Frequency: 1	Frequency: 1	Frequency: 1	Frequency: 1
react	native	operations	developer	I	System	am	UI	software

Word Suggestions

dev :	Frequency: 643	Frequency: 303	Frequency: 135	Frequency: 121	Frequency: 72
	development	develop	developer	developing	devil
rea :	Frequency: 272	Frequency: 250	Frequency: 248	Frequency: 221	Frequency: 217
	really	read	ready	reached	react

Output 3 :

Word Suggestion for multiple search terms when the user enters a partial word


Salary Stratos

dev rea rem dim

Recent Searches With Frequency Count

Frequency: 10	Frequency: 8	Frequency: 7	Frequency: 4	Frequency: 1	Frequency: 1	Frequency: 1	Frequency: 1	Frequency: 1
react	native	operations	developer	I	System	am	UI	software

Word Suggestions

dev :	Frequency: 643	Frequency: 303	Frequency: 135	Frequency: 121	Frequency: 72
	development	develop	developer	developing	devil
rea :	Frequency: 272	Frequency: 250	Frequency: 248	Frequency: 221	Frequency: 217
	really	read	ready	reached	react
rem :	Frequency: 231	Frequency: 178	Frequency: 169	Frequency: 161	Frequency: 157
	remained	removed	remarked	remember	remote
dim :	Frequency: 44	Frequency: 24	Frequency: 23	Frequency: 19	Frequency: 18
	diminished	dim	dimly	diminish	dimmler

8. Search frequency :

Code Explanation :

The SearchFrequency class manages search term frequencies using an LRUCache, maintaining a limited size to track the most recent and frequently used search terms. The LRUCache, implemented as a custom class extending LinkedHashMap, enforces a Least Recently Used behavior. It stores search terms and their frequencies, automatically removing the least recently used entry when the cache exceeds the specified size. The displaySearchFrequencies method retrieves the last 10 search term frequencies in descending order using a custom SortedArray data structure, providing a dynamically updated view of recent search term activities. The updateSearchFrequency method ensures accurate tracking of user interactions by updating the search term frequencies based on an array of search terms. Overall, this component efficiently combines LRUCache and sorted arrays to offer insights into recent user search patterns.

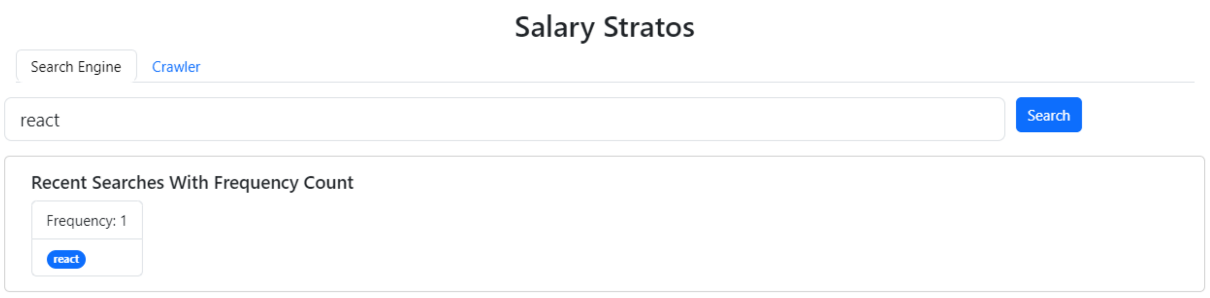
LRU Cache size is 50

The return array size to ui is 10

If the Cache size becomes bigger than 50 it will remove the least recently used word from the LRU Cache

Output 1 :

First word entered into the LRU Cache "react"


Salary Stratos

Search Engine | Crawler

react

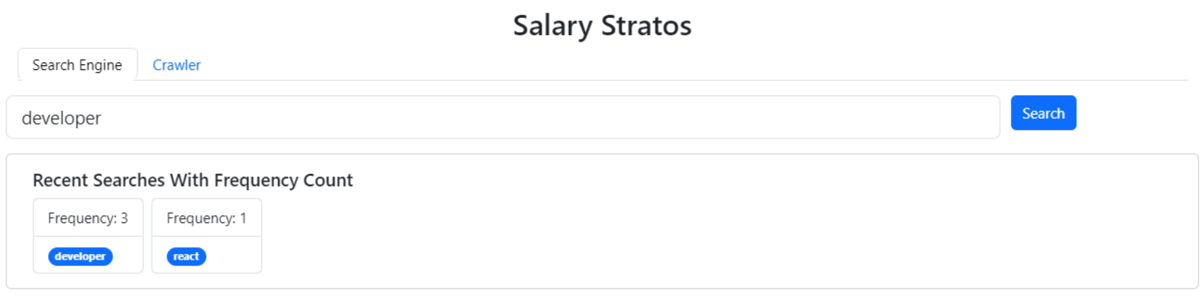
Search

Recent Searches With Frequency Count

Frequency: 1 — react

Output 2 :

Second word entered into the LRU cache "developer" which is searched 3 times


Salary Stratos

Search Engine | Crawler

developer

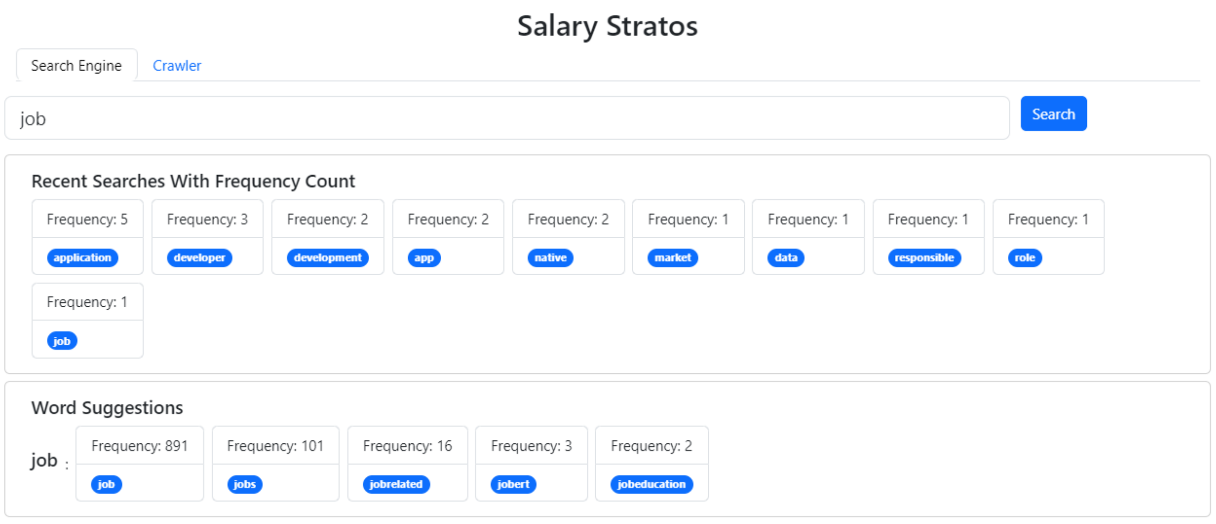
Search

Recent Searches With Frequency Count

Frequency: 3 — developer
Frequency: 1 — react

Output 3 :

After inserting 10 words we see that "react" is not shown in the ui anymore because it is not used for a while, but it still remains in the cache till the cache size becomes 50, when this happens if "react" is the least recently used word it will get removed from the cache.


Salary Stratos

Search Engine | Crawler

job

Search

Recent Searches With Frequency Count

Frequency: 5 — application
Frequency: 3 — developer
Frequency: 2 — development
Frequency: 2 — app
Frequency: 2 — native
Frequency: 1 — market
Frequency: 1 — data
Frequency: 1 — responsible
Frequency: 1 — role
Frequency: 1 — job

Word Suggestions

job :
Frequency: 891 — job
Frequency: 101 — jobs
Frequency: 16 — jobrelated
Frequency: 3 — jobert
Frequency: 2 — jobeducation

9. Finding patterns using regular expressions :

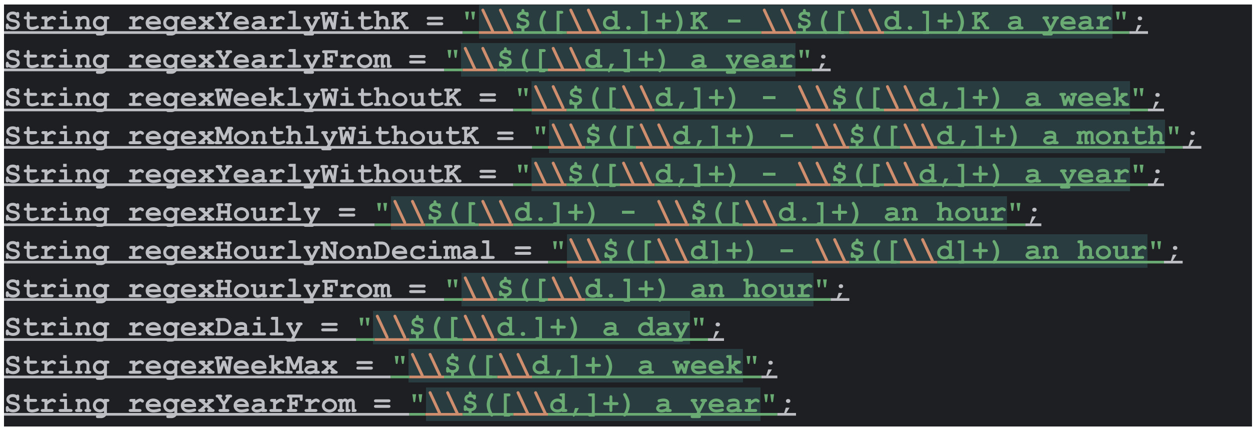
Code Explanation :

The scrapeWebPage method in classes such as SimplyHiredScrapper, RemoteOk, and GlassDoorScrapper utilizes regular expressions (Regex) to identify various salary formats and convert them into a consistent yearly format. The code initializes multiple regular expression patterns (regexYearlyWithK, regexYearlyFrom, etc.) designed to match different salary structures, such as hourly, weekly, monthly, or yearly, with or without the use of 'K' (thousands) and other variations.

After extracting the raw salary string from the web page, the code checks which regex pattern matches the format and proceeds with appropriate processing. For instance, it replaces the matched pattern with specific placeholders and then uses replaceAll to extract numeric values. It further adjusts these values based on the specific time unit (hours, weeks, etc.) to convert them into a standardized yearly representation. The resulting minimum and maximum salary values are then utilized in the application for analysis and comparison. This approach ensures accurate extraction and uniform representation of salary data, providing consistency in further processing and analysis

<u>Output 1 :</u>

If salary in Simply Hired is anything other than the below it will print the failure message



```
String regexYearlyWithK = "\\$([\\d.]+)K - \\$([\\d.]+)K a year";
String regexYearlyFrom = "\\$([\\d,]+) a year";
String regexWeeklyWithoutK = "\\$([\\d,]+) - \\$([\\d,]+) a week";
String regexMonthlyWithoutK = "\\$([\\d,]+) - \\$([\\d,]+) a month";
String regexYearlyWithoutK = "\\$([\\d,]+) - \\$([\\d,]+) a year";
String regexHourly = "\\$([\\d.]+) - \\$([\\d.]+) an hour";
String regexHourlyNonDecimal = "\\$([\\d]+) - \\$([\\d]+) an hour";
String regexHourlyFrom = "\\$([\\d.]+) an hour";
String regexDaily = "\\$([\\d.]+) a day";
String regexWeekMax = "\\$([\\d,]+) a week";
String regexYearFrom = "\\$([\\d,]+) a year";
```

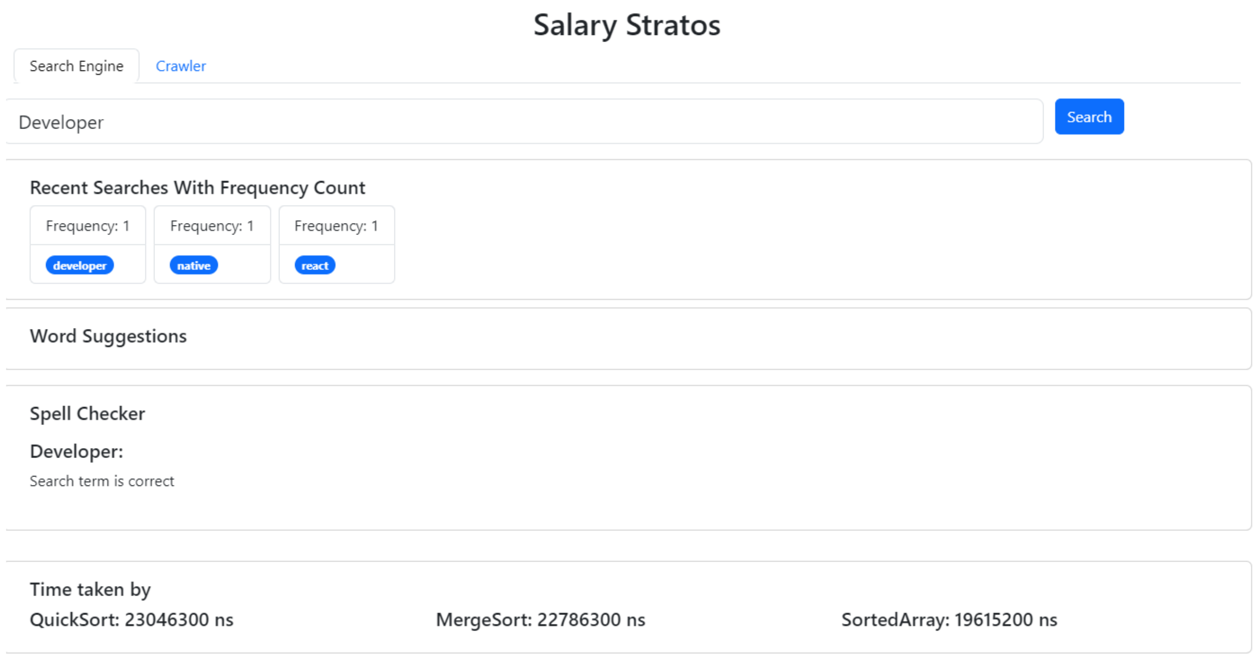
10. Compare sorting algorithms for Page ranking :

<u>Code Explanation :</u>

The CompareRunTimesData class is designed to store and manage runtimes for different sorting algorithms, particularly Merge Sort, Binary Search, and QuickSort, within the context of inverted indexing and page ranking. Each sorting algorithm's runtime is tracked through dedicated variables (mergeSort, binarySearch, quickSort) along with respective getter and setter methods. This implementation is intended to facilitate runtime comparison between these algorithms when applied to the specific task of sorting jobs retrieved from the inverted index. By setting and retrieving the runtimes for each sorting operation, this class provides a streamlined way to evaluate and contrast the efficiency of these sorting methods in the context of page ranking tasks.

In the context of maintaining a sorted array for inverted indexing and page ranking, binary search holds advantages over Merge Sort and QuickSort. Binary search excels in scenarios where the data is already partially sorted, which aligns with the situation encountered when maintaining a sorted array. Given its time complexity of O(log n) for finding the correct insertion position in a sorted array, binary search proves notably faster than both QuickSort (O(n log n)) and MergeSort (O(n log n)) in this specific context. Binary search's efficiency in locating insertion positions within a sorted array directly complements the task of maintaining order in the context of inverted indexing, making it a more suitable choice in scenarios where incremental sorting is required while minimizing time complexity.

<u>Output 1 :</u>



Developer

Recent Searches With Frequency Count

Frequency: 1 developer
Frequency: 1 native
Frequency: 1 react

Word Suggestions

Spell Checker

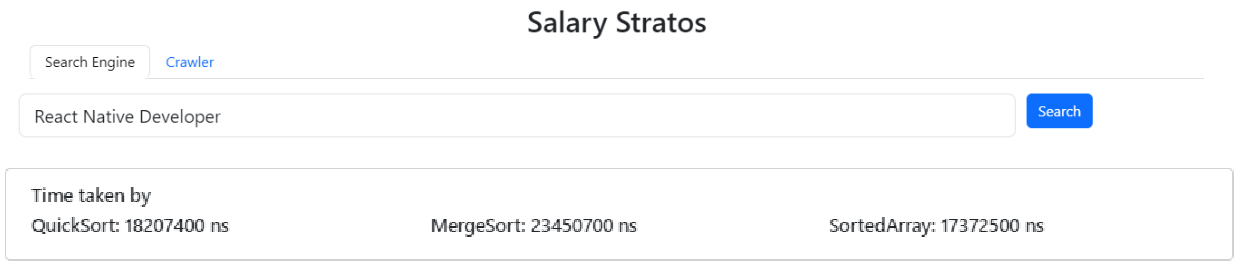
Developer:

Search term is correct

Time taken by

QuickSort: 23046300 ns MergeSort: 22786300 ns SortedArray: 19615200 ns

<u>Output 2 :</u>



React Native Developer

Time taken by

QuickSort: 18207400 ns MergeSort: 23450700 ns SortedArray: 17372500 ns

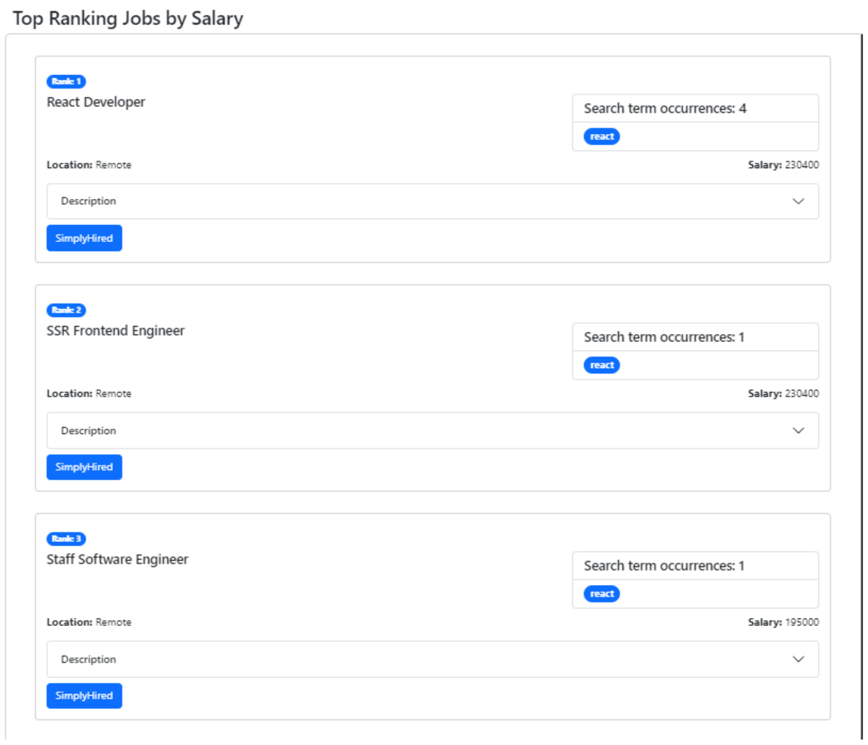
10. Recommend top paid jobs based on search results :

<u>Code Explanation :</u>

The searchInvertedIndexedDataBySalary method in the PageRanking.java class efficiently ranks jobs based on a combination of their maximum salary and word frequency. Leveraging a Trie data structure (TrieDS) for inverted indexing of job-related terms, the function iterates through the provided search terms, retrieves relevant information from the trie, and calculates the cost for each job. The cost is determined by adding the maximum salary and the word frequency associated with each job. Jobs are then inserted into a SortedArray based on their calculated cost, ensuring that the array remains sorted in descending order of cost. This approach enables the quick identification of top-ranking jobs, and in cases where multiple jobs have the same salary, they are further sorted based on their word frequency. The result is an efficient mechanism for users to explore and discover jobs that align with their preferences, particularly focusing on the combined factors of salary and word frequency.

Output 1 :

Top paid jobs for "React"



Top Ranking Jobs by Salary

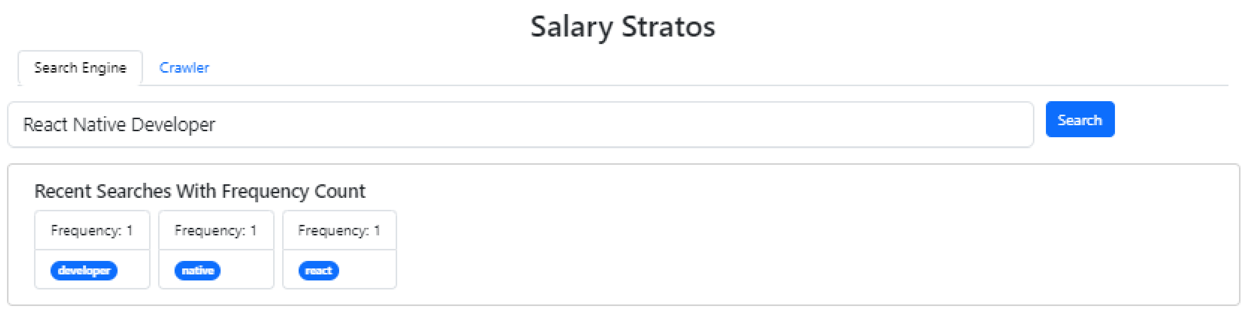
Rank: 1
React Developer
Search term occurrences: 4
react
Location: Remote
Salary: 230400
Description
SimplyHired

Rank: 2
SSR Frontend Engineer
Search term occurrences: 1
react
Location: Remote
Salary: 230400
Description
SimplyHired

Rank: 3
Staff Software Engineer
Search term occurrences: 1
react
Location: Remote
Salary: 195000
Description
SimplyHired

Output 2 :

Top paid jobs for "React Native Developer"



Salary Stratos

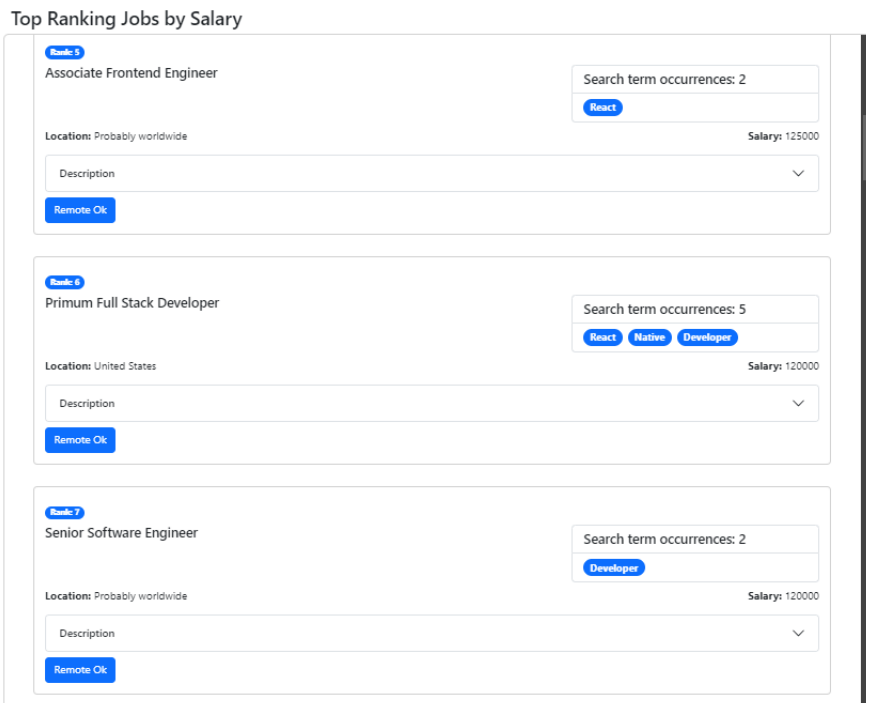
Search Engine | Crawler

React Native Developer

Search

Recent Searches With Frequency Count

Frequency: 1 — developer
Frequency: 1 — native
Frequency: 1 — react


Top Ranking Jobs by Salary

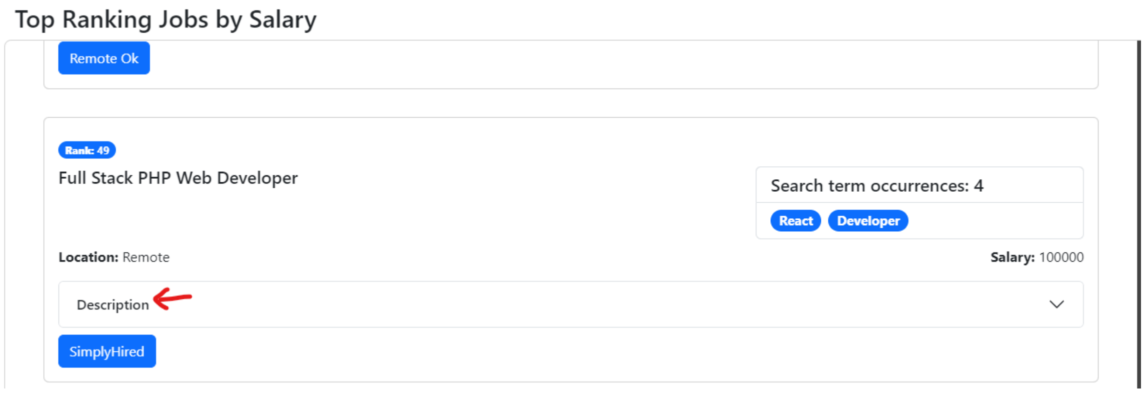
Rank: 5
Associate Frontend Engineer
Search term occurrences: 2
React
Location: Probably worldwide
Salary: 125000
Description
Remote Ok

Rank: 6
Primum Full Stack Developer
Search term occurrences: 5
React Native Developer
Location: United States
Salary: 120000
Description
Remote Ok

Rank: 7
Senior Software Engineer
Search term occurrences: 2
Developer
Location: Probably worldwide
Salary: 120000
Description
Remote Ok

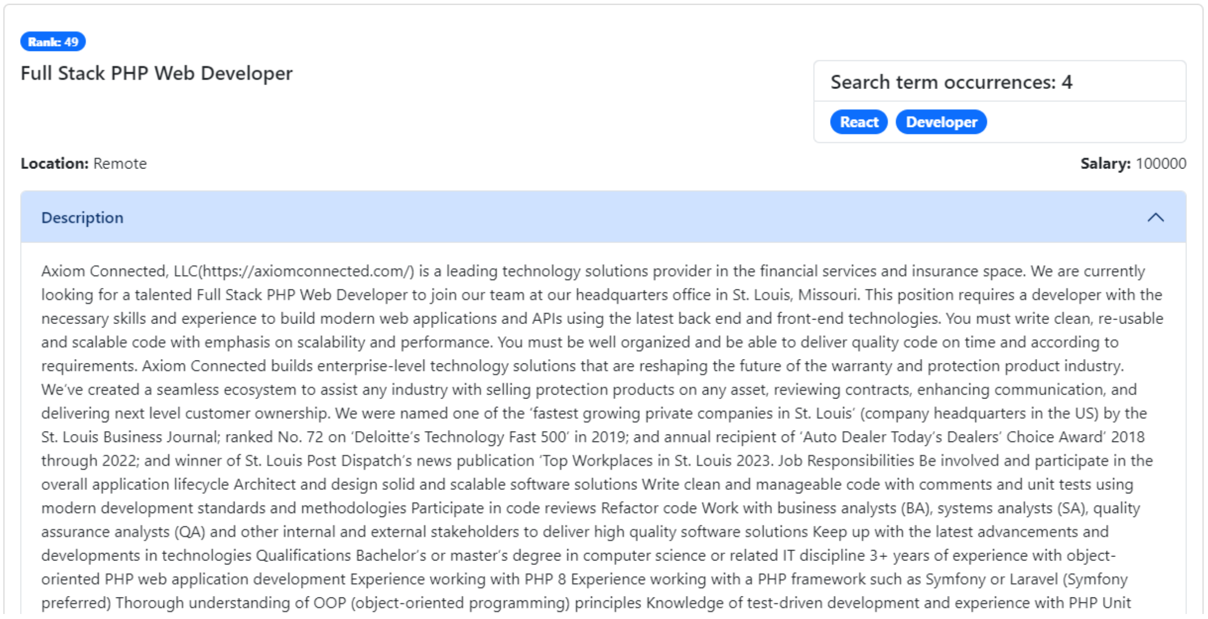
Output 3 :

We also show the description of the job along with the location, title, salary etc. When we click on the description it shows the job description.


Top Ranking Jobs by Salary

Remote Ok

Rank: 49
Full Stack PHP Web Developer
Search term occurrences: 4
React Developer
Location: Remote
Salary: 100000
Description
SimplyHired



Rank: 49

Full Stack PHP Web Developer

Search term occurrences: 4

React Developer

Location: Remote

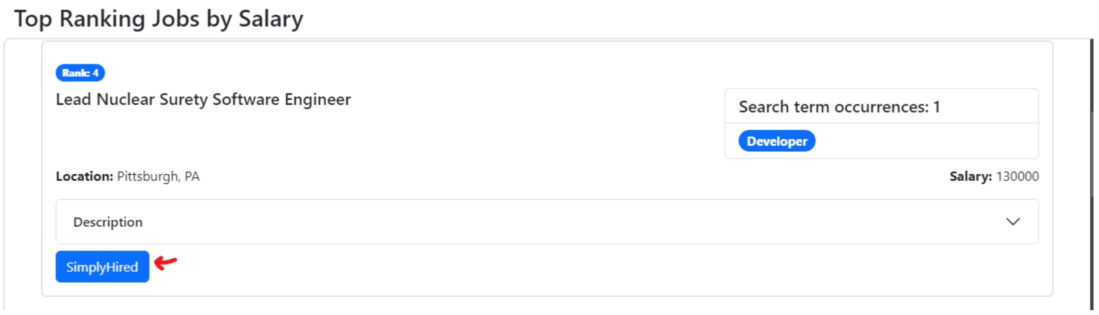
Salary: 100000

Description

Axiom Connected, LLC(https://axiomconnected.com/) is a leading technology solutions provider in the financial services and insurance space. We are currently looking for a talented Full Stack PHP Web Developer to join our team at our headquarters office in St. Louis, Missouri. This position requires a developer with the necessary skills and experience to build modern web applications and APIs using the latest back end and front-end technologies. You must write clean, re-usable and scalable code with emphasis on scalability and performance. You must be well organized and be able to deliver quality code on time and according to requirements. Axiom Connected builds enterprise-level technology solutions that are reshaping the future of the warranty and protection product industry. We've created a seamless ecosystem to assist any industry with selling protection products on any asset, reviewing contracts, enhancing communication, and delivering next level customer ownership. We were named one of the 'fastest growing private companies in St. Louis' (company headquarters in the US) by the St. Louis Business Journal; ranked No. 72 on 'Deloitte's Technology Fast 500' in 2019; and annual recipient of 'Auto Dealer Today's Dealers' Choice Award' 2018 through 2022; and winner of St. Louis Post Dispatch's news publication 'Top Workplaces in St. Louis 2023. Job Responsibilities Be involved and participate in the overall application lifecycle Architect and design solid and scalable software solutions Write clean and manageable code with comments and unit tests using modern development standards and methodologies Participate in code reviews Refactor code Work with business analysts (BA), systems analysts (SA), quality assurance analysts (QA) and other internal and external stakeholders to deliver high quality software solutions Keep up with the latest advancements and developments in technologies Qualifications Bachelor's or master's degree in computer science or related IT discipline 3+ years of experience with object-oriented PHP web application development Experience working with PHP 8 Experience working with a PHP framework such as Symfony or Laravel (Symfony preferred) Thorough understanding of OOP (object-oriented programming) principles Knowledge of test-driven development and experience with PHP Unit

<u>Output 4 :</u>

The job website name is a button which has the link to the job. When clicked on the job website name it takes to the job website where the job is posted.



Top Ranking Jobs by Salary

Rank: 4

Lead Nuclear Surety Software Engineer

Search term occurrences: 1

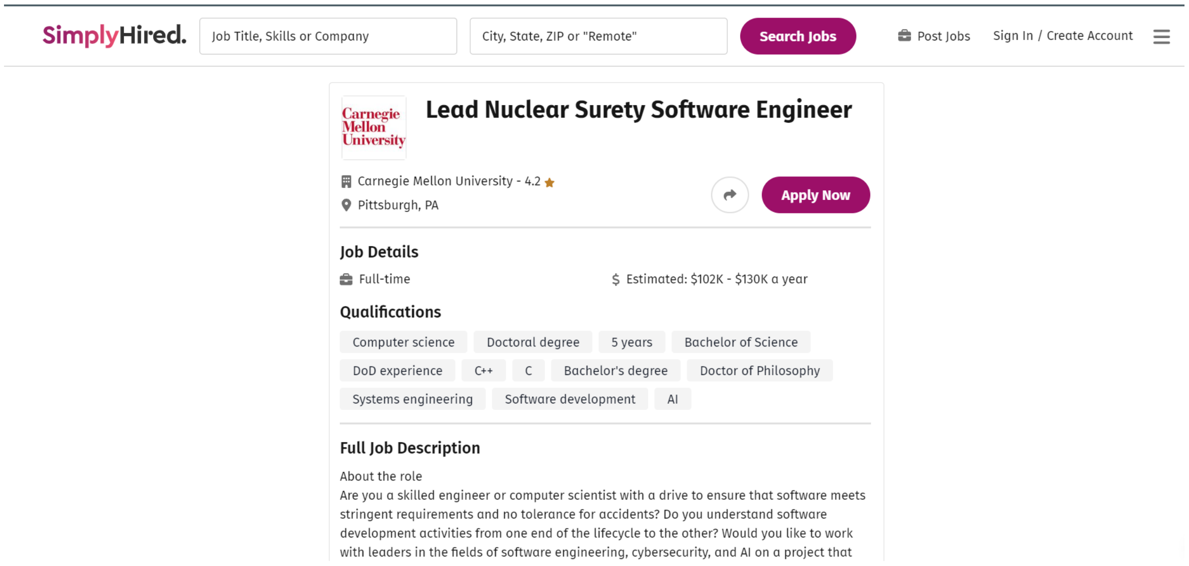
Developer

Location: Pittsburgh, PA

Salary: 130000

Description

SimplyHired



SimplyHired. Job Title, Skills or Company | City, State, ZIP or "Remote" | Search Jobs | Post Jobs | Sign In / Create Account

Lead Nuclear Surety Software Engineer

Carnegie Mellon University - 4.2 ★

Pittsburgh, PA

Apply Now

Job Details

Full-time

$ Estimated: $102K - $130K a year

Qualifications

Computer science | Doctoral degree | 5 years | Bachelor of Science | DoD experience | C++ | C | Bachelor's degree | Doctor of Philosophy | Systems engineering | Software development | AI

Full Job Description

About the role

Are you a skilled engineer or computer scientist with a drive to ensure that software meets stringent requirements and no tolerance for accidents? Do you understand software development activities from one end of the lifecycle to the other? Would you like to work with leaders in the fields of software engineering, cybersecurity, and AI on a project that