

Search Results Linkage Analyzer

Project (S.R.L.A)

DSAI 103 Data Acquisition

2023 Batch

Made By: Youssef Mohammad - 202300220

Under the Supervision of Prof. Mohammad Maher

Abstract:

What did we make?

The Search Results Linkage Analyzer (S.R.L.A) app retrieves results links with the results relationships and displays them in a dashboard. The app analyzes the results relationships when a user enters a query and returns all results and statistics into a Dashboard for ease of results analysis.

How Does the App works? :

The S.R.L.A. app focuses on identifying the most relevant search results for a user's query and presenting them through visual dashboards.

1. Data Acquisition and Initial Processing:

- The app starts by taking two inputs from the user: their search query and a Serp API key.
- It then leverages the API key to retrieve the organic search results for the given query.
- Each retrieved result comes with a snippet, a concise summary of the corresponding webpage.

2. Extracting Key Information and Assigning Relevance Scores:

- The app dives deeper into these snippets by extracting key statements from each one.
- It employs Term Frequency (TF) analysis to assess the relevance of each statement to the other results snippets statements. Essentially, it calculates how often terms from the statement of each snippet of each result appear within the total statements of total snippets of total results.
- To further refine relevance, the app considers the position of each result in the search rankings. Results appearing higher on the page are generally considered more relevant, so their TF-based relevance scores are adjusted accordingly.

3. Identifying Highly Relevant Results:

- To establish a baseline for relevance, the app calculates the average relevance score of all retrieved results.
- It then sets a threshold value that is half of this average.
- Only results with a relevance score exceeding this threshold are deemed highly relevant to the user's query.
- Titles of these highly relevant results are compiled into a specific data structure called a "node list."

4. Network Construction and Centrality Analysis:

- Since the focus is on highly relevant results, the app establishes connections (edges) between all the nodes in the node list. This creates a network where each node (a highly relevant result title) is interconnected with every other node.
- It's important to note that due to the interconnected nature of this network (all nodes are considered highly relevant), a specific network analysis metric called "betweenness centrality" will be zero for every node. Betweenness centrality typically measures the influence a node has in connecting other nodes within a network, but in this case, the complete interconnectedness renders this value irrelevant.
- Another network analysis metric, "degree centrality," is also calculated, but it might not provide much insight due to the same interconnectedness. Degree centrality simply counts the number of connections (edges) a node has, and in this case, all nodes will have the same high degree value.

5. Identifying Communities within Relevant Results:

- After calculating betweenness centrality, S.R.L.A. goes a step further to uncover potential thematic groupings within the highly relevant results. It achieves this by employing the Girvan-Newman algorithm.
- In essence, the Girvan Newman algorithm works by iteratively removing the edges with the highest "betweenness centrality" from the network. Betweenness centrality, though zero in this case due to the network structure, conceptually refers to how often a connection (edge) lies on the shortest paths between other nodes.
- By removing these high-betweenness edges, the algorithm progressively breaks down the network into smaller, more tightly knit clusters. These clusters represent groups of highly relevant results that likely share similar themes or topics.

6. Visualization with Heatmap and 3D Network Graph:

- To visually represent the pre-calculated relevance scores of the highly relevant results, the app creates a heatmap. The Y-axis of this heatmap corresponds to the Title of each result within the list (remember, only highly relevant results are included), and the X-axis represents the actual relevance score calculated earlier. By using colour intensity or shading, the heatmap effectively highlights results with higher relevance scores.
- Finally, the app constructs a 3D network graph to provide a spatial representation of the connections between these highly relevant results. It utilizes a function within the NetworkX library (a Python package for network analysis) called "random layout" to assign random X, Y, and Z coordinates to each node (result title). These coordinates essentially determine the placement of each result within the 3D space, and by plotting them with connections (edges), the app visualizes the network of highly relevant search results.

7. Enhanced Features of the Search Results Linkage Analyzer (S.R.L.A) App:

1. Knowledge Graph Scraping:

- The app scrapes the knowledge graph associated with the user's query, extracting rich data such as names, locations, entities, parents, awards, education, and more.
- This data is displayed on a separate page of the dashboard, providing users with a comprehensive overview of the query's context and related information.

2. Image Scraping and Display:

- The app employs an algorithm to scrape the first applicable image related to the user's query.
- The scraped image is displayed on the same page as the knowledge graph data, enhancing the visual appeal of the dashboard and providing users with a more immersive experience.

These enhancements significantly expand the capabilities of the S.R.L.A. app, making it a more versatile tool for users seeking to gain deeper insights from their search results.

Exporting for Visualization and Streamlit Dashboard:

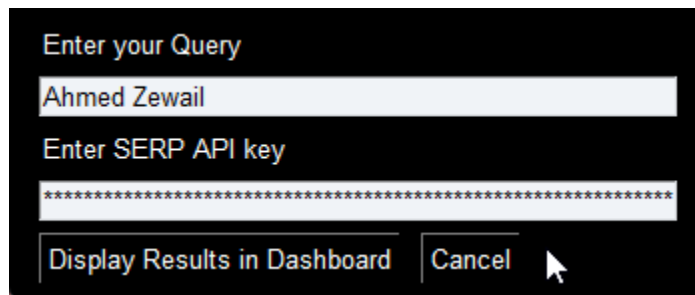
- After generating the different visualizations, S.R.L.A. ensures they can be displayed effectively within dashboards. To achieve this:
 - The network graph, 3D network graph, Clustered Community Graph by (Girvan Newman) and heatmap graph are all exported as PNG images. These image files are a common format for displaying visuals on web pages.
 - The app also exports crucial data into JSON files. These files include:
 - Relevancy values for each result.
 - Positions of the results in the search rankings.
 - Degree centrality values.
 - Betweenness centrality values (which will likely be all zeros due to the network structure).
 - Links of Most Relevant Results (separately)
 - All Links of Results
- Finally, all the generated files (PNG images and JSON data) are imported into a Streamlit dashboard. Streamlit is a Python framework for creating web apps. Within the dashboard:
 - The JSON data is converted into Pandas DataFrames, a popular data structure in Python for data analysis.
 - Streamlit then utilizes these DataFrames to display the data (relevancy values, positions, etc.) as bar charts or line charts,

allowing for easy comparison of results and the links are displayed in tables.

- The imported PNG images are also incorporated into the dashboard using Streamlit's image display functionalities.
- To ensure the images render correctly within the dashboard, Streamlit employs a conversion from BGR to RGB colour format.

Step By Step Guide to display how the S.R.L.A Works:

- 1.Run The S.R.L.A App, Enter your Query, Serp Api Key and
Click Display Results in Dashboard:

A screenshot of a web application interface with a black background. It features two text input fields. The first field is labeled 'Enter your Query' and contains the text 'Ahmed Zewail'. The second field is labeled 'Enter SERP API key' and contains a series of asterisks. Below the input fields are two buttons: 'Display Results in Dashboard' and 'Cancel'. A mouse cursor is pointing at the 'Cancel' button.

Enter your Query

Ahmed Zewail

Enter SERP API key

Display Results in Dashboard Cancel

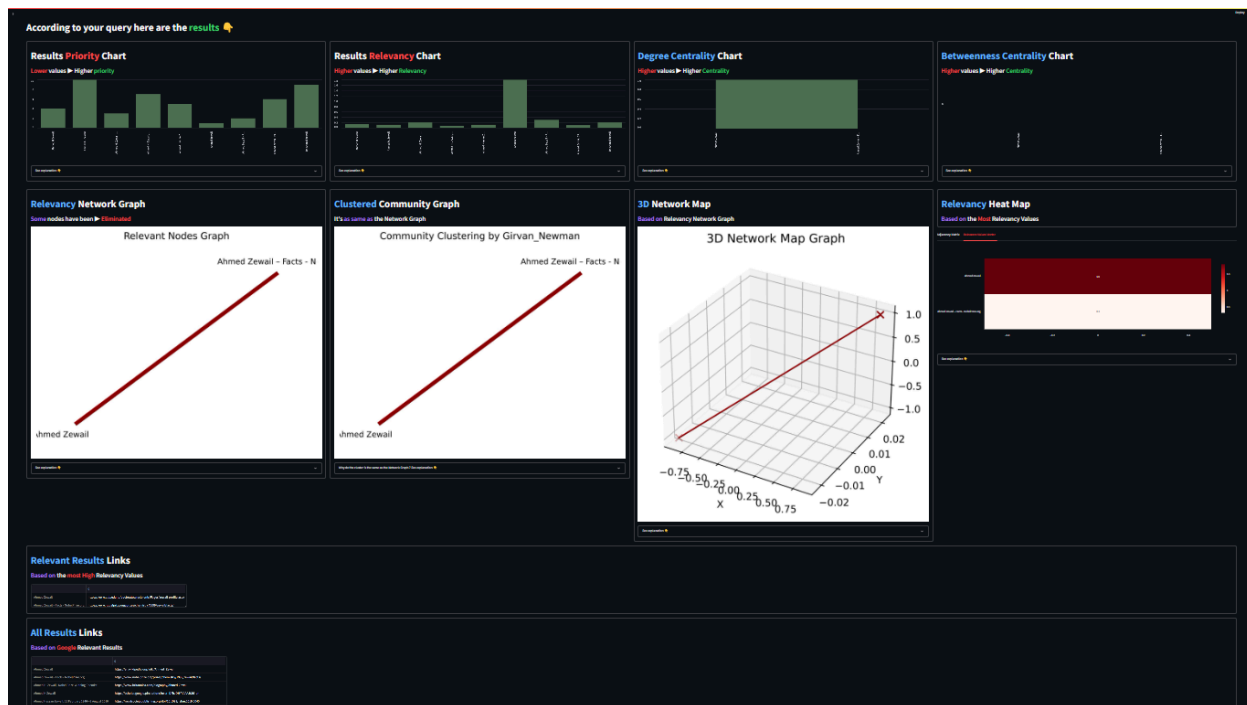
2. Wait For less than a minute until this message appears to
you:

You can now view your Streamlit app in your browser.

Local URL: <http://localhost:8501>

Network URL: <http://192.168.1.95:8501>

3. The DashBoard will appear automatically in your default browser: (Note that these dims of the Dashboard webpage were set to least to capture the whole possible Area of the Dashboard)



Welcome to **Search-Results-linkage-Analyzer** 📈 (Don't forget to reach out to us 😊)

Some Additional Data About Your Query 📈

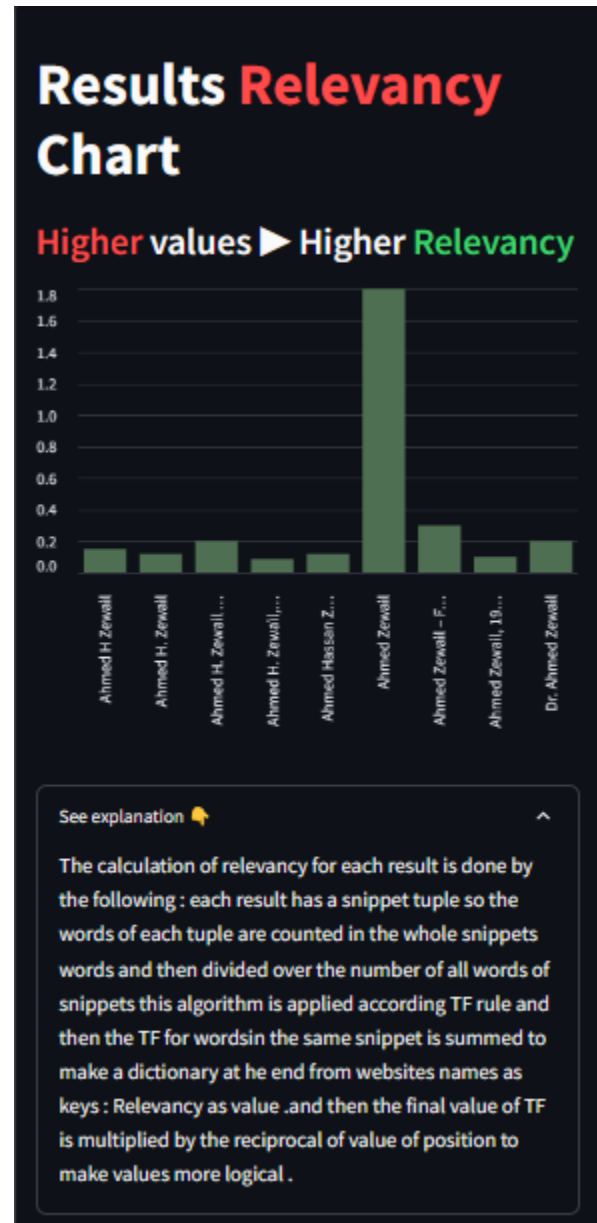
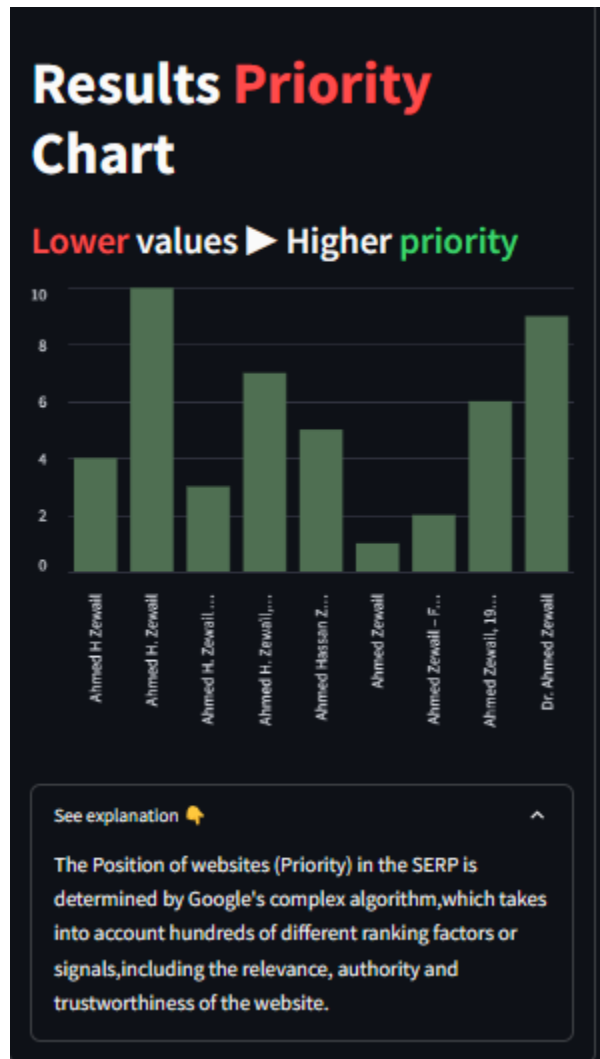
Additional Data About Your Query

If you Didn't find any Info here Try to scrap it by yourself through the provided links in the another page

| | |
|-----------------------------|---|
| Name | Ahmed Zewail |
| Entity Type | people, scholars, people |
| Description | Ahmed Hassan Zewail was an Egyptian and American chemist, known as the "father of femtochemistry". He was awarded the 1999 Nobel Prize in Chemistry for his work on femtochemistry and became the first Egyptian and Arab to win a Nobel Prize in a scientific field, and the second African to win a Nobel Prize in Chemistry. |
| Dead | August 3, 2018 (age 70 years), Pasadena, CA |
| Education | University of Pennsylvania (1969–1973), Alexandria University (1965), Alexandria University (1967) |
| Parents | Bawika Say, Hassan Zewail |
| Founder | Not Available |
| President | Not Available |
| Location | Not Available |
| Establishment | Not Available |
| Known For | Femtochemistry |
| Born | February 26, 1946, Damanshour, Egypt |
| Spouse | Derna Fakham (m. 1989–2016) |
| Awards | Nobel Prize in Chemistry, Order of the Nile, MORE |
| Knowledge Graph Search Link | https://www.google.com/search?igmid=ny0ayhmA&icon-US&q=Ahmed+Zewail&gs=ls396131a1061d9d&skind=17&source=sh/v/loop/m/1/ |
| Hood Quarters | Not Available |
| Sales | Not Available |
| Stock Price | Not Available |
| Subsidiaries | Not Available |
| Executives | Not Available |

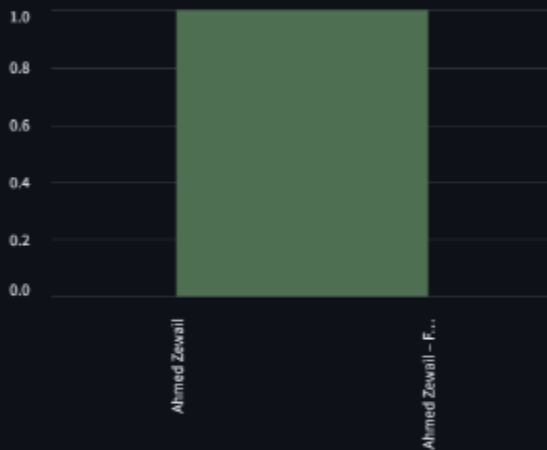
4. A Quick view for each data container on the DashBoard

according to users' query = "Ahmed Zewail":



Degree Centrality Chart

Higher values ► Higher Centrality

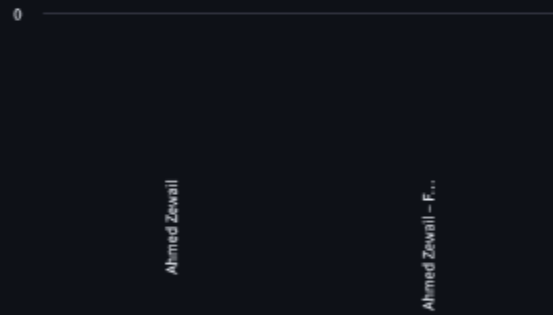


See explanation 📌

The Degree Centrality of each website is calculated by counting the number of connections each website has in the network map. Also you should realize that they are all equal as the threshold have removed some lower relative results and all websites are relevant to each other.

Betweenness Centrality Chart

Higher values ► Higher Centrality

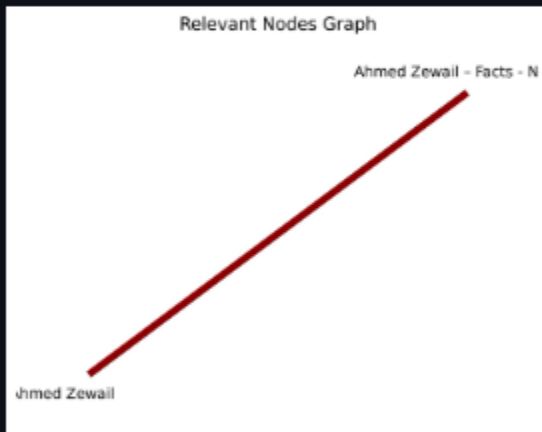


See explanation 📌

Betweenness centrality measures a node's importance in a network by counting how many shortest paths between other nodes pass through it. Imagine traffic flowing through a city - a central bridge would have high betweenness centrality because many routes use it.

Relevancy Network Graph

Some nodes have been ►
Eliminated

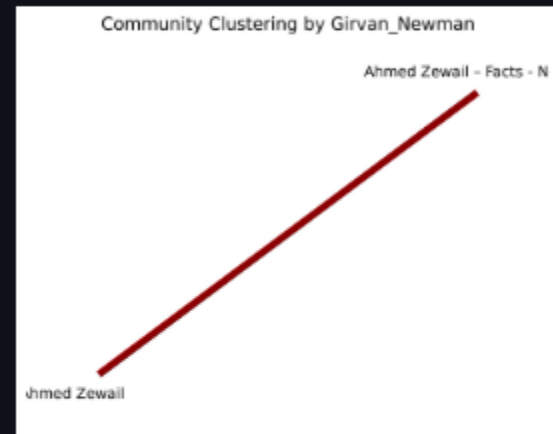


See explanation 🗨

The Network Graph is an interactive graph that shows the relationships between the websites and some websites are eliminated by comparing the value of each node with a threshold values.

Clustered Community Graph

It's as same as the Network Graph



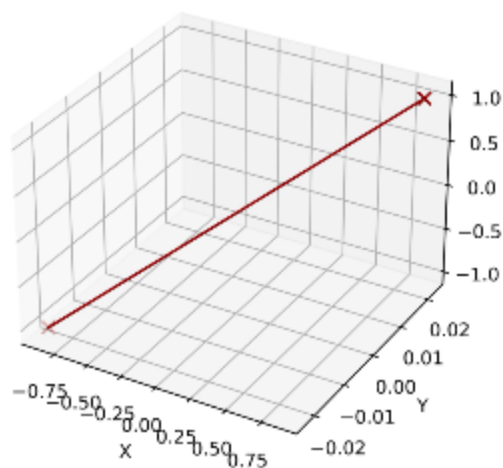
Why do the cluster is the same as the Network Graph ? See explanation 🗨

How Does Girvan Newman work ? The idea was to find which edges in a network occur most frequently between other pairs of nodes by finding edges betweenness centrality. The edges joining communities are then expected to have a high edge betweenness. The underlying community structure of the network will be much more fine-grained once the edges with the highest betweenness are eliminated which means that communities will be much easier to spot.

3D Network Map

Based on Relevancy Network Graph

3D Network Map Graph



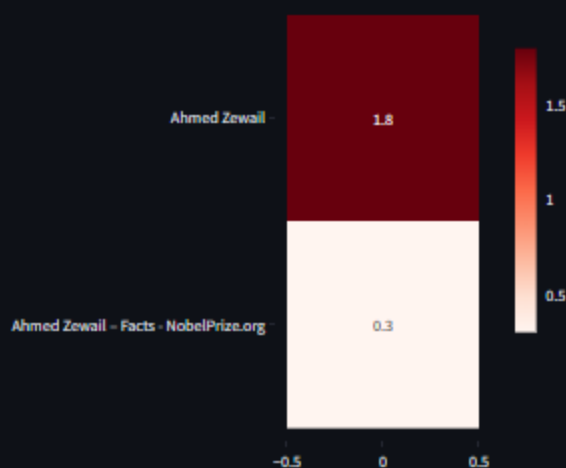
See explanation 📖

For every node, a position is generated randomly by choosing each of dim coordinates uniformly at random on the interval $[0.0, 1.0]$.

Relevancy Heat Map

Based on the Most Relevancy Values

Adjacency Matrix Relevancy Values Vector



See explanation 📖


The Relevancy Heat Map shows the most relevant results according to the other results. You will realize that as you are reaching the greater value of relevancy the color of the result in heat map will be more denser .

Results Relations Of Your Query
Additional Data About User Query

Additional Data About Your Query

If you Didn't find any Info here Try to scrap it by yourself through the provided links in the another page

| | |
|-----------------------------|---|
| | 0 |
| Name | Ahmed Zewail |
| Entity Type | people, scholars, people |
| Description | Ahmed Hassan Zewail was an Egyptian and American chemist, known as the "father of femtochemistry". He was awarded the 1999 Nobel Prize in Chemistry for his work on femtochemistry and became the first Egyptian and Arab to win a Nobel Prize in a scientific field, and the second African to win a Nobel Prize in Chemistry. |
| Died | August 2, 2016 (age 70 years), Pasadena, CA |
| Education | University of Pennsylvania (1969–1973), Alexandria University (1969), Alexandria University (1967) |
| Parents | Rawhia Dar, Hassan Zewail |
| Founder | Not Available |
| President | Not Available |
| Location | Not Available |
| Establishment Date | Not Available |
| Known For | Femtochemistry |
| Born | February 26, 1946, Damanhour, Egypt |
| Spouse | Dema Faham (m. 1989–2016) |
| Awards | Nobel Prize in Chemistry, Order of the Nile, MORE |
| Knowledge Graph Search Link | https://www.google.com/search?kgmid=/m/0syhm&hl=en-US&q=Ahmed+Zewail&kgm=b39b333a10e61d9d&shndl=17&source=sh/x/kp/osrp/m5/1 |



5. You Can customize your experience by choosing Either a light or dark mood of the dashboard from the settings:

Settings

Development

☒ Run on save
Automatically updates the app when the underlying code is updated.

Appearance

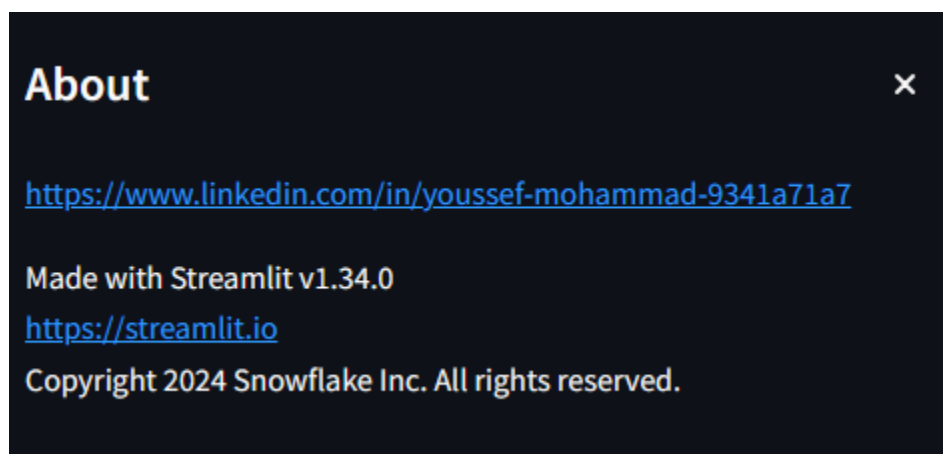
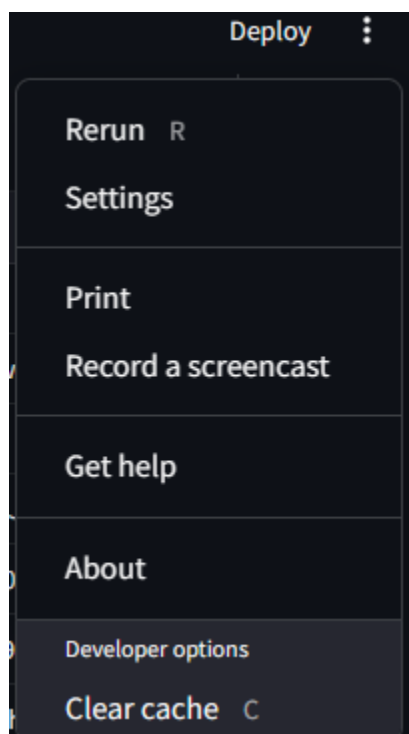
☒ Wide mode
Turn on to make this app occupy the entire width of the screen

Choose app theme, colors and fonts

Use system setting

Edit active theme

7. It's Appreciated if to keep in contact with the developer
for any Inquiries by browsing the “About” section or “Get
Help” section:



For Further experimenting You can download the app from the link below:

[S.R.L.A. app](#)

*Note that The APP may need libraries for functionality.

*Note that not all queries may get accurate results which may disturb app functionality.