

Part 4: User Interface and Web Analytics

1. User Interface

In the development of our search engine's user interface, we undertook several crucial steps to ensure a seamless and intuitive user experience:

1.1 Data Preparation

- Meticulously read and adapted the load corpus file to read the .json file about the Ukrainian-Russian war, ensuring a seamless integration into our code. To do this, we read and made sure we understood the code of the skeleton and then merged it with the knowledge we gained from the first lab, where we also had to read this file with a required format. Here, we save each tweet as a document object to be used on the search engine to generate results once a query is given.

1.2 Algorithm Integration

The previously developed algorithms, those used in Part 3, have been harmoniously integrated into our search engine, carefully tuned for compatibility.

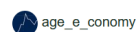
- The TF-IDF (Term Frequency-Inverse Document Frequency) algorithm is a numerical representation technique in natural language processing that we use. It measures the importance of a word in a tweet relative to the collection of tweets. We calculated a weight for each word based on its frequency in the tweet (TF) and its rarity across all tweets (IDF). High TF-IDF scores indicate words that are specific to a document, which helps us in information retrieval and text analysis by highlighting key terms while reducing common words. Following the previous score, we applied cosine similarity to generate the most similar results to the user's query.
- Apart from this, we implemented our algorithm, which combines the essence of the TF-IDF algorithm and adds some weight to the likes and retweets in the score.

To do this, we took the algorithm we implemented in previous parts of the project and adapted it to the structure of the skeleton.

1.3 Results Presentation:

- Engineered a visually refined presentation of search results, ordered intelligently based on algorithmic scores. Drawing inspiration from the streamlined design principles observed in the results page of search engines like Google.
- Each result is thoughtfully composed, featuring user image, names, a clickable title leading to the complete tweet, the publish date of the tweet, along with a concise description and date.

Found 511 results...



age_economy

Lithuania, we will win this war! #UkraineRussiaWar #ukrainecounteroffensive #UkraineWillWin #poland...
[doc_details?id=1575810627448475651&search_id=72737¶m2=2](#)

Fri Sep 30 11:31:45 +0000 2022 — Lithuania, we will win this war! #UkraineRussiaWar
#ukrainecounteroffensive #UkraineWillWin #poland #Lithuania #war <https://t.co/wGWIMHvcid>



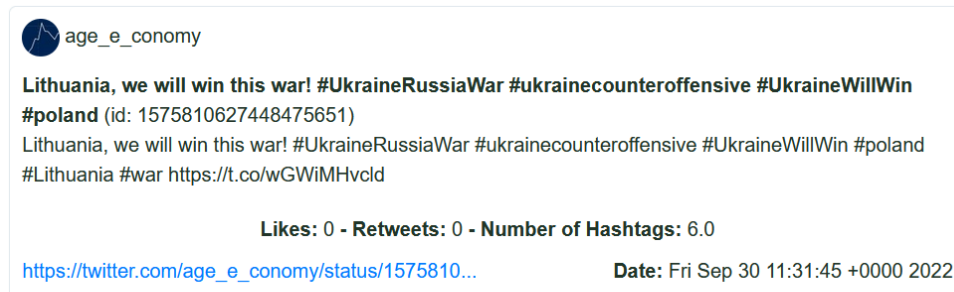
LittleLeighXoxo

The war crime, for The Holocaust was The Nuremberg trials. (Which is a small part of this hearing.)...
[doc_details?id=1575157054938955777&search_id=72737¶m2=2](#)

Wed Sep 28 16:14:41 +0000 2022 — The war crime, for The Holocaust was The Nuremberg trials. (Which is a small part of this hearing.) It took place on 11/20/1945 to 10/1/1946. Now #Putin should be indicated & prosecuted as a war criminal for his war crimes. (#Ukraine, #UkraineRussiaWar) <https://t.co/SLyuRZukof>

1.4 Tweet Display

Increased user interaction by implementing a feature that allows users to view the full tweet by simply clicking on its title. We included all the information we considered relevant to the tweet, such as the number of likes, retweets or hashtags, as well as the URL and the date the tweet was published.



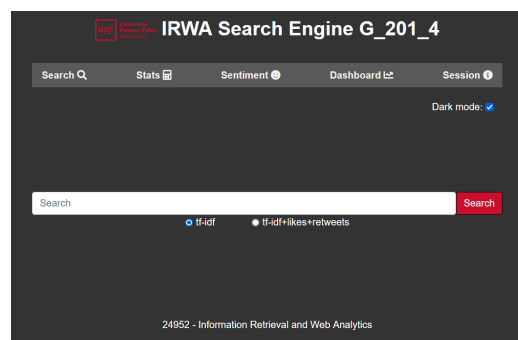
1.5 Navigation Bar

Designed an elegant and functional navigation bar to ensure a seamless and intuitive journey for users exploring the application, allowing you to navigate between the search page, tweet statistics, sentiment analysis, dashboard and your session statistics. All of these are explained in more detail later in this document.



1.6 Additional Features:

- Empowered users with easy access to session information and emotional analysis directly from the navigation bar.
- Strategically implemented a performance boost by persistently storing the generated index in a .pkl file after the initial run, substantially enhancing subsequent load times.
- We added radio buttons to allow the user select which algorithm want to use to obtain his results.
- Enhanced user customization through the thoughtful implementation of a Light/Dark mode.



2. Web Analytics

Our web analytics implementation encompasses a diverse set of features and insightful visualizations to unravel the intricacies of user behavior:

2.1 Stats

- Dynamically showcased the tweets that resonated with users, offering a tangible snapshot of their interaction.

Clicked docs:

(2 visits) — id: 1575810627448475651 — Lithuania, we will win this war! #UkraineRussiaWar #ukrainecounteroffensive #UkraineWillWin #poland #Lithuania #war <https://t.co/wGWMHwld>

(2 visits) — id: 1575157054938955777 — The war crime, for The Holocaust was The Nuremberg trials. (Which is a small part of this hearing.) It took place on 11/20/1945 to 10/1/1946. Now #Putin should be indicated & prosecuted as a war criminal for his war crimes. (#Ukraine, #UkraineRussiaWar) <https://t.co/SLyuRZukof>

(1 visits) — id: 1575916461620690977 — 🇪🇺 Berlin is Ukraine Brussels is Ukraine London is Ukraine Stockholm is Ukraine Paris is Ukraine Europe is Ukraine Just like Kyiv is #Ukraine and Ukraine is Ukraine This is a continental threat, let's face it <https://t.co/aRQwsKprwd> #WARINUKRAINE #UKRAINEWAR #UKRAINERUSSIAWAR <https://t.co/OAAYnNDMDm> <https://t.co/ttsYk0eGhx>

- Introduced an intuitive dashboard that visually portrays the frequency of clicks for each document, delivering a user-friendly analytical experience.

Ranking of Visited Documents



- Featured an aesthetically pleasing word cloud, offering a visual representation of the queries searched. Take into account that only is available if the user has engaged in a search.

WordCloud

war war2
ukraine

2.2 Sentiment Analysis

- Introduced a novel feature empowering users to subject any phrase or word to emotion analysis, fostering a deeper understanding of user sentiments.
- Visualized user-clicked tweets, enabling users to copy tweets and explore the predicted emotional nuances.

NLTK Sentiments

Type a sentence, click on the submit button and wait for your prediction.

Clicked docs:

(Here you have the clicked docs, you can copy them to see which is the sentiment of the tweet)

(2 visits) — id: 1575810627448475651 — Lithuania, we will win this war! #UkraineRussiaWar #ukrainecounteroffensive #UkraineWillWin #poland #Lithuania #war <https://t.co/wGWMHwclD>


(2 visits) — id: 1575157054938955777 — The war crime, for The Holocaust was The Nuremberg trials. (Which is a small part of this hearing.) It took place on 11/20/1945 to 10/1/1946. Now #Putin should be indicated & prosecuted as a war criminal for his war crimes. (#Ukraine, #UkraineRussiaWar) <https://t.co/SLyuRZukof>

(1 visits) — id: 1575916461620690977 — 🇺🇦 Berlin is Ukraine Brussels is Ukraine London is Ukraine Stockholm is Ukraine Paris is Ukraine Europe is Ukraine Just like Kyiv is #Ukraine and Ukraine is Ukraine This is a continental threat, let's face it <https://t.co/aRQwsKprwd> #WARINUKRAINE #UKRAINEWAR #UKRAINERUSSIAWAR <https://t.co/OAAYnNDMDm> <https://t.co/ttsYk0eGhx>

24952 - Information Retrieval and Web Analytics


NLTK Sentiments

Type a sentence, click on the submit button and wait for your prediction.



NLTK Sentiments

Type a sentence, click on the submit button and wait for your prediction.



Clicked docs:

(Here you have the clicked docs, you can copy them to see which is the sentiment of the tweet)

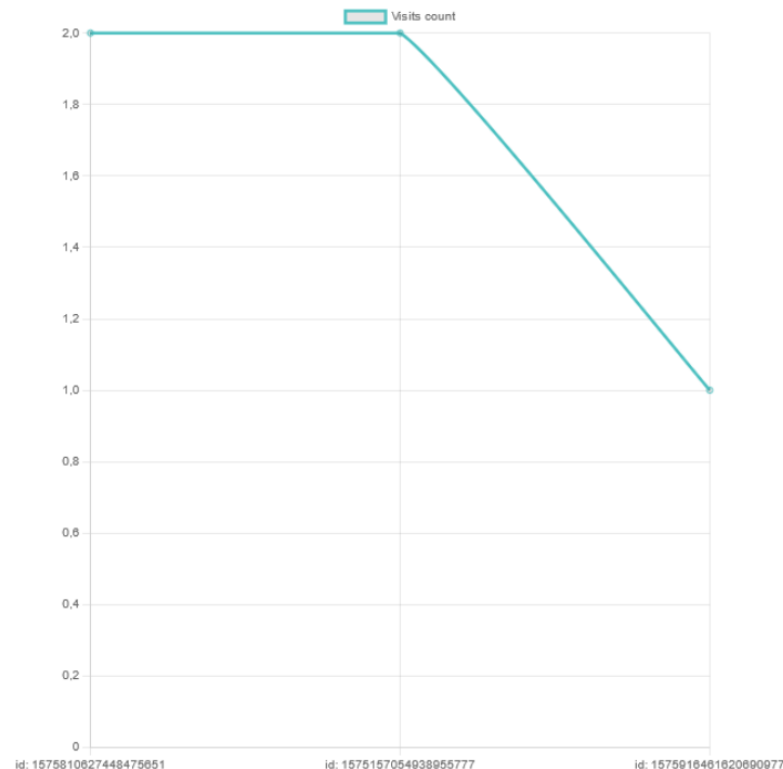
(1 visits) — id: 157575699941586945 — I happy to #Ukraine defense well against Putin criminal mind

- Hope positive as #Ukrainian will win soon & will success to capture all lands from #RussianArmy
- Support #humanity FIRST THEN POLITICS <https://t.co/AaZh8UDFdn> #UkraineRussiaWar #UkraineWar #UkrainianArmy #win #EUA

2.3 Dashboard

- Crafted an engaging graphical representation illustrating the frequency of user-clicked tweets, providing a nuanced perspective on user preferences.

Ranking of Visited Documents



- Thoughtfully presented a comprehensive overview of user interactions by displaying the tweets clicked during their engagement.

Tweets visited:

(2 visits)— id: 1575810627448475651 — Lithuania, we will win this war! #UkraineRussiaWar
#ukrainecounteroffensive #UkraineWillWin #poland #Lithuania #war <https://t.co/wGWiMHvclD>

(2 visits)— id: 1575157054938955777 — The war crime, for The Holocaust was The Nuremberg trials. (Which is a small part of this hearing.) It took place on 11/20/1945 to 10/1/1946. Now #Putin should be indicated & prosecuted as a war criminal for his war crimes. (#Ukraine, #UkraineRussiaWar) <https://t.co/SLyuRZukof>

(1 visits)— id: 1575916461620690977 — 🇺🇦 Berlin is Ukraine Brussels is Ukraine London is Ukraine Stockholm is Ukraine Paris is Ukraine Europe is Ukraine Just like Kyiv is #Ukraine and Ukraine is Ukraine This is a continental threat, let's face it <https://t.co/aRQwsKprwd> #WARINUKRAINE #UKRAINEWAR #UKRAINERUSSIAWAR <https://t.co/OAAYnNDMDm> <https://t.co/ttsYk0eGhx>

2.4 Session information

- Furnished users with essential session details, including IP, accessing platform, and browser information.
- Thoughtfully displayed the duration of the user's ongoing session in minutes, providing real-time insights into user engagement.

Session

IP: 127.0.0.1 — **Platform:** Windows 10 — **Browser:** Firefox

Session Time: 25 min

- Offered a chronological list of sessions, showcasing IP, operating system, and browser details, sequenced from the most recent to the oldest. The list of sessions is saved in a text document called *sessions.txt*.

All Sessions

(your actual session is the first one)

- **IP:** 127.0.0.1 — **Platform:** Windows 10 — **Browser:** Firefox,
- **IP:** 127.0.0.1 — **Platform:** Windows 10 — **Browser:** Firefox,
- **IP:** 127.0.0.1 — **Platform:** Windows 10 — **Browser:** Chrome,
- **IP:** 127.0.0.1 — **Platform:** Windows 10 — **Browser:** Firefox,
- **IP:** 127.0.0.1 — **Platform:** Windows 10 — **Browser:** Firefox,
- **IP:** 127.0.0.1 — **Platform:** Windows 10 — **Browser:** Firefox,
- **IP:** 127.0.0.1 — **Platform:** Windows 10 — **Browser:** Firefox,
- **IP:** 7.0.0.1 — **Platform:** Windows 10 — **Browser:** Firefox.

- Enriched the user experience by displaying a list of searched queries, complemented by a visually compelling word cloud and histogram depicting query frequency.

Queries Done:

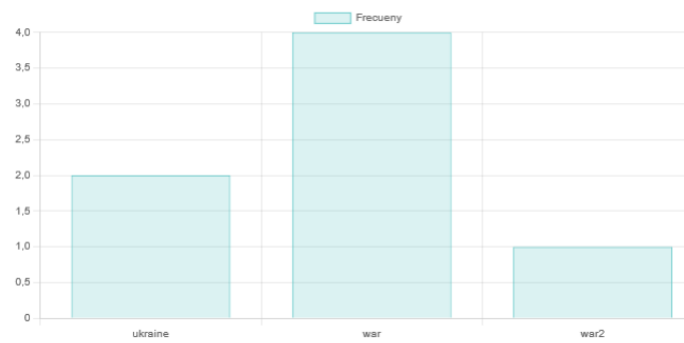
ukraine, war, war, war2, ukraine, war, war.

WordCloud



A word cloud visualization where the words 'ukraine', 'war2', and 'war' are displayed in different colors (yellow, blue, and green respectively) and sizes, indicating their relative frequency in the search queries.

Frequency of the queries



3. Future Works

The following outlines potential avenues for future development, promising to elevate both the user interface and web analytics capabilities of our search engine.

- **Enhanced User Interface:** Implement user feedback mechanisms to continuously refine and improve the user interface based on user interactions and preferences.
- **Real-Time Updates:** Explore the feasibility of implementing real-time updates to keep users informed of the latest relevant information.
- **Machine Learning Integration:** Investigate the integration of machine learning models to enhance search result relevance and improve the overall user experience.
- **Extended Analytics:** Expand the scope of web analytics by incorporating additional metrics, such as click-through rates, popular search queries, and user demographics.
- **Enhanced Sentiment Analysis:** Enhance the sentiment analysis feature by incorporating advanced natural language processing techniques for more accurate emotion prediction.