



Bingle — critical search engine on **Ukraine**

Zhendong Yin, Yifan Zhu



Purpose



1. Focus on Ukraine, search with time span.
2. Show different positions and angles for the related queries (determined by their languages and the article publishers).
3. Show sentiment analysis on each documents
4. Enable comment function and let comments to influence the articles ranking



Method

- Dataset
- Database
- Functions of the websites



Corpus acquisition and processing



A short horizontal bar with a teal left half and an orange right half.

Corpus acquisition

- Time span: 02/22/2022 - the start date of Ukraine - Russian war till today
- Size: 13,670 news
- Languages:
 - English -> most widely used language
 - Russian -> one of the main parties of the war
 - Chinese (simplified, traditional)-> most used in population
 - French -> attitudes from European

A short horizontal bar with a teal left half and an orange right half, located above the title.

Raw data collected in file - Data format

- headline of the news
- author
- news source
- Date
- Article text
- Language

Raw data collected in file



title	author	source	time	content	language
【不按國際劇本走的烏克蘭總統】8年前的克里米亞戰...	上報 UP Media	Buzzorange.com	2022-03-02 00:00:00	◇ 24 Volodymyr Zelensky...	zh
Guerre en Ukraine : à Irpin, près de Kiev, la bataille...	NULL	Francetvinfo.fr	2022-03-05 00:00:00	◇ La capitale ukrainienne constitue l'un des principau...	fr
Varsovie signe l'achat de 250 chars de combat amé...	NULL	BFMTV	2022-04-05 00:00:00	◇ La Pologne a signé un contrat de 4,3 milliards d'eur...	fr
【IFCN TALK 2022】 來自戰地的聲音 烏克蘭查核組...	台灣事實查核中心	Yahoo Entertainment	2022-04-06 00:00:00	◇ 4174300080...	zh
Хакеры взломали сайты ФСИН, Роскомнадзора,...	gagadget.com	Gagadget.com	2022-03-09 00:00:00	◇ , , , . Telegram- IT ARMY of Ukraine....	ru
Tour news: Battles, Fly Anakin, Caspian, Daniel Rom...	BrooklynVegan Staff	Brooklyn Vegan	2022-03-11 00:00:00	◇ Here's a roundup of recent tour news. Check the To...	en
Guerre en Ukraine : la détresse des faubourgs de Ki...	NULL	Francetvinfo.fr	2022-03-13 00:00:00	◇ Qu'importe les risques, Brovary (Ukraine), Sacha Se...	fr
IT-армия Украины создала онлайн-карту с адрес...	Максим Григорьев	ltc.ua	2022-04-14 00:00:00	◇ . IT- , ,	ru
В Варшаве назвали главную цель визита президе...	Григорий Воронцов	Life.ru	2022-04-14 00:00:00	◇ Getty Images / STATE EMERGENCY SERVICE OF UK...	ru
На Украине заявили о желании Киева обменять...	Варвара Кошечкина	Lenta	2022-04-14 00:00:00	◇ : Press service of State Security Service of Ukraine...	ru
Ukraine says street battles ongoing in Mariupol	Reuters	FX Empire	2022-04-15 00:00:00	◇ (Reuters) Ukraines defence ministry spokesman Ole...	en
Zelensky: "The battle for Donbas is very important"	Ivana Saric	Axios	2022-04-17 00:00:00	◇ Ukrainian President Volodymyr Zelensky stressed th...	en
Ukraine: Frappe de missile à Brovary, près de Kyiv,...	Reuters	Challenges	2022-04-17 00:00:00	◇ (Reuters) - Une frappe de missile aux premières he...	fr
Battle Rages in Mariupol's Streets	Bob Cronin	Newser	2022-03-19 00:00:00	◇ (Newser)...	en
頓巴斯戰役開打！美智庫分析：俄軍不太可能比先前...	自由時報	Ltn.com.tw	2022-04-19 00:00:00	◇ ISW...	zh
Ukraine : la bataille pour le Donbass a démarré	NULL	Euronews	2022-04-19 00:00:00	◇ La Russie a annoncé mardi matin avoir mené durant...	fr
極音速武器首戰，俄軍發射匕首飛彈攻擊烏克蘭彈藥庫	Alan Chen	Technews.tw	2022-03-20 00:00:00	◇ Destruction of a weapons depot of the Armed Force...	zh
馬斯克低軌衛星「星鏈」助攻，烏克蘭發動無人機之...	中央社	Buzzorange.com	2022-03-21 00:00:00	◇ Starlink...	zh



Corpus processing

Text preprocessing: normal text preprocessing: stem, small letter, stopwords...

Translation - google translation API



Translation AI - google translation



- Data: Russian, French, Chinese(simplified, traditional) and English
- Translate **headlines and news content** into English



Database design and indexing



We used three databases.

1. PostgreSQL
2. Redis
3. Elastic search





Database design - PostgreSQL

We stored the original data with this database.

We design the database with the following 5 tables:

- **User Table:** Name, Email, Password -> for registration and comments function
- **News Table:** Title, Author, Media Source, Publish Time, Content, ElasticSearch_ID, Sentiment, Language -> most important, store original news article
- **News Translation Table:** Title, Content -> translated data in this table
- **Comments Table:** News_ID, User_ID, Content, Publish Time -> store user's comments
- **Timeline Table:** Date, Event -> store all the events of Ukraine - Russian war according to time sequence.



PostgreSQL



Database design and indexing - Redis



We use it to map query and search results, cache the results from elasticsearch to improve search efficiency

- Redis is a in-memory Key-Value Database.
 - Format of key: " keyword-start_date-end_date-media_source" -because we have 4 search options in search bar
 - Transform the 4 search options into a string
- Every search result will be stored in Redis for 24 hours.

Using redis, there is no need to search again if the query has showed up before.

-> to boost the search speed. It could reduce the search time even many people are using the engine at time.





Database design and indexing - ElasticSearch

We use elasticsearch for search.

With the following attributes, we built data.

ID(as news id in postgresql), Title(as news title), Author, Source(media source), Time(published date), Content(article itself), Title_Translation(translation of the title), Content_Translation(translation of the article), Stemmer content(stemmed article), Comments Number(number of comments), Sentiment score(score of sentiment analysis)



elasticsearch



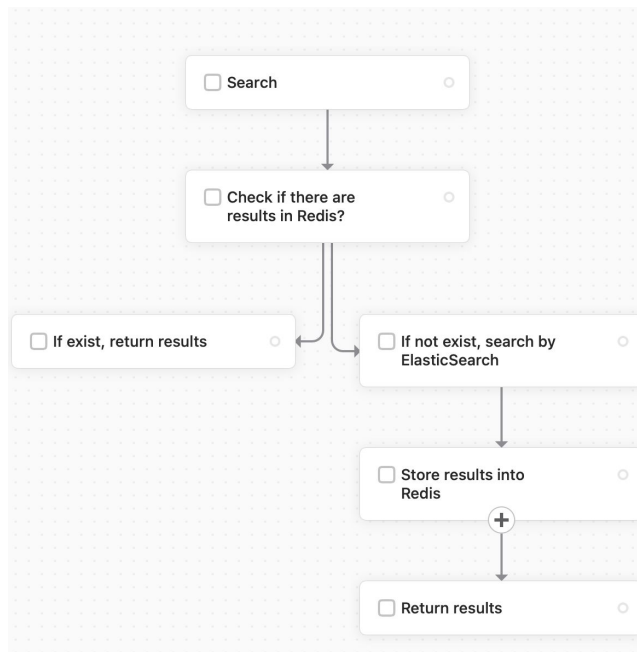
search flow with redis and elasticsearch



elasticsearch



redis



Runtime function



- Index page
- Sentiment analysis
- Comments





Bingle

mm/dd/yyyy

mm/dd/yyyy

MEDIA

SEARCH

[HOME](#) [TIMELINE](#)

BEMTV

April 05, 2022 9

Varsovie signe l'achat de 250 chars de combat améri...

La Pologne a signé un contrat de 4,3 milliards d'euros pour 250 chars de combat américains Abrams avec "une quantité sérieuse de munitions". Le minist...

Lastest news

Korrespondent.net

May 03, 2022 Q3

В бою за Харьков захватчики убили призера "Игр Непокоренных"

-, 2, " " . 2 . Warrior Games: Team Ukraine. " , Warrior Games . . . , , , " - . . " . 2021 " " . 25 2021 . .net
Telegram. <https://t...>

Fox News

📅 April 21, 2022 🗨️ 7

Mariupol battle rages with thousands injured, Ukraine official calls it 'key moment' in war



Index page

- The index page contains the search bar, time selection and media selection area.
- It also contains latest news area.
- A timeline button to show events in the war according to time sequence.
- In the middle, it shows slides of the most popular three articles in the database.



Runtime functionality - search bar

Index page:

- Keyword input as search query
- Select starting time and end time
- Select media source

The screenshot shows a search bar interface with three main input fields. Arrows from the list items point to these fields: 'Battle of Brovary' for the keyword, '03/09/2022' for the start time, and '05/08/2022' for the end time. A dropdown menu is open, showing 'MEDIA' as the selected source, with 'Francetvinfo.fr' and 'Buzzorange.com' as options.

Battle of Brovary	03/09/2022	05/08/2022	MEDIA Francetvinfo.fr Buzzorange.com	SEARCH
-------------------	------------	------------	--	--------



Bingle

mm/dd/yyyy



mm/dd/yyyy



MEDIA

SEARCH

[HOME](#)[TIMELINE](#)

Timeline

[Home](#) → [Timeline](#)

2022-02-22

01. [Donetsk People's Republic](#)02. [Luhansk People's Republic](#)

2022-02-24

01. [Attack on Snake Island](#)02. [Siege of Mariupol](#)03. [Battle of Antonov Airport](#)04. [Siege of Chernihiv](#)05. [Battle of Okhtyrka](#)06. [Chuhuiv air base attack](#)07. [Battle of Kharkiv](#)

Click timeline button

Arrange events
according to the time it
happened in the russian-
ukraine war



Runtime functionality - Timeline

Timeline:

- Events in ukraine - russian war displayed in time sequence
- Could be used as keyword for search

Purpose:

- When the user click it, it could be used as search query
- Different articles with their sources and sentiment analysis pop up
- Different attitudes are showed

2022-02-24

01. [Attack on Snake Island](#)
02. [Siege of Mariupol](#)
03. [Battle of Antonov Airport](#)
04. [Siege of Chernihiv](#)
05. [Battle of Okhtyrka](#)
06. [Chuhuiv air base attack](#)
07. [Battle of Kharkiv](#)
08. [Battle of Konotop](#)
09. [Battle of Sumy](#)
10. [Battle of Chernobyl](#)
11. [Battle of Kherson](#)
12. [Kyiv offensive](#)
13. [Southern Ukraine offensive](#)
14. [Northeastern Ukraine offensive](#)
15. [Eastern Ukraine offensive](#)
16. [On conducting a special military operation](#)

Ukraine: Missile strike in Brovary, near Kyiv, says its mayor

Origin language (French): Ukraine: Frappe de missile à Brovary, près de Kyiv, dit son maire

Challenges

Search By Google

Reuters

April 17, 2022

(Reuters) - A missile strike in the early hours of Sunday damaged infrastructure in the town of Brovary, near the Ukrainian capital Kyiv, the mayor of Brovary said in a me... [+813 tanks]

Origin language (French): (Reuters) - Une frappe de missile aux premières heures de dimanche a endommagé des infrastructures dans la ville de Brovary, près de la capitale ukrainienne Kyiv, a dit le maire de Brovary dans un me... [+813 chars]





Comment function:


Registered users post their comments.


It also shows the date they posted them

Comments



Tony:
World peace !!!!
 2022-05-08 22:23:19 UTC



Tony:
World peace !!!!
 2022-05-08 22:21:52 UTC

Enter your comment here

submit



Runtime functionality - Comments



- News articles are open for registered users to comment
- All comments are from tweets dataset, so they are fake ones built by us.
- Every time, when a comment is posted, a 'hook' would be synchronized to elasticsearch to update the rank score.
- Number of comments could influence ranking of the search results



Bingle

Battle of Brovary

mm/dd/yyyy

mm/dd/yyyy

MEDIA

SEARCH

[HOME](#) [TIMELINE](#)

Results

[Home](#) → [Results](#)

bingle

Media source

Search Result

0.001512 seconds

Challenges

April 17, 2022

1

[Ukraine: Missile strike in Brovary, near Kyiv, says its mayor](#)

Origin language (French): Ukraine: Frappe de missile à Brovary, près de Kyiv, dit son maire (Reuters) - Une frappe de missile aux premières heures de dimanche a endommagé des infrastructures dans la ville de Brovary, près de la capitale ukrain...

Franceinfo.fr

March 13, 2022

1

[War in Ukraine: the distress of the suburbs of kyiv, surrounded by the Russians](#)

Origin language (French): Guerre en Ukraine : la détresse des faubourgs de Kiev, encerclés par les Russes Qu'importe les risques, Brovary (Ukraine), Sacha Serputko charge les voitures pour les autres, avec des gâteaux, des soupes et quelques produits de pre...

Search results:

For each search result, it shows the media source, the sentiment analysis of the article, the translation of the headline and the original texts.




Runtime functionality - sentiment analysis


Sentiment analysis: range(-1,1),


-1 as the most negative(red), 1 as the most positive(blue)

We built a neural network model to train and predict

the sentiment. -> machine learning

Bfmtv April 05, 2022 
[Warsaw signs the purchase of 250 American Abrams battle tanks for 4.3 billion euros](#)
Origin language (French): Varsovie signe l'achat de 250 chars de combat américains Abrams pour 4,3 milliards d'euros
La Pologne a signé un contrat de 4,3 milliards d'euros pour 250 chars de combat américains Abrams avec "une quantité sérieuse de munitions". Le minist...

Fox News April 21, 2022 
[Mariupol battle rages with thousands injured, Ukraine official calls it 'key moment' in war](#)
Thousands are reportedly injured as the battle for Mariupol continues with Russia claiming victory Thursday as Ukrainian officials warning world leader...

Brooklyn.Vegan March 11, 2022 
[Tour news: Battles, Fly Anakin, Caspian, Daniel Romano / Carson McHone, more](#)
Here's a roundup of recent tour news. Check the Tour Dates category for more. BATTLES NYC art rock group Battles have announced spring tour dates tha...

A horizontal bar with a teal left half and an orange right half.

Ranking Algorithm

$\text{BM25} + \log(1 + 0.1 * \text{number of comments})$

All ranking scores = BM25 results + $\log(1 + 0.1 * \text{number of comments})$

And the result updates with the change of number of the comments.

Logic behind the algorithm: if a news is more discussed, then it is more like a hit. It is more possible to be searched and to be wanted to be seen by the users. -> more comments, higher rank



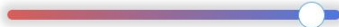
Ranking algorithm



FX Empire

April 15, 2022

5



[Ukraine says street battles ongoing in Mariupol](#)

(Reuters) Ukraines defence ministry spokesman Oleksandr Motuzanyk said on Friday that street battles were ongoing in the besieged port city of Mariupo...

Newser

March 19, 2022

2



[Battle Rages in Mariupol's Streets](#)

(Newser) The intense battle between Ukrainian and Russian forces for Mariupol has moved to the streets in the center of the city, officials said Satur...

news with 5
comments is
ranked higher
than the one
with 2
comments

DEMO TIME



- Searching scenarios
- UI display
- results
- explanation



A short horizontal bar with a teal left half and an orange right half.

primary contributions

- Data collection: Yin Zhendong
- corpus processing: Yin Zhendong, Zhu Yifan
- UI development: Yin Zhendong
- elasticsearch schema definition: Yin Zhendong
- information extraction: Zhu Yifan
- knowledge base development: Yin Zhendong, Zhu Yifan
- algorithm implementation: Yin Zhendong, Zhu Yifan
- machine learning components: Zhu Yifan



LESSONS

—



What worked well:

We managed to collect news data, design and construct our own database, implement sentiment analysis, and enable the comment function. We are able to use events in the Russian - Ukraine war as topic to display critical positions.



A short horizontal bar with a teal segment on the left and an orange segment on the right, located to the left of the title.

What didn't work as well as planned (and why):

Critical thinking of the Russian - Ukraine war is our starting point, when collecting all data news, we forgot to transform news documents into events first. When we are talking about critical thinking, it should be based on an event with the same time, location and even participating pirates. And our data collection failed to take care of that as it should be the preparation before the building of the database.

improving this application in the future:

Next time, we would like to form events first and then make that as the input of our database. In that way, all search will be around events and the searched results could be more critical indeed.