UKRAINIAN CATHOLIC UNIVERSITY

BACHELOR THESIS

Aggregation system of news search with configurable notifications by keywords

Author:
Arsen Ilchyniak

Supervisor:
Dmytro Pryimak

A thesis submitted in fulfillment of the requirements for the degree of Bachelor of Science

in the

Department of Computer Sciences
Faculty of Applied Sciences

Lviv 2022

Declaration of Authorship

I, Arsen Ilchyniak, declare that this thesis titled, "Aggregation system of news search with configurable notifications by keywords" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

"*quote*"

Author

UKRAINIAN CATHOLIC UNIVERSITY

Faculty of Applied Sciences

Bachelor of Science

Aggregation system of news search with configurable notifications by keywords

by Arsen Ilchyniak

Abstract

This thesis focuses on analysing current news resources and optimization of accessibility. Other focus is performance of program, algorithms and tools connected to it. One of main things is implementation of solution with usage of Java programming language and spring framework. Data integrity is implemented with Apache Kafka. Two types of algorithm are proposed to implement main logic. As result, one of them is chosen and effective telegram bot with its implementation is created.

Acknowledgements

I appreciate help from my supervisor Dmytro Pryimak, who supported me during this time and was reminding to do this work and not to drop it.

Contents

Declaration of Authorship **i**

Abstract **iii**

Acknowledgements **iv**

1 Introduction **1**
1.1 Motivation . . . 1
1.2 Reasons of resource improvement . . . 1
1.3 Goals . . . 1

2 Analysis of existing sources **2**
2.1 Introduction . . . 2
2.2 Google trends . . . 2
2.3 Similarweb . . . 3
2.4 tgstat . . . 5
2.5 Conclusions . . . 6

3 Architecture and implementation **8**
3.1 Architecture planning . . . 8
3.2 Used tools and frameworks . . . 8
3.3 Algorithms testing . . . 10

4 Conclusions **13**

5 References **14**
5.1 References . . . 14
5.2 Analytical websites . . . 14

List of Figures

2.1 My architecture 2
2.2 Google analytic about most popular search topics 3
2.3 Censor.net statistics 4
2.4 Ukr.net official statistics 4
2.5 Pravda statistics 4
2.6 TSN statistics 5
2.7 Truexanewsua statistics 6
2.8 Zelenskiy official statistics 6

3.1 Architecture 9
3.2 User interactions 9
3.3 Database model 11
3.4 Milliseconds it takes to process n words on 10000 characters. 12

Chapter 1

Introduction

1.1 Motivation

Last month I spent hours reading the news and becoming more dependent on them. I decided to watch through information connected to news resources. As a result, short research helped me discover that many users have the same problem and became a new idea for my work.

1.2 Reasons of resource improvement

News always was a popular theme. However, in our country, the number of people interested in the news has significantly increased over the last time. This has led to creating an uncountable number of news resources. However, in modern reality, not everyone can spend a significant amount of time reviewing each resource separately and looking through information that interests him. Web users have become much pickier than they used to be. That is why many people began to switch from the usual sources of information such as websites to alternatives, the central place among which took Telegram, including channels, chats and bots. However, users still spend much time looking through different sources to find the information that interests them.

1.3 Goals

- Analyze and collect information about the available media resources.
- Plan and architect the solution based on advantages gotten.
- Implement an alternative solution.

Chapter 2

Analysis of existing sources

2.1 Introduction

In this part different analytical tools are used to show trends and demand of different news resources.

2.2 Google trends

Google trends - https://trends.google.com/

A resource allows viewing and analyzing specific queries that users have typed in Google and other related systems. This tool will help see a reasonably accurate picture of changes in the popularity of specific topics and search trends for each day. The popularity of news in Ukraine has grown significantly since the war started, as indicated by statistics, see Fig. 2.1. Firstly, it was a panic growth, but now we can talk about a steady audience growth, which is currently three times larger than before the war.

Many people are surfing the internet searching for information during this period. They want to be confident in the information and choose a source for themselves or find a place where more compelling or exclusive news is. Another important factor is which queries are still the most popular. People are more likely to query specific news than watch news in general, as shown in Fig. 2.2. The general conclusion is that individual moments and details are much more interesting to people than giving them a broad picture. These results suggest that the information distributed in individual parts is more popular than when all the news follows one another. This information will be taken and compared with the results of further analysis.

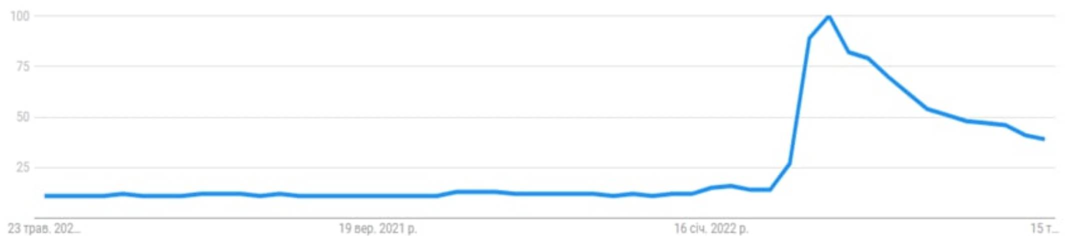
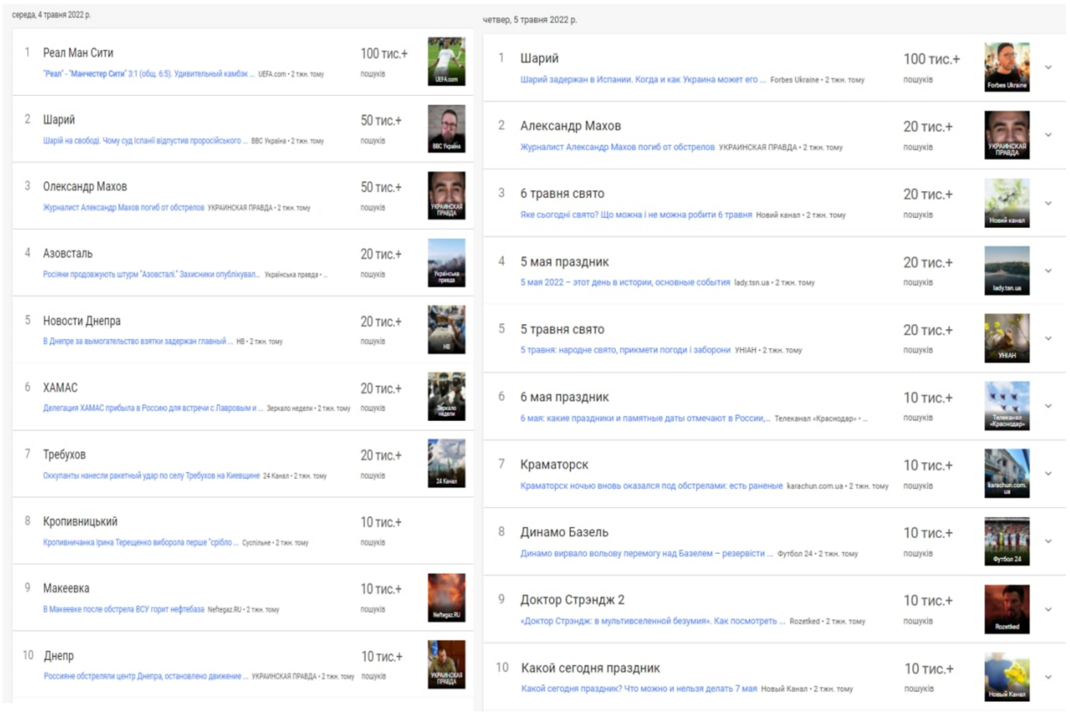


FIGURE 2.1: My architecture



середа, 4 травня 2022 р.

1 Реал Ман Сити — 100 тис.+ пошуків
"Реал" - "Манчестер Сити" 3:1 (общ. 6:5). Удивительный камбэк ... UEFA.com • 2 тжн. тому

2 Шарий — 50 тис.+ пошуків
Шарій на свободі. Чому суд Іспанії відпустив проросійського ... BBC Україна • 2 тжн. тому

3 Олександр Махов — 50 тис.+ пошуків
Журналист Александр Махов погиб от обстрелов УКРАИНСКАЯ ПРАВДА • 2 тжн. тому

4 Азовсталь — 20 тис.+ пошуків
Росіяни продовжують штурм "Азовсталі." Захисники опублікували... Українська правда • ...

5 Новости Днепра — 20 тис.+ пошуків
В Днепре за вымогательство взятки задержан главный ... НВ • 2 тжн. тому

6 ХАМАС — 20 тис.+ пошуків
Делегация ХАМАС прибыла в Россию для встречи с Лавровым и ... Зеркало недели • 2 тжн. тому

7 Требухов — 20 тис.+ пошуків
Оккупанты нанесли ракетный удар по селу Требухов на Киевщине 24 Канал • 2 тжн. тому

8 Кропивницький — 10 тис.+ пошуків
Кропивничанка Ірина Терещенко виборола перше "срібло ... Суспільне • 2 тжн. тому

9 Макеевка — 10 тис.+ пошуків
В Макеевке после обстрела ВСУ горит нефтебаза Neftegaz.RU • 2 тжн. тому

10 Днепр — 10 тис.+ пошуків
Россияне обстреляли центр Днепра, остановлено движение ... УКРАИНСКАЯ ПРАВДА • 2 тжн. тому

четвер, 5 травня 2022 р.

1 Шарий — 100 тис.+ пошуків
Шарий задержан в Испании. Когда и как Украина может его ... Forbes Ukraine • 2 тжн. тому

2 Александр Махов — 20 тис.+ пошуків
Журналист Александр Махов погиб от обстрелов УКРАИНСКАЯ ПРАВДА • 2 тжн. тому

3 6 травня свято — 20 тис.+ пошуків
Яке сьогодні свято? Що можна і не можна робити 6 травня Новий канал • 2 тжн. тому

4 5 мая праздник — 20 тис.+ пошуків
5 мая 2022 – этот день в истории, основные события lady.tsn.ua • 2 тжн. тому

5 5 травня свято — 20 тис.+ пошуків
5 травня: народне свято, прикмети погоди і заборони УНІАН • 2 тжн. тому

6 6 мая праздник — 10 тис.+ пошуків
6 мая: какие праздники и памятные даты отмечают в России,... Телеканал «Краснодар» • ...

7 Краматорск — 10 тис.+ пошуків
Краматорск ночью вновь оказался под обстрелами: есть раненые karachun.com.ua • 2 тжн. тому

8 Динамо Базель — 10 тис.+ пошуків
Динамо вирвало вольову перемогу над Базелем – резервісти ... Футбол 24 • 2 тжн. тому

9 Доктор Стрэндж 2 — 10 тис.+ пошуків
«Доктор Стрэндж: в мультивселенной безумия». Как посмотреть ... Rozetked • 2 тжн. тому

10 Какой сегодня праздник — 10 тис.+ пошуків
Какой сегодня праздник? Что можно и нельзя делать 7 мая Новый Канал • 2 тжн. тому

FIGURE 2.2: Google analytic about most popular search topics

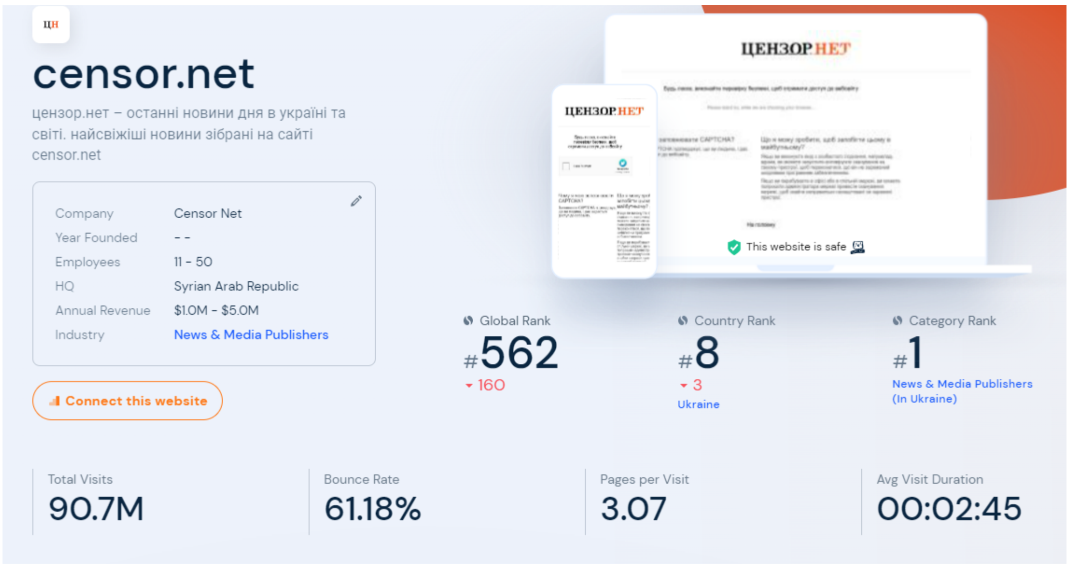
2.3 Similarweb

Similarweb - https://www.similarweb.com/

This part uses several parameters for evaluation, namely the number of visitors, average visit time, page attendance, bounce rate, features. Censor.net is used to be the most rated system among others. With only ninety million visits per month and most rated among other websites, it has several features that make it so favoured. First of all, it is an independent resource where the editor-in-chief is Yuriy Butusov, who is known for being on the front lines of war. Secondly, it can find texts by predefined keywords. Compared to others, it has better access to the previously mentioned feature. Lastly, it has pretty unique materials about the situation on the front line and is different from the official resources view. On the other side, it has a bigger than average bounce rate caused by people checking the website for new information and closing when they do not see anything new. The same thing can be said about all information resources.

TSN is the most visited website. It has two hundred millions visits per month. A pretty similar one has Pravda. Both resources have a small average page visit rate. At the same time, the main difference is the interface. Pravda has a simpler one, which causes a lower bounce rate, while TSN has the biggest one caused by site complexity. TSN got its audience from TV, partly because of its connection to the president in the past. One interesting feature that both of them have is the up bar of losses of the occupiers' army.

UKR.NET has the biggest average visit time and the number of pages visited, but it can be due to its usage as email. However, Similarweb still rates it as second among news resources. So, it is not only about email usage but about the news too. Furthermore, the primary thing that makes it popular is topic-based information. Usage of diverse information resources is a reason too.


censor.net

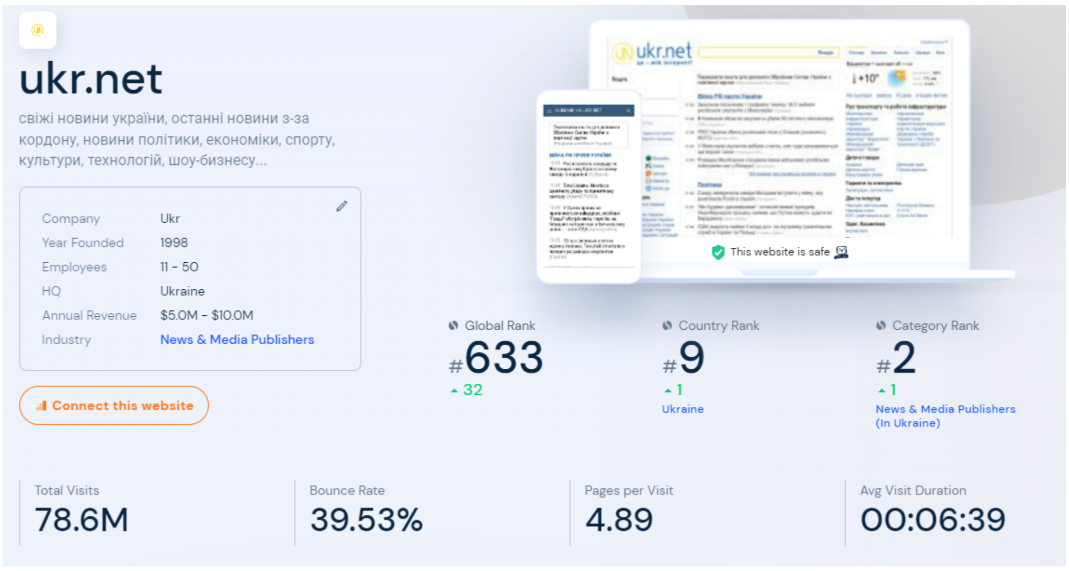
цензор.нет – останні новини дня в україні та світі. найсвіжіші новини зібрані на сайті censor.net

Company	Censor Net
Year Founded	- -
Employees	11 - 50
HQ	Syrian Arab Republic
Annual Revenue	$1.0M - $5.0M
Industry	News & Media Publishers

Global Rank	Country Rank	Category Rank
#562	#8	#1
▾ 160	▾ 3 Ukraine	News & Media Publishers (In Ukraine)

Total Visits	Bounce Rate	Pages per Visit	Avg Visit Duration
90.7M	61.18%	3.07	00:02:45

FIGURE 2.3: Censor.net statistics


ukr.net

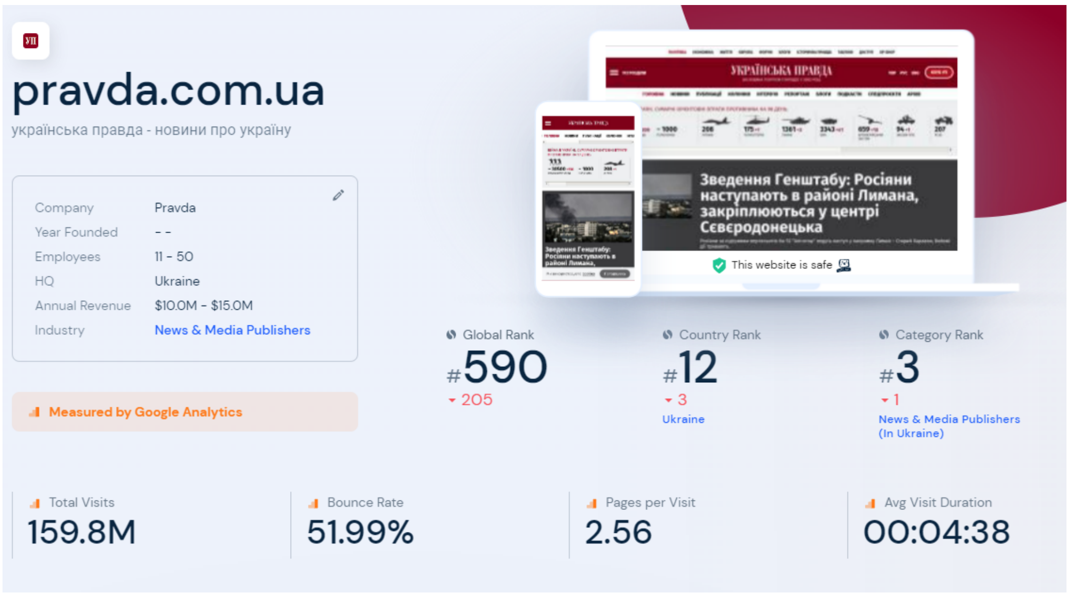
свіжі новини україни, останні новини з-за кордону, новини політики, економіки, спорту, культури, технологій, шоу-бизнесу...

Company	Ukr
Year Founded	1998
Employees	11 - 50
HQ	Ukraine
Annual Revenue	$5.0M - $10.0M
Industry	News & Media Publishers

Global Rank	Country Rank	Category Rank
#633	#9	#2
▴ 32	▴ 1 Ukraine	▴ 1 News & Media Publishers (In Ukraine)

Total Visits	Bounce Rate	Pages per Visit	Avg Visit Duration
78.6M	39.53%	4.89	00:06:39

FIGURE 2.4: Ukr.net official statistics


pravda.com.ua

українська правда - новини про україну

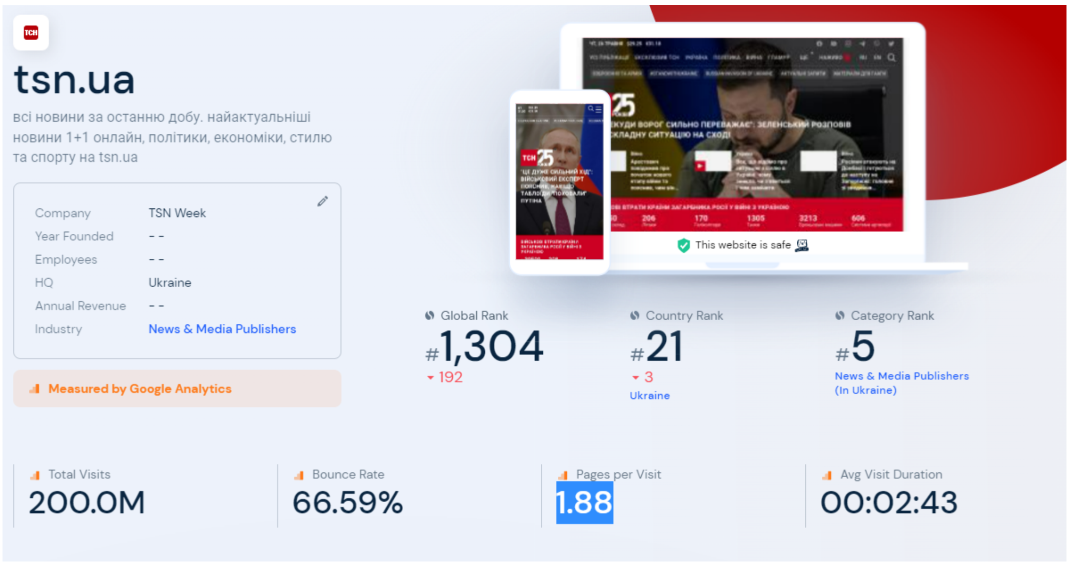
Company	Pravda
Year Founded	- -
Employees	11 - 50
HQ	Ukraine
Annual Revenue	$10.0M - $15.0M
Industry	News & Media Publishers

Measured by Google Analytics
This website is safe
Зведення Генштабу: Росіяни наступають в районі Лимана, закріплюються у центрі Сєвєродонецька

Global Rank	Country Rank	Category Rank
#590	#12	#3
▾ 205	▾ 3 Ukraine	▾ 1 News & Media Publishers (In Ukraine)

Total Visits	Bounce Rate	Pages per Visit	Avg Visit Duration
159.8M	51.99%	2.56	00:04:38

FIGURE 2.5: Pravda statistics


tsn.ua

всі новини за останню добу. найактуальніші новини 1+1 онлайн, політики, економіки, стилю та спорту на tsn.ua

Company	TSN Week
Year Founded	- -
Employees	- -
HQ	Ukraine
Annual Revenue	- -
Industry	News & Media Publishers

Measured by Google Analytics

Global Rank #1,304 ▾ 192

Country Rank #21 ▾ 3 Ukraine

Category Rank #5 News & Media Publishers (In Ukraine)

Total Visits	Bounce Rate	Pages per Visit	Avg Visit Duration
200.0M	66.59%	1.88	00:02:43

FIGURE 2.6: TSN statistics

Based on previous results, several conclusions about the efficiency and disadvantages of different resources were made. First of all, this is the number of visited pages. Web sites with separated topics have a lower bounce rate. Due to that, UKR.NET and Censor.net have higher results by this parameter. Moreover, Pravda has a bigger audience due to its straightforward interface. Despite the fact that TSN is the most visited one, its bounce rate and small average page visit rate tell us about its weaknesses and show that a significant part of the audience will leave soon.

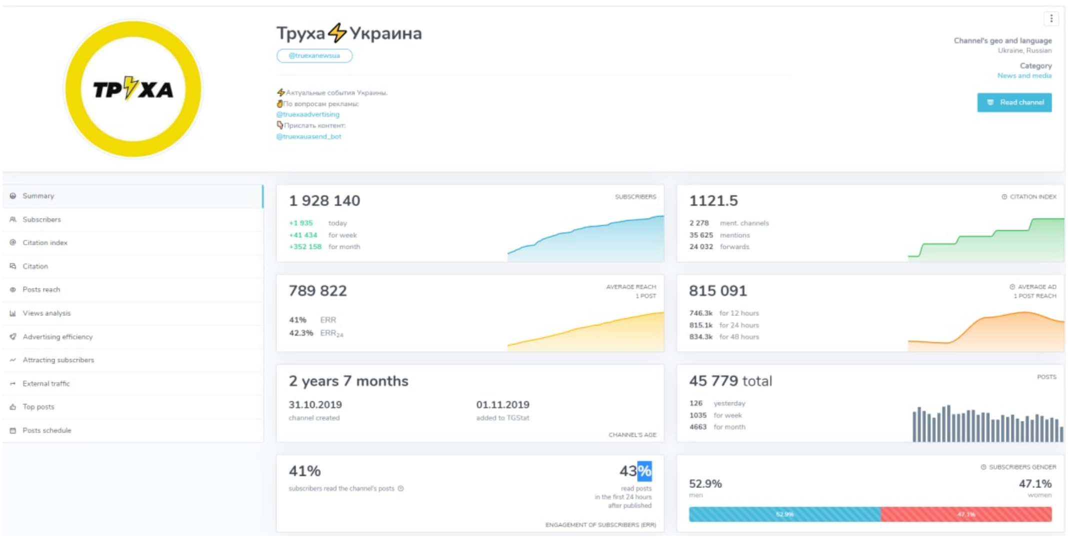
2.4 tgstat

Tgstat - https://uk.tgstat.com/ratings/channels/news?sort=members

For subscribers' analysis and news popularity in Telegram, the work compares the two biggest channels in Ukraine with several official ones. U_Now provides information in quite an ordinary manner. All information is verified and comes almost already after official publication on the internet. On another side, Truexanewsua quite often publishes not verified information. However, this allows him to publish many exclusive materials at the exact moment after some event happens, making it popular among users.

This work takes @V_Zelenskiy_official as the biggest official information resource. Generally, the subscribers' number of official information resources is between 0.6 and 1.3 million when biggest unofficial channels have close to 2 million subscribers. Mostly it is connected to the ability of unofficial channels to public information from as many sources as they can process. However, the total number of views of official posts is often higher than others. It is connected to resend of official sources. For example, views of one post in @V_Zelenskiy_official is 1.15 higher than subscribers. At the exact moment, @Truexanewsua have only 40 per cent views compared to its subscribers. Despite this, the number of subscribers of @Truexanewsua continues to rise when @V_Zelenskiy_official decreases.

The other resource that is popular in telegram is @UASmartNewsBot. Opened information about these bot users is missing. However, due to its spread among different media when KitSoft launched it and its new way of accessing data, it is assumed that the audience is quite significant. The main advantages are different



Труха⚡Украина
@truexanewsua

⚡Актуальные события Украины.
По вопросам рекламы:
@truexaadvertising
Прислать контент:
@truexauasend_bot

Channel's geo and language
Ukraine, Russian
Category
News and media

SUBSCRIBERS
1 928 140
+1 935 today
+41 434 for week
+352 158 for month

CITATION INDEX
1121.5
2 278 ment. channels
35 625 mentions
24 032 forwards

AVERAGE REACH 1 POST
789 822
41% ERR
42.3% ERR$_{24}$

AVERAGE AD 1 POST REACH
815 091
746.3k for 12 hours
815.1k for 24 hours
834.3k for 48 hours

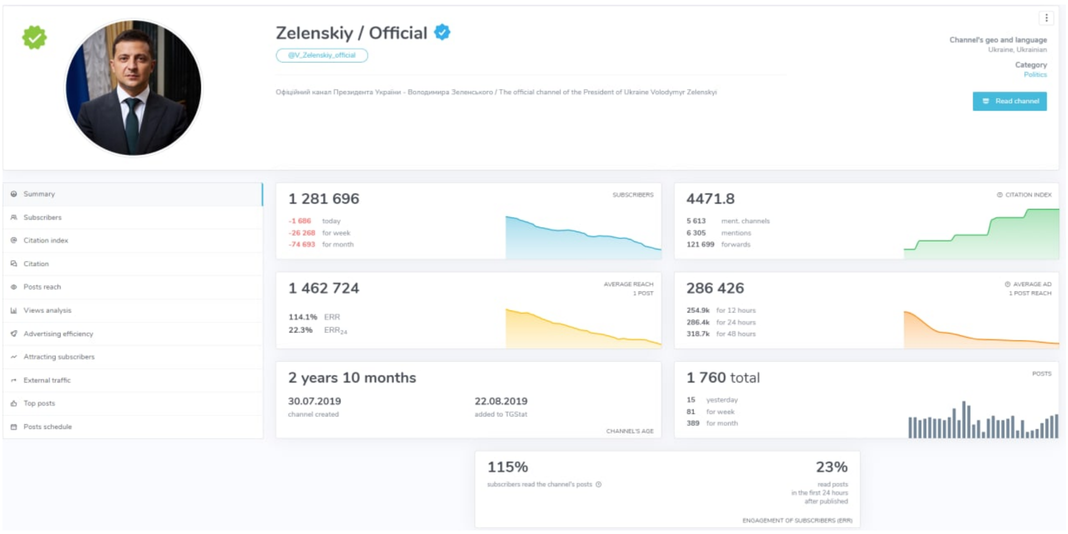
CHANNEL'S AGE
2 years 7 months
31.10.2019 channel created
01.11.2019 added to TGStat

POSTS
45 779 total
126 yesterday
1035 for week
4663 for month

ENGAGEMENT OF SUBSCRIBERS (ERR)
41% subscribers read the channel's posts
43% read posts in the first 24 hours after published

SUBSCRIBERS GENDER
52.9% men
47.1% women

FIGURE 2.7: Truexanewsua statistics



Zelenskiy / Official
@V_Zelenskiy_official

Офіційний канал Президента України - Володимира Зеленського / The official channel of the President of Ukraine Volodymyr Zelenskyi

Channel's geo and language
Ukraine, Ukrainian
Category
Politics

SUBSCRIBERS
1 281 696
-1 686 today
-26 268 for week
-74 693 for month

CITATION INDEX
4471.8
5 613 ment. channels
6 305 mentions
121 699 forwards

AVERAGE REACH 1 POST
1 462 724
114.1% ERR
22.3% ERR$_{24}$

AVERAGE AD 1 POST REACH
286 426
254.9k for 12 hours
286.4k for 24 hours
318.7k for 48 hours

CHANNEL'S AGE
2 years 10 months
30.07.2019 channel created
22.08.2019 added to TGStat

POSTS
1 760 total
15 yesterday
81 for week
389 for month

ENGAGEMENT OF SUBSCRIBERS (ERR)
115% subscribers read the channel's posts
23% read posts in the first 24 hours after published

FIGURE 2.8: Zelenskiy official statistics

sources of information collected in one place and that customers can write and get news by keywords that interest them. Because of the decision to take information only from official resources and not add new resources after launch, the channel became worthless resource for a significant part of the audience. Another big drawback is its base of 86 official channels that post similar information, and users get notifications about the same action but from many similar sources.

2.5 Conclusions

- People require structured information.
- People want to see information from different resources.
- Alternative resources have become more popular and continue to improve.
- Telegram is one of the significant resources.
- Information delivery has to be fast to interest subscribers.

- A variety of sources make information sound more confident.
- Exclusive materials make the resource more attractive.

Chapter 3

Architecture and implementation

3.1 Architecture planning

First of all, I have thought about the type of information delivery. Channels this time have too much information and are too generalized. I have chosen the "Telegram" bot, which has potential. However, as mentioned in the last part, "@UASmartNews-Bot" has not used this potential. Commonly, users have two primary needs. Firstly, subscribers want to get notifications when information that interests them is published. Secondly, they want to get information about some topic in the past.

As a result, this work needs to have a database to save information about two different types of information. The first one is subscribers' keywords needed to make information delivered user-specific. Another one is articles that have to be delivered to users interested in past events.

It is possible to make a new resource that will interest people, but this work is about improving news accessibility. Therefore, I have decided to scrap and process different popular resources.

Another mandatory part is the effectiveness of this work. The number of subscribers and keywords will increase with time, and naively making this work will lead to failure. According to this, I have tested several algorithms for keyword checking in texts. Last but not least, according to data consistency and modules that process the information, I have to add a message broker for the transmission user requests reliability.

There are a set of commands that users will use:

- Add keyword. To do this, send a keyword to the bot. Then the user is subscribed to the news that contains this word. Later, he has to wait for when new news containing his keywords appear
- Delete keyword. Works with /delete keyword.
- Get posts by keyword and dates. A user enters command /get_posts_by_keyword_and_dates. After that, he has to enter a keyword and two dates in the next message.

3.2 Used tools and frameworks

Java was chosen as the programing language.

Java is a programming language and computing platform first released by Sun Microsystems in 1995. It has evolved from humble beginnings to power a large share of today's digital world, by providing the reliable platform upon which many services and applications are built. New, innovative products and digital services designed for the future continue to rely on Java, as well.

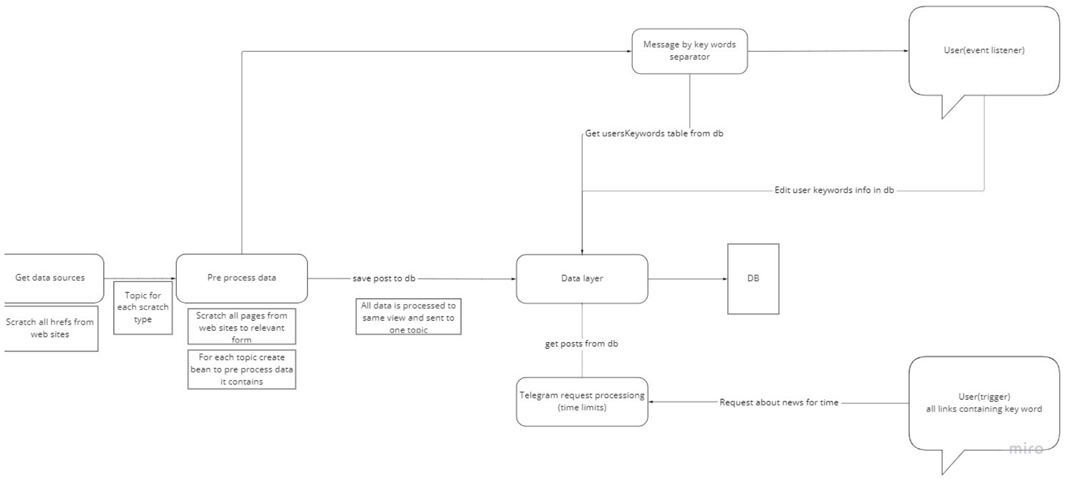

Message by key words separator
User(event listener)
Get usersKeywords table from db
Edit user keywords info in db
Get data sources
Scratch all hrefs from web sites
Topic for each scratch type
Pre process data
Scratch all pages from web sites to relevant form
For each topic create bean to pre process data it contains
save post to db
All data is processed to same view and sent to one topic
Data layer
DB
get posts from db
Telegram request processiong (time limits)
Request about news for time
User(trigger) all links containing key word
miro

FIGURE 3.1: Architecture

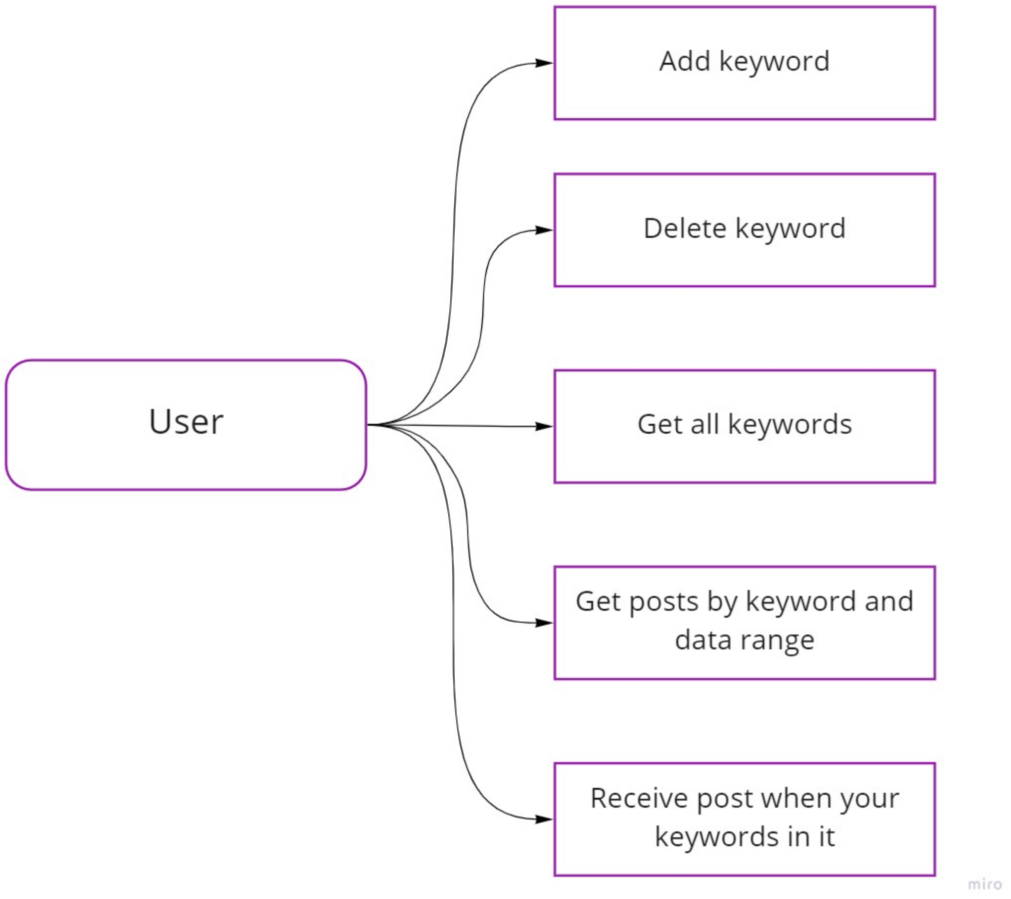

User
Add keyword
Delete keyword
Get all keywords
Get posts by keyword and data range
Receive post when your keywords in it
miro

FIGURE 3.2: User interactions

Its libraries are pretty helpful and easy to use.

It has many tools to simplify work and use different annotations to replace tones of code.

Kafka was chosen as a message broker. Combining it and Spring for Apache Kafka makes it the most suitable option.

Apache Kafka — Robust Queue Broker

" Kafka combines three key capabilities so you can implement your use cases for event streaming end-to-end with a single battle-tested solution:

- To publish (write) and subscribe to (read) streams of events, including continuous import/export of your data from other systems.
- To store streams of events durably and reliably for as long as you want.
- To process streams of events as they occur or retrospectively.

And all this functionality is provided in a distributed, highly scalable, elastic, fault-tolerant, and secure manner. Kafka can be deployed on bare-metal hardware, virtual machines, and containers, and on-premises as well as in the cloud. You can choose between self-managing your Kafka environments and using fully managed services offered by a variety of vendors. " Description of kafka at official page.

" The Spring for Apache Kafka (spring-kafka) project applies core Spring concepts to the development of Kafka-based messaging solutions. It provides a "template" as a high-level abstraction for sending messages. It also provides support for Message-driven POJOs with @KafkaListener annotations and a "listener container". These libraries promote the use of dependency injection and declarative. " Description of spring-kafka at official page.

I have chosen "PostgreSQL" as a database because it is standardized and easy to work with.

"PostgreSQL is a powerful, open source object-relational database system with over 30 years of active development that has earned it a strong reputation for reliability, feature robustness, and performance. ". Description of PostgreSQL at official page.

3.3 Algorithms testing

There were two main algorithms that this work tests.
The first one is - "Aho-Corasick algorithm".

" Let there be a set of strings with the total length M (sum of all lengths). The Aho-Corasick algorithm constructs a data structure similar to a trie with some additional links, and then constructs a finite state machine (automaton) in O(MK) time, where K is the size of the used alphabet.

It was considered that code would need to process big text files. However, practice shows that the average amount of chars in one file is less than 1000. Moreover, there is a variety of letters in the alphabet. As a result, this algorithm was the worst among others. ". Description of suffix-automation from cp-algorihms.
The second one is - "Suffix Automation".

" A suffix automaton is a powerful data structure that allows solving many string-related problems.

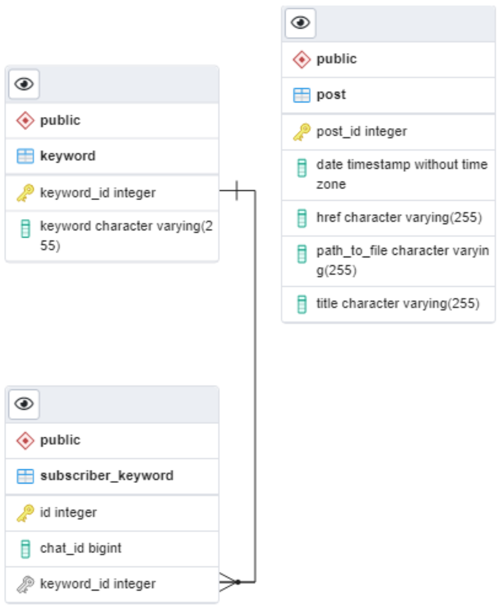

public
keyword
keyword_id integer
keyword character varying(255)
public
post
post_id integer
date timestamp without time zone
href character varying(255)
path_to_file character varying(255)
title character varying(255)
public
subscriber_keyword
id integer
chat_id bigint
keyword_id integer

FIGURE 3.3: Database model

For example, you can search for all occurrences of one string in another, or count the amount of different substrings of a given string. Both tasks can be solved in linear time with the help of a suffix automaton.

Intuitively a suffix automaton can be understood as a compressed form of all substrings of a given string. An impressive fact is, that the suffix automaton contains all this information in a highly compressed form. For a string of length n it only requires O(n) memory. Moreover, it can also be built in O(n) time (if we consider the size k of the alphabet as a constant), otherwise both the memory and the time complexity will be O(n*log(k)).

The linearity of the size of the suffix automaton was first discovered in 1983 by Blumer et al., and in 1985 the first linear algorithms for the construction was presented by Crochemore and Blumer.

". Description of aho-corasick from cp-algorihms.

Naive method. String method - ".contains()" is used to check whether a keyword is in text. Its complexity is O(n*k), where n - text length, k - the number of keywords.

Comparing these two left algorithms, we can easily see that Suffix automation is faster with a larger number of keywords. However, with a small number of keywords, it deals better. This is because a naive algorithm does not have to prepare some structure to give results about keywords much faster. According to this I can say that naive algorithm will spend too much time and resources to execute the same piece of work. If to watch from perspective of small program with limited number of users it is quite satisfying. General results can be even better then Suffix Automation, but users would not see them. And when the number of subscribers will increase and number of keywords too, the program will need much more time to execute all queries that come. As result information will not be delivered fast and exclusive as it was planed for.

As the result, Suffix automation were taken, because of its overall result and good complexity.

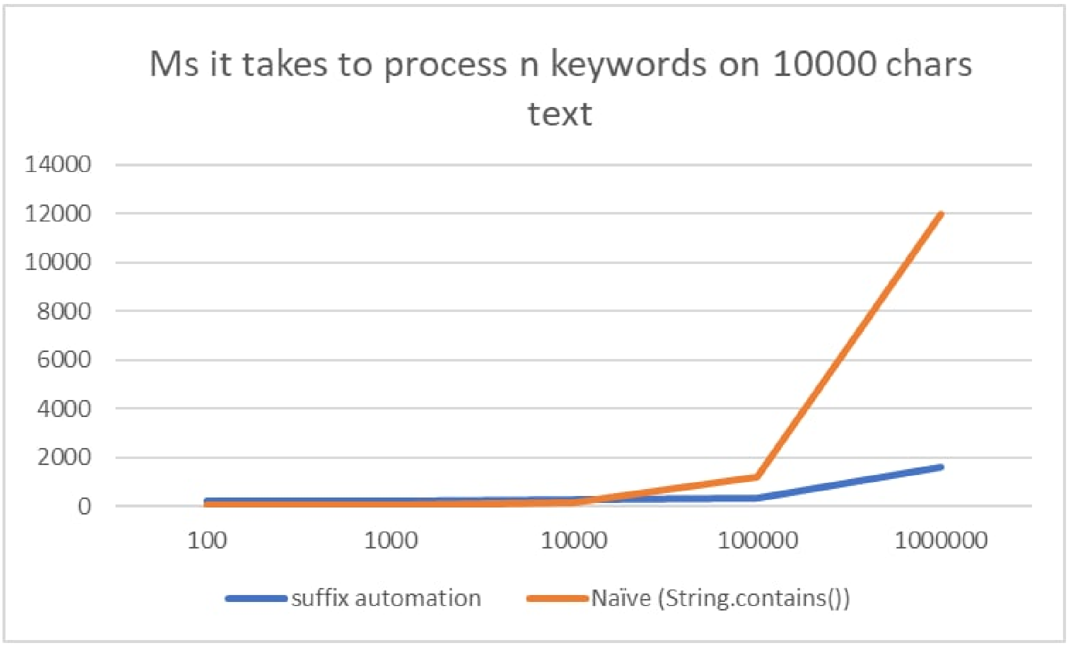


FIGURE 3.4: Milliseconds it takes to process n words on 10000 characters.

Chapter 4

Conclusions

As a result, I got a ready program that sends new news to subscribers by keywords they entered. It has a simple interface and can run on any machine globally with Telegram API. This work analyses algorithms and detects their weak and strong sides. Aho-Corasick, which I have used much before, turned out to be useless here because of its specification and data amount that incomes from news resources.

I have already partly launched it and gave bots link to several people to test it. As a result, several of my acquaintances started to browse news sources less if to believe their words. The most effect it had was on my acquaintance, who is sensitive to everything, and this bot helps her see only good news and avoid bad by entering specific keywords. Hope after deploy it will attract more people and I will show statistics of its usage during presentation.

Chapter 5

References

5.1 References

- Telegram api documentation - https://core.telegram.org/bots/api
- Best message brokers - https://blog.containerize.com/2021/07/09/top-5-open-source-message-queue-software-in-2021
- Aho-corasick algorithm - https://cp-algorithms.com/string/aho_corasick.htmlreferences
- Suffix-automation - https://cp-algorithms.com/string/suffix-automaton.html
- General Suffix Automaton Construction - https://static.googleusercontent.com/media/research.goo
- Kafka documentation - https://kafka.apache.org/documentation/
- Spring framework for Kafka documentation - https://docs.spring.io/spring-kafka/docs/current/reference/html/
- Spring data jpa documentation - https://docs.spring.io/spring-data/jpa/docs/current/reference/ht
- Lombok documentation - https://projectlombok.org/features/all
- Jsoup documentation - https://jsoup.org/
- Date formats - https://jenkov.com/tutorials/java-date-time/parsing-formatting-dates.html

5.2 Analytical websites

- Telegram statistics - https://uk.tgstat.com/ratings/channels/news?sort=members
- Similarweb - https://www.similarweb.com/
- Google trends - https://trends.google.com/
- Pravda advertise - https://www.pravda.com.ua/cdn/cd1/advertising/