УДК004.04

РОЗРОБКА ІНФОРМАЦІЙНОЇ СИСТЕМИ ДЛЯ ЗБОРУ ІНФОРМАЦІЇ З ДОСТОВІРНИХ ТА ПЕРЕВІРЕНИХ ДЖЕРЕЛ,ДЛЯ ІНФОРМУВАННЯ ТА АНАЛІЗУ НАСЕЛЕННЯ УКРАЇНИ ПІД ЧАС ВІЙНИ

Голопотилюк Євгеній Андрійович
Кандидат технічних наук,доцент Рудніченко Микола Дмитрович
Національний Університет «Одеська Політехніка», УКРАЇНА

АНОТАЦІЯ. Розробка парсингової системи для збору достовірної інформації під час війни. Категоризація інформації за відповідними категоріями та запитами. Аналіз користувачів за демографічними даними та пристроями.

Вступ. Простір інтернету кожного дня заповнює велика кількість інформації. Кожен з нас заходить до пошукачу за відповідями на наші запитання.Планування відпустки, пошук нових знань, перевірка достовірностей фактів, все це відбувається в міжнародній павутині. Переглядаючи інформацію можна побачити дві новини з різним результатом, або різними історичними зносками.

Після 24 лютого російські засоби масової інформації почали створювати велику кількість недостовірної інформації - викривлення історії, підробка медіа матеріалів та багато іншого.

Засобами пропаганди російська сторона мала частку контролю на українській території до початку тапід час війни. Дезінформація під час воєнного стану може призвести до смертельних ризиків для людини.

Місцеві групи у месенджерах сповіщали про небезпеку та наближення ракет за короткий час. Отримуючи швидко достовірну інформацію з декількох історників дає можливість застерегти себе, або бути в курсі загальної картини.

Мета роботи. Розробка інформаційної систем, що збирає інформацію з достовірних джерел, обробляє та категоризує і аналізує поведінкові фактори людей під час війни.

Основна частина роботи. Розбираючи дану проблему, потрібно виділити основні модулі для проектування-технологія збору даних,база даних для обробки великої кількості інформації, список джерел,платформа для відображення інформації.Обробка даних включає в себе лексичний та синтаксичний аналіз. Загалом це можна охарактеризувати як процес аналізу рядка символів на мові, яка відповідає правилам формальної граматики. Аналіз з точки зору аналізу даних розширює це визначення до двоетапного процесу, який показує програмний парсер, що може прочитати, проаналізувати або перетворити.Результатом є більш структурований формат. Інструмент SeleniumWebDriver використовується для автоматизації тестування веб-додатків,щоб переконатися,що вони працюють належним чином.Він підтримує багато браузерів,таких як Firefox,Chrome,IEtaSafari.Однак,використовуючи SeleniumWebDriver,ми можемо автоматизувати тестування лише веб-додатків.Він не відповідає вимогам для віконних програм[1].

Вибираючи базу даних для обробки великої кількості інформації, потрібно орієнтуватись наNoSQLрішення.Використовуючи технології наданій системі не потрібно видаляти весь об’єм даних з таблиці, перероблювати структуру та переносити дані. В такому алгоритмі існують шанси загублення даних.

Elasticsearch - це сучасна пошуково-аналітична система, яка базується на Apache Lucene. Побудована на Java, Elasticsearch є базою даних NoSQL. Це означає, що вона зберігає дані в не структурованому вигляді і що ви не можете використовувати SQL для запитів до неї.

Інформаційні джерела були зібрані з ютубу та офіційних сайтів України для подальшого збору інформації.В списку знаходяться такі ресурси :“ТСН”, “УНІАН”,“ЗНАЙ.UA”,“Хлопці з лісу”та інші.Загальний обсяг джерел налічує 65 ютуб каналів та4медіасайту.Для відображення

інформації потрібно створити портал з новинами та відео, що будуть мати основні категорії: “Війна”, “Політика”, “Економіка”, “Світ”, “Україна”,”Історія”,”Життя”.

Проектуючи дану систему, потрібно створити чіткий механізм для взаємодії усіх модулей даного проекту. На рисунку 1 відображена система сбору інформації.

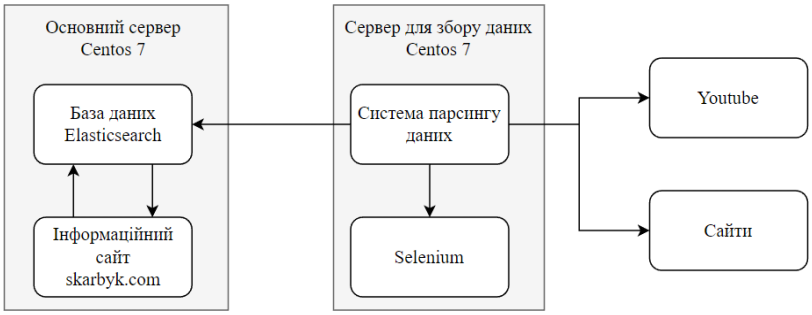

Рисунок1 –Система сбору інформації з достовірних джерел

При розробці інформаційного сайту потрібно використати традиційні українські кольори, створити сторінки зі списком відеотастатей,використовувати достовірні джерела для збору інформації.Сайт повинен бути розроблений на технологіях:HTML5,CSS,Js,JQuery,PHP7.4[2]. Для аналізу користувачів потрібно використати Google Analytics та Google Console. Для генерації користувачів потрібно використати FaceBook ADS та Google ADS. За допомогою цих рекламних систем,можна згенерувати користувачів котрим цікава тема зу країнськими новинами та історією.

На основі технічного завдання був створений сайт під назвою “Скарбик”, що дає можливість збирати інформацію з відповідних джерел.Скріншот з готового сайту відображений на рисунку 2.

Сайт виконує всі задані функції та на 15.05.2023 рік містить більше ніж 100 000 зібраних даних у вигляді відео та статей. Кожного дня додається близько 50-80 інформаційних додатків.

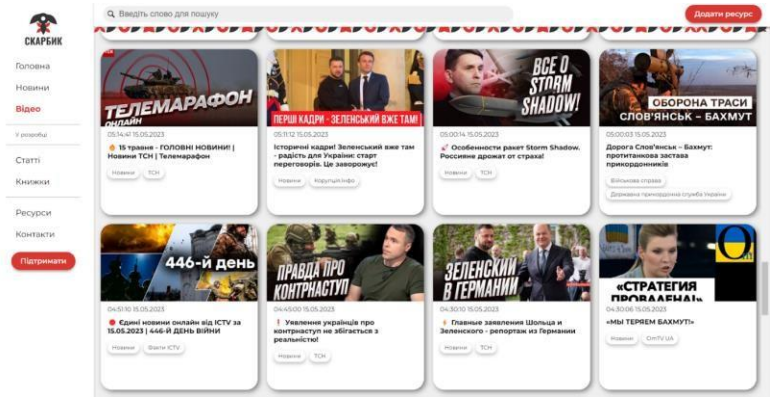

Рисунок2–Функціонуючий прототип інформаційної системи збору інформації з достовірних джерел

Після розробки та запуску реклами на сайт, почався етап збору інформації про користувачів, завдяки аналітичним ресурсам, що були встановлені на сайт, можно було зібрати відповідні факти.Так як велика кількість українців виїхала за межі України та локальна пошукова система інших стран видає в основному російську інформацію, було вирішено запустити рекламу на 5 країн для перегляду попиту на український контент [3].

Перший аналіз користувачів за демографією Google ADS. Запускаючи дану рекламу було використано контекстно та контекстно медійну рекламу.У таблиці1 відображені дані по країнам [3].

Таблиця1-Демографічні дані покраїнам

Країна	Кількістькористувачів
Poland	1175
Latvia	1035
Romania	383
Lithuania	314
Germany	227
Ukraine	208

За допомогою аналітичних інструментів можна збирати міста та їх популярність серед інших показників. У таблиці 2 відображені дані по самим популярним містам [3].

Таблиця2-Демографічні дані по містам

Місто	Кількість користувачів
Riga	742
Warsaw	342
Vilnius	166
Wroclaw	131
Poznan	109

На території України активні дві мови - українська та російська. В таблиці 3 показана кількість язиків та їх частка в браузерах [3].

Таблиця3-Демографічні дані помові в браузері

Мова браузеру	Кількість користувачів
Russian	2207
Ukrainian	509
English	314
Latvian	227

Використовуючи всі дані, що були здобуті за допомогою маркетингових інструментів - демографія, пристрої, карта дій користувача, можна зробити висновок. Після початку війни в Україні попит на новини та український контент виріс у різних країнах,а більш за всього у Польщі, Латвії та Румунії. Близько 5% браузерів користувачів були переведені на локальні мови,що може свідчити про прискорення адаптації в іншій країні.Середній вік користувача був 40-60відсотків.Молодше покоління знаходиться в соціальних мережах.Зібрані дані показують, що попит на український контент росте, але не завжди цей контент доходить до людини котра його шукає.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. UnmeshGundecha,SeleniumTestingToolsCookbook/GundechaUnmesh,02Novt.2012
2. AvinashKaushik,WebAnalytics2.0:TheArtofOnlineAccountability/KaushikAvinash,27Oct.2009
3. MohamedAlibi,MasteringCentOS7LinuxServer(EnglishEdition)/AlibiMohamed,05Jun2016

DEVELOPMENT OF AN INFORMATION SYSTEM FOR COLLECTING INFORMATION FROM RELIABLE AND VERIFIED SOURCES TO INFORM AND ANALYZE THE POPULATION OF UKRAINE DURING THE WAR

Yevhen Holopotyliuk,
Candidate of Technical Sciences,Associate Professor Mykola Rudnichenko
Odessa Polytechnic National University, UKRAINE

ANNOTATION. Developing a parsing system to collect reliable information during the war. Categorization of information into relevant categories and queries. Analysis of users by demographics and devices.