

Програмний засіб для децентралізованого формування наборів відкритих даних

Вінницький національний технічний університет

Анотація

Розроблено програмний засіб для децентралізованого формування наборів відкритих даних різних категорій. Реалізація базується на інтелектуальному хмарному сервісі Firebase ML.

Ключові слова: відкриті дані, геолокація, розпізнавання об'єктів, мобільний додаток.

Abstract

The software for decentralized forming open data sets of different categories has been developed. The implementation is based on the intelligent cloud service Firebase ML.

Keywords: Open Data, Geolocation, Object Recognition, Mobile Application.

Вступ

Відкриті дані (або open data) — підхід до зберігання інформації, згідно з яким набір даних має бути у вільному доступі, вільно використовуватися та розповсюджуватися. При цьому такі дані можуть використовувати як неприбуткові організації, так і комерційні установи [1]. Вони створюють можливість швидко отримати першочергові дані, які дозволяють приймати правильні рішення, які базуються на достовірній, а не заангажованій інформації.

Дослідження організації TAPAS спільно з міжнародними партнерами показало, що відкриті дані вже принесли понад 700 мільйонів доларів до економіки країни у 2017 році [2]. З них близько 200 мільйонів — це прибутки українських компаній за нові продукти або послуги на базі відкритих даних. Ще близько півмільярда — непряма вигода від більш ефективної роботи. Відкриті дані можуть слугувати першочерговим інструментом для творення інформаційного суспільства. За їхньою допомогою можна швидко отримувати першочергову інформацію про важливі аспекти життя суспільства. Відтак рішення компаній та громадян будуть ґрунтуватися на достовірній інформації з першоджерел (а не на суб'єктивній оцінці експертів, ЗМІ чи компаній). Основний результат використання таких даних — прозорість процесів у суспільстві. Саме тому актуальним є створення наборів відкритих даних.

Постановка задачі

Для забезпечення сталого демократичного розвитку у державі необхідно надати вільний доступ громадянам, громадським об'єднанням та іншим зацікавленим особам до публічної інформації, що має суспільно важливе значення. Міністерства та держпідприємства збирають первинні дані без врахування єдиних методологічних стандартів, що ускладнює роботу з даними й унеможливує їх об'єднання. Відсутня інфраструктура відкритих даних – немає належних серверних потужностей задля підтримання Національного порталу відкритих даних. Відсутніми є також спеціалісти з досвідом роботи з профільними відкритими даними, тому функції з оприлюднення даних покладаються на працівників прес-служби. Ще однією з проблем є те, що відсутнє оперативне оновлення даних; не всі існуючі набори передбачають прив'язку до геокоординат; недостатня кількість наданих державними структурами даних та категорій даних.

Виходячи з вищеписаних ознак, було прийнято рішення розробити програмний засіб, що б дозволяв формувати набори відкритих даних по різних категоріям.

Розробка системи

Програмний засіб базується на використанні геоінформаційної технології [3]. Структурну схему засобу наведено на рисунку 1. Складовими системи є сервер (містить базу даних), систему інтелектуального аналізу зображень (СІА), ряд об'єктів та клієнтів.

Кожен клієнт представляє собою будь-який пристрій, на якому є фотокамера, GPS-навігатор, доступ до мережі Інтернет та дозволяє встановити розроблений додаток (мобільний телефон, планшет). Камера призначена для фото фіксації об'єкта, GPS – для отримання геолокації, яка буде прив'язана до зображення при додаванні даних у набір. Також відбувається прив'язка до моменту часу, коли було здійснено отримання зображення. Оскільки дані формуються будь-яким користувачем на відміну від єдиного державного органу, то такий підхід до децентралізованого формування даних дозволяє підвищити ефективність створення наборів даних.

Далі зображення надходить до СІА та аналізується. Там відбувається створення опису зображення, а саме списку знайдених об'єктів та вірогідності їх наявності. Далі відбувається диференціація отриманих результатів. Залишаються лише ті об'єкти, які було визначено з вірогідністю більше 85%. Після цього необхідні дані надсилаються клієнту.

Процес формування наборів даних складається з двох етапів. Під час першого етапу отриманий список об'єктів розподіляється на категорії (наприклад пожежа, дорожньо-транспортна пригода тощо). Далі отримані дані разом з координатами обробляються та надсилаються на сервер. Сервер в свою чергу обробляє дані та записує їх у базу даних. Оскільки клієнти можуть також отримувати створені набори даних, то зв'язок між клієнтом та сервером є двостороннім.

Також створювана система дозволяє візуалізувати набори даних з прив'язкою їх до місцевості. Кожен набір представляється набором маркерів різних кольорів, що відображаються на карті. Координати кожного маркера – ті, що були отримані під час обробки зображення.

Структура наборів даних створена на основі методичних рекомендацій Державного агентства з питань електронного урядування України [4].

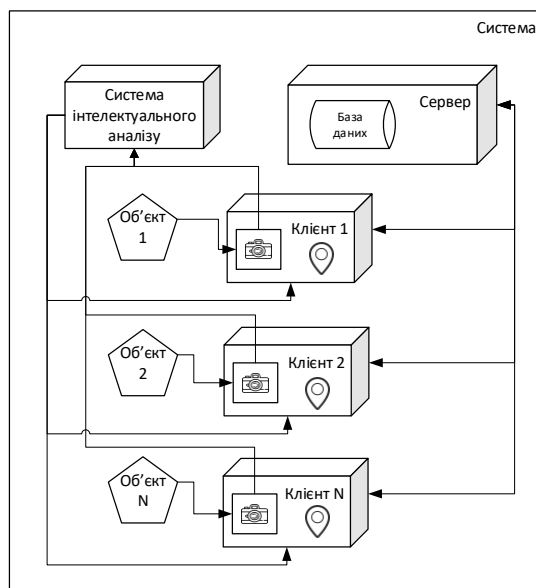


Рисунок 1 – Структурна схема роботи програмного засобу

Для розробки серверної частини додатку було використано мову програмування C# та такі технології: .NET Core, ASP, Web API, MS SQL, Elasticsearch [5]. Elasticsearch використовується для збільшення швидкодії та зменшення складності, адже ця технологія дозволяє індексувати збережені зображення.

Розроблено архітектуру клієнтської частини додатку, що складається з чотирьох основних модулів: модуля авторизації, модуля обробки зображень, модуля геовізуалізації та модуля статистики.

Для реалізації клієнтської частини програмного додатку було вирішено використовувати фреймворк Flutter та мова програмування Dart. Flutter – це програмний каркас із відкритим кодом, для створення додатків для платформ Android та iOS, а також для створення веб-застосунків, розроблений компанією Google [6]. На відміну від багатьох відомих на сьогоднішній день мобільних платформ, Flutter не використовує JavaScript ні в якому вигляді. В якості мови програмування для Flutter вибрали

Dart, який компілюється в бінарний код, за рахунок чого досягається швидкість виконання операцій порівнянн з Objective-C, Swift, Java, або Kotlin.

Модуль авторизації призначений для реєстрації користувача в додатку та подальшого входу в додаток. Для авторизації використовується Firebase Authentication. Вхід можна здійснити також через облікові записи Facebook або Google.

Модуль обробки зображень полягає в тому, що користувач робить фото на камеру, в цей момент до фото підв'язується геолокація, далі додаток взаємодіє з Firebase та за допомогою ML Kit розпізнає, що знаходиться на фото та формує відповідні масиви об'єктів та за допомогою API передає ці дані на сервер.

Метою модуля геовізуалізації є отримання через API даних із сервера та відображення їх геолокацій на карті відповідно до категорій. Вікно завантаження фото зображене на рисунку 2. Вікно карти – на рисунку 3.

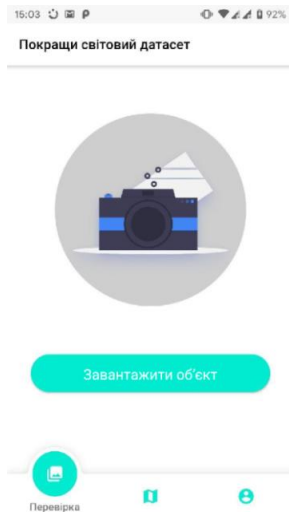


Рисунок 2 – Вікно завантаження зображення

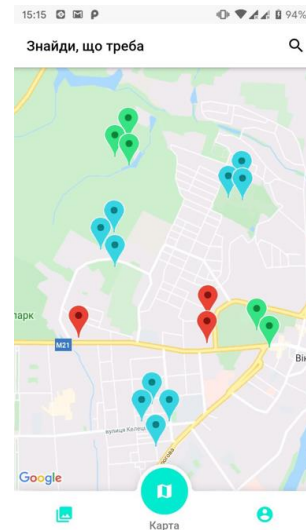


Рисунок 3 – Вікно карти

Модуль статистики використовується для відображення інформації про користувача, але основною метою є показ статистики згідно з вибраною категорією. Щоб вибрати категорію потрібно перейти на сторінку категорій та виділити зі списку ті категорії, які цікавлять. Після цього є можливість підписатись на обрані категорії (рис. 4). Далі ми можемо переглянути статистику по кожній з категорій (рис. 5). Також там можна завантажити набір даних у форматі json чи csv.

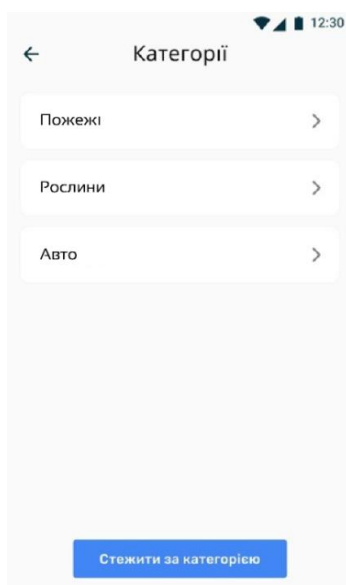


Рисунок 4 – Вікно списку категорій

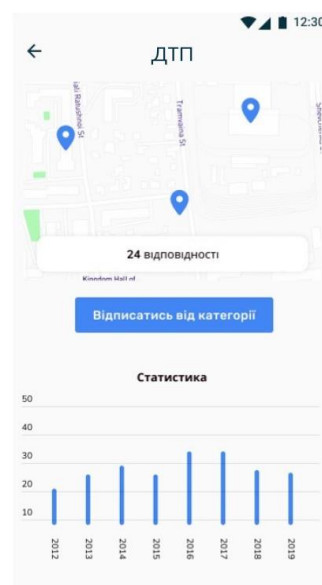


Рисунок 5 – Вікно статистики

Висновки

Проаналізувавши актуальність і важливість досліджуваної теми, було розроблено програмний засіб для децентралізованого формування наборів відкритих даних. Розроблений додаток дозволяє формувати набори відкритих даних на основі зображень, що роблять користувачі клієнтського додатку. Дані можуть бути використані різними дослідниками, аналітиками, державними структурами та будь-якими іншими громадянами. Створена система дозволяє вирішувати такі проблеми, як автоматизація формування даних, отримання статистичних даних про протікання певних процесів на території та відсутність потужної бази для систем машинного навчання.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Головна сторінка - Data.gov.ua [Електронний ресурс] Режим доступу: <https://data.gov.ua/>
2. Відкриті дані українських міст. Рекомендації щодо впровадження політики [Електронний ресурс]. – Режим доступу: <https://storage.decentralization.gov.ua/uploads/library/file/377/2019.02.11.pdf>.
3. Інтелектуальна геоінформаційна технологія формування наборів відкритих даних // М.Д. Кренцін, Л.М. Куперштейн, І.А. Колос, В.О. Ніколайчук, С.О. Васильков // Матеріали IV міжнародної науково-практичної конференції «PRIORITY DIRECTIONS OF SCIENCE DEVELOPMENT» – Львів, 2020 [Електронний ресурс] – Режим доступу: http://sci-conf.com.ua/wp-content/uploads/2020/02/priority-directions-of-science-development_3-4.02.2020.pdf
4. Методичні рекомендації Державного агентства з питань електронного урядування України [Електронний ресурс] – Режим доступу: <https://data.gov.ua/uploads/files/2018-08-11-104353.234698Recomendation.pdf>
5. Почему Elasticsearch — хороший выбор для сбора и анализа данных среднего объёма [Електронний ресурс] – Режим доступу: <https://tproger.ru/blogs/why-elasticsearch-is-a-good-choice/>
6. Flutter - Beautiful native apps in record time [Електронний ресурс] – Режим доступу: <https://flutter.dev/>

Куперштейн Леонід Михайлович, канд. тех. наук., доцент кафедри захисту інформації, Вінницький національний технічний університет, Вінниця, e-mail: kupershtein.lm@gmail.com

Кренцін Михайло Дмитрович, студент групи ІПІ-19м, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, Вінниця, e-mail: mishatron98@gmail.com

Колос Ірина Андріївна, студентка групи ІПІ-19м, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, Вінниця, e-mail: kolos.irena@gmail.com

Ніколайчук Владислав Олександрович, студентка групи 2ПІ-19м, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, Вінниця, e-mail: erengard97@gmail.com

Васильков Сергій Олександрович, студент групи ІКІ-16б, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, Вінниця, e-mail: s.vasilkov1998@gmail.com

Leonid Kupershtein, Cand. Sc., Associate Professor of the Department of Information Protection, Vinnytsia National Technical University, Vinnytsia, e-mail: kupershtein.lm@gmail.com

Mykhailo Krentsin, student of group ІPI-19m, Faculty for Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, e-mail: mishatron98@gmail.com

Kolos Iryna, student of group ІPI-19m, Faculty for Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, e-mail: kolos.irena@gmail.com

Nikolaichuk Vladyslav, student of group 2PI-19m, Faculty for Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, e-mail: erengard97@gmail.com

Vasylykov Serhii, student of group ІKI-16b, Faculty for Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, e-mail: s.vasilkov1998@gmail.com