

Магістерська кваліфікаційна робота на тему:

Пошук місць проведення дозвілля з використанням інтелектуального аналізу даних

Виконав студент гр. 1КН-17м Ліщинський М.В.
Керівник: к.т.н., старший викладач кафедри КН
Петришин С.І.

Спеціальність 122 “Комп’ютерні науки”

МЕТА ТА ЗАВДАННЯ ДОСЛІДЖЕННЯ

Метою дослідження магістерської кваліфікаційної роботи є підвищення інформативності веб-орієнтованого пошуку місць проведення дозвілля.

Для досягнення поставленої мети необхідно розв'язати такі завдання:

- провести аналіз проблеми розв'язання задачі пошуку місць проведення дозвілля, проаналізувати поняття, основні показники та фактори, що впливають на пошук місця проведення дозвілля;
- розглянути існуючі методи вирішення задачі пошуку місць проведення дозвілля та обрати й обґрунтувати вибір методу, який задовольняє мету даної магістерської кваліфікаційної роботи;
- розробити метод пошуку місць проведення дозвілля на основі інтелектуального аналізу даних;
- сформулювати стадії інформаційної технології, розробити структуру та алгоритм роботи програмного засобу;
- виконати програмну реалізацію запропонованої інформаційної технології;
- провести тестування програмного продукту та виконати аналіз отриманих результатів.

ОБ'ЄКТ, ПРЕДМЕТ ТА МЕТОДИ ДОСЛІДЖЕННЯ

Об'єкт дослідження – процес пошуку місць для проведення дозвілля комп'ютерними засобами.

Предмет дослідження – інформаційна технологія та програмні засоби пошуку місць проведення дозвілля та їх інформативність.

Методи дослідження

У роботі використані наступні методи наукових досліджень:

- системного аналізу,
- теорії інтелектуального аналізу даних для реалізації інформаційної технології пошуку місць для проведення дозвілля,
- методи математичної статистики, ,
- об'єктно-орієнтованого програмування.

НАУКОВА НОВИЗНА ОДЕРЖАНИХ РЕЗУЛЬТАТІВ

1. Набула подальшого розвитку інформаційна технологія пошуку місць для проведення дозвілля, яка відрізняється використанням модифікованого методу оцінювання місця проведення дозвілля, що привело до підвищення інформативності пошуку.
2. Удосконалено метод оцінювання місця проведення дозвілля, який відрізняється використанням методу к найближчих сусідів, що підвищило інформативність пошуку.

· ПРАКТИЧНЕ ЗНАЧЕННЯ ОДЕРЖАНИХ РЕЗУЛЬТАТІВ

1. розроблено алгоритми роботи програмного забезпечення пошуку місць проведення дозвілля на основі інтелектуального аналізу даних;
2. розроблено програмні засоби пошуку місць для проведення дозвілля на основі інтелектуального аналізу даних;

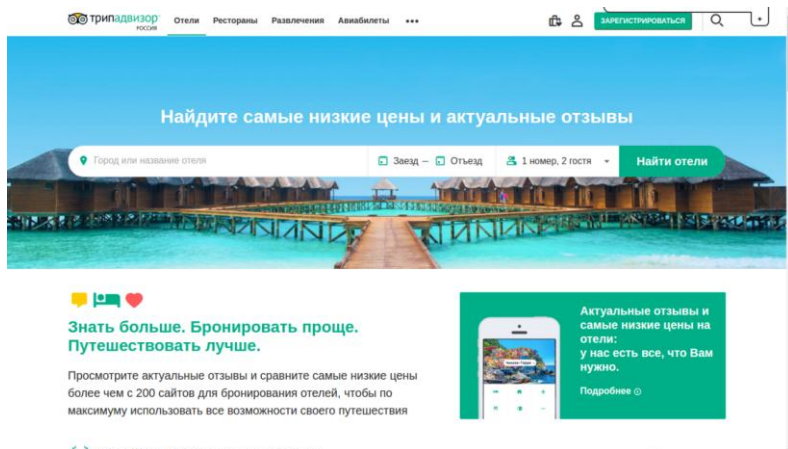
Аналіз предметної області пошуку місць проведення дозвілля

Серед методів класифікації виділяються наступні:

- статистичні методи (лінійна регресія, логістична регресія);
- різні варіанти лінійного програмування;
- дерево класифікації;
- нейронні мережі;
- генетичний алгоритм;
- метод к найближчих сусідів.

Було обрано метод к найближчих сусідів

Вибір і обґрунтування аналогу



Головна сторінка сайту TripAdvisor

Існуючі рішення

- Skyscanner;
- momondo;
- KAYAK;
- TripAdvisor;
- zomato;
- tastecard.

Розробка методу оцінювання місць проведення дозвілля на основі методу класифікації k-найближчих сусідів

Метод класифікації k найближчих сусідів полягає у такому:

1. Знайти K найближчих об'єктів до досліджуваного об'єкту (цифрового вектору) X в навчальній вибірці. Під відстанню між векторами може розумітися, наприклад, Евклідова відстань:

$$\rho(x, x') = \sqrt{\sum_{i=1}^n (x_i - x'_i)^2}$$

або максимум модулів:

$$\rho(x, x') = \max_i |x_i - x'_i|$$

або інша функція відстані.

2. Віднести X до того класу серед K сусідів, який частіше інших зустрічається.

а) випадок K=1: Алгоритм найближчого сусіда (nearest neighbour)

б) випадок K=N: Константний прогноз найбільш частим класом у вибірці.

в) у разі регресії потрібно усереднити характеристики k найближчих сусідів.

Коли користувач реєструється на сайті, він вводить свої параметри, які заносяться у таку форму:

Ідентифікатор користувача	Вік	Стать	K_активність повсякденного життя	K_сон	K_активність відпочинку	K_заробітку	K_дозволяєних витрат	K_населення території проживання	K_завоженості
ID3548	29	ч							

Коли користувач шукає місце проведення дозвілля, то серед найближчих до нього будуть обиратися тільки ті, у яких максимальну оцінку поставили користувачі, подібні йому за параметрами. Це відбувається таким чином:

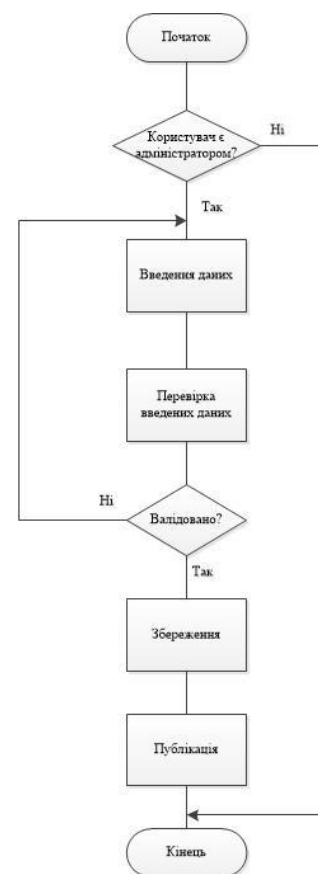
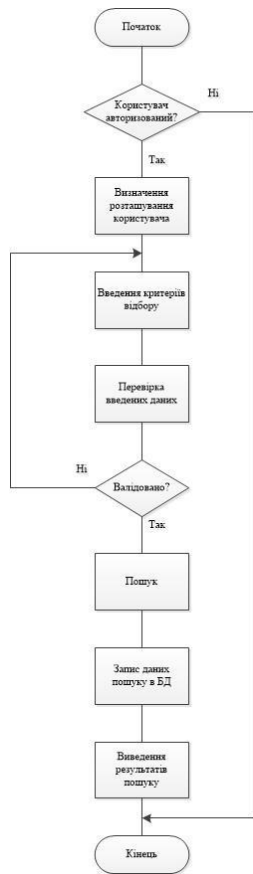
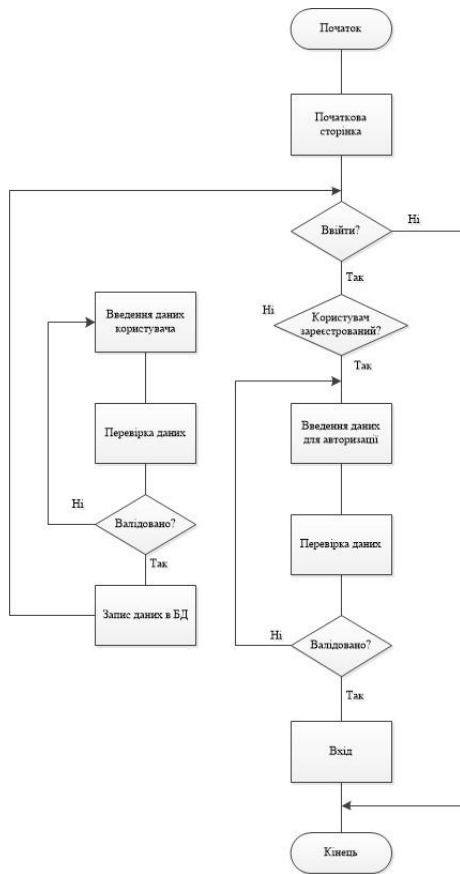
Користувач задає критерій пошуку: категорія місця дозвілля, мінімальна оцінка та радіус. У вказаному радіусі шукаються заклади дозвілля потрібного типу, але по них виводиться не середня оцінка по всіх відгуках користувачів про цей заклад, а серед всіх відгуків про цей заклад обираються тільки відгуки тих користувачів, що подібні (близькі по відстані між векторами ознак) нашому по таких параметрах як вік, стать, спосіб життя, доходів і т. п.

Тобто пропозиція буде більш інформативна, оскільки відображає не загальні враження про цей заклад, а враження тільки тих, до кого подібний наш користувач. І це більш правильно, тому що одній категорії людей заклад проведення дозвілля може подобатись, а іншій категорії – ні.

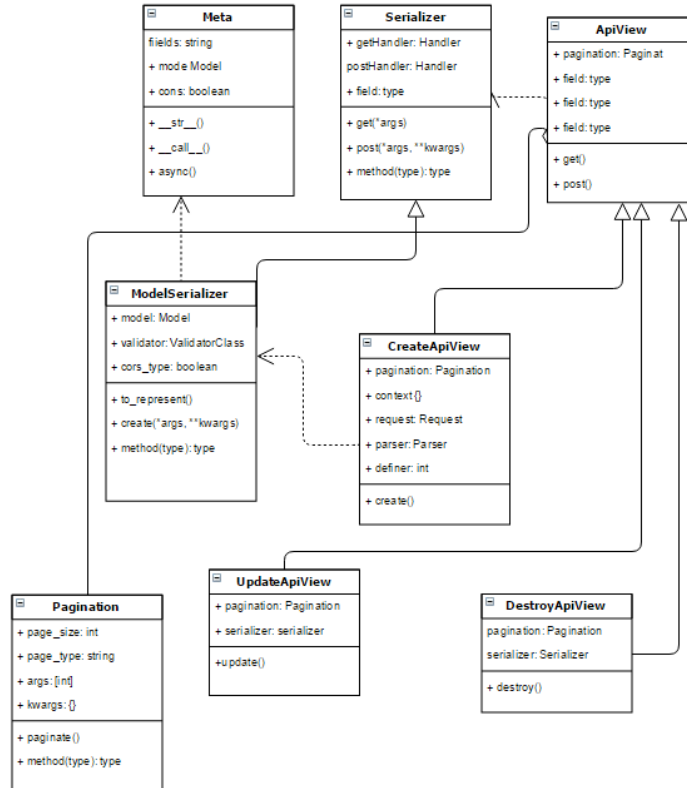
Структура інформаційної технології пошуку місць проведення дозвілля на основі інтелектуального аналізу даних



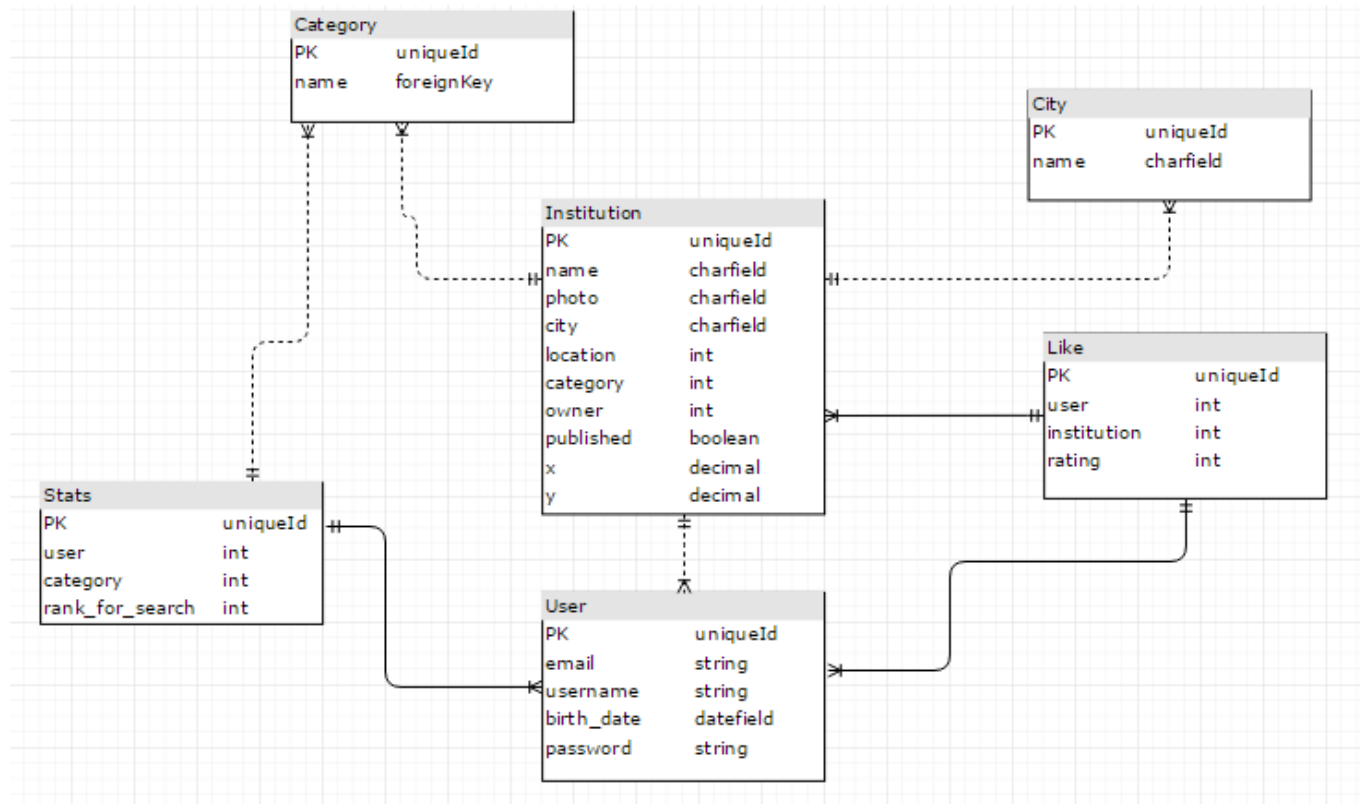
Алгоритми роботи програмної реалізації інформаційної технології пошуку місць проведення дозвілля



UML-діаграма класів



Діаграма баз даних



Засоби розробки (Backend)

Технологія	Ruby on Rails	Django
Рік створення	2004	2005
Кількість зірок на github.com	35 900	26 200
Принцип конфігурації	Model View Controller	Model Template View
JavaScript	Ruby on Rails має вбудований статичний модуль компіляції, який використовує CoffeeScript за замовчуванням.	Django обробляє статичні файли за умовчанням. Популярні програми сторонніх розробників компілюють і стискають файли.

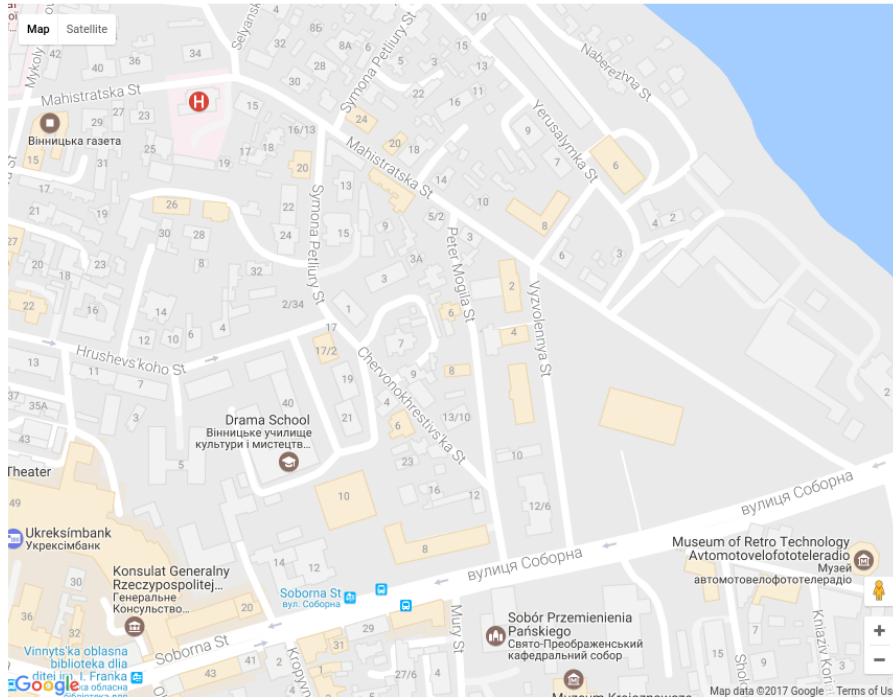
Засоби розробки (Frontend)

Технологія	AngularJS	React.js
Розробник	Google	Facebook
Тип технології	Повністю реалізований MVC фреймворк, написаний на JavaScript	Бібліотека JavaScript (Перегляд в MVC; необхідний Flux для реалізації архітектури)
Кількість зірок на github.com	56 100	68 900
Концепція	Вводить JavaScript в HTML Працює з реально DOM. Рендеринг зі сторони клієнта	Вводить HTML в JavaScript працює з віртуальною DOM. Рендеринг зі сторони сервера
Залежності	Керує залежностями автомаічно	Необхідні додаткові інструменти для керування залежностями

Було обрано мови програмування Python, JavaScript та технологій Django та AngularJS

Головна сторінка

Leisure Search Sign Out



Params

Find nearest place

Category

Min Rank

The nearest place

Find places in radius

Category

Min Rank

Radius

Find places in radius

Сторінка адміна сервісу

Administration WELCOME, ADMIN@EXAMPLE.COM [VIEW SITE](#) / [CHANGE PASSWORD](#) / [LOG OUT](#)

Site administration

ACCOUNTS

Email addresses + Add Change

AUTH TOKEN

Tokens + Add Change

AUTHENTICATION AND AUTHORIZATION

Groups + Add Change

LEISURE

Categories + Add Change

Cities + Add Change

Institutions + Add Change

Likes + Add Change

Stats + Add Change

Temp photos + Add Change

SITES

Sites + Add Change

Recent actions

My actions

- + Sasha_2017-06-10
14:32:03.263304+00:00
Like
- + Mc Donalds
Institution
- Shawerma
Institution
- Shawerma
Institution
- Shawerma
Institution
- Shawerma
Institution
- + Shawerma
Institution
- + Lviv
City
- Francua
Institution
- + Francua
Institution

localhost8000/admin/password_change/

Досягнення мети роботи

Метою магістерської кваліфікаційної роботи було підвищення інформативності пошуку місць проведення дозвілля.

Цю характеристику важко оцінити кількісно. Кількісною оцінкою інформативності сервісу пошуку місць проведення дозвілля можна було б вважати кількість позитивних відгуків користувачів або усереднену оцінку користувачів.

Для того, щоб отримати такі дані, потрібно, аби цим сервісом користувались певний час, щоб отримати такі відгуки та оцінки та порівняти їх з відгуками та оцінками про аналог. Але цей сервіс щойно розроблений і ми не маємо таких оцінок та відгуків. Тому можемо тільки якісно оцінити інформативність розробленого веб-сервісу пошуку місць проведення дозвілля.

Він буде більш інформативним за аналог для конкретного користувача тому, що видає не просто усереднену оцінку знайденого закладу дозвілля по всіх користувачах, а видає середню оцінку тих користувачів, які подібні конкретному за своїми параметрами віку, статі, інтересів, доходів, стилю життя та ін..

Тому мета роботи досягнута, інформативність пошуку місць проведення дозвілля підвищена.

ЕКОНОМІЧНА ЧАСТИНА

Було проведено економічне обґрунтування доцільності розробки ПЗ пошуку місць проведення дозвілля з використанням інтелектуального аналізу даних. Нова розробка має рівень комерційного потенціалу вище середнього. Загальна сума витрат на виконання робіт склала 26929,12 грн. Загальні витрати на виконання та впровадження результатів виконаної наукової роботи – 31681,32 грн. Абсолютна ефективність вкладених інвестицій становить 283788,12грн, і це свідчить про те, що вкладання коштів на виконання та впровадження результатів НДДКР є доцільним. Відносна (щорічна) ефективність вкладених в наукову розробку інвестицій – 115 %, отже інвестор буде зацікавлений у фінансуванні даної наукової розробки. Термін окупності складає 0,87 року, тобто фінансування розробки ПЗ пошуку місць проведення дозвілля є економічно доцільним проектом.

АПРОБАЦІЯ РЕЗУЛЬТАТІВ РОБОТИ ТА ПУБЛІКАЦІЇ

Апробація результатів роботи.

Результати досліджень апробовані на V міжнародній науково-практичній конференції «Інформаційні технології та взаємодії» м. Київ, 20-21 листопада 2018 року.

Публікації.

За результатами магістерської кваліфікаційної роботи опубліковано 1 тези доповідей на конференції та готується заявка на авторське свідоцтво на твір (програму).

ВИСНОВОК

В результаті виконання роботи було розроблено інформаційну технологію пошуку місць проведення дозвілля на основі інтелектуального аналізу даних. Для програмної реалізації веб-технології були використані мови програмування Python, JavaScript та технології Django і AngularJS. Аналіз результатів роботи веб-системи пошуку місць проведення дозвілля показав, що розроблене програмне забезпечення має більшу інформативність пошуку місць проведення дозвілля, ніж аналогічна інформаційно-пошукова система, а значить мета роботи досягнута.

Дякую за увагу!