

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ  
ІМЕНІ ІГОРЯ СІКОРСЬКОГО»

Факультет прикладної математики

Кафедра прикладної математики

Пояснювальна записка

до курсової роботи

із дисципліни “Бази даних та інформаційні системи”

на тему

**“Бот для збирання інформації з вказаних ресурсів за темою лекції”**

Виконав:

студент групи КМ-62

Козирєв А. Ю.

Перевірили:

Терещенко І. О.

Ковальчук-Химюк Л. О.

Київ — 2019

## ЗАВДАННЯ НА ВИКОНАННЯ КУРСОВОЇ РОБОТИ

Основним завданням курсової роботи є створення багатокористувацького сервісу архітектури “клієнт-сервер”, головна задача якого є генерація тексту лекції за темою вказаною користувачем, а також надання можливості зберегти згенеровані ресурси та лекції, редагувати та видаляти їх. За допомогою технології машинного навчання користувачу необхідно лише вказати тему лекції (як словосполучення або кілька словосполучень), яку він хоче згенерувати.

## АНОТАЦІЯ

Додаток буде дуже популярним серед студентів та викладачів університетів. Більше не матиме необхідності шукати в інтернеті різні ресурси та перечитувати велику кількість літератури задля створення лекцій, тепер системи автоматизації будуть виконувати роботу по аналізу та написанню текстів. Для отримання “brand-new” лекції лише треба зробити одну просту дію: визначитись з темою лекції.

## РЕФЕРАТ

У студентів та викладачів багато часу для підготовки доповідей та лекцій саме фільтрація контенту та групування по змісту. Тепер все це можна довірити штучному інтелекту! Нейронна мережа може генерувати осмислені корисні тексти, і створювати унікальний контент на задану тематику. Вже зараз такі відомі проекти як Botnic.AI можуть генерувати сіквели до серії книг “Гаррі Поттер”. Але нейромережі можуть генерувати не тільки нові романи, тексти пісень, а ще й ресурси для навчання. Незважаючи на великий потенціал даної сфери, подібні додатки до сих пір викликають недовіру серед користувачів в мережі Інтернет; все через єдиний вагомий недолік: *нейронні мережі є непередбачуваними*. У сучасних мережах знаходяться мільярди компонентів-нейронів, кожен з яких так чи інакше впливає на результат обчислення, тому точно визначити алгоритм, за яким працює мережа майже неможливо. Але у сфери машинного навчання великий потенціал, саме тому вони мають місце у сучасних мікро-сервісних додатках. Принцип даного додатка є простим: опишіть тему лекції - отримайте вашу лекцію. І нічого зайвого! Штучний інтелект все зробить за вас!

# ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ.....	5
ВСТУП.....	6
ОСНОВНА ЧАСТИНА.....	7
1. АНАЛІЗ ПІДПРИЄМСТВА АВТОМАТИЗАЦІЇ .....	7
a. Передпроектне дослідження .....	7
b. Мета .....	7
2. ПОСТАНОВКА ЗАДАЧІ.....	8
a. Визначення категорії користувачів.....	8
b. Бізнес-правила.....	8
3. МОДЕЛЮВАННЯ БІЗНЕС-ПРОЦЕСІВ.....	8
a. Use-case.....	9
b. Component diagram.....	10
c. UML-diagram.....	11
4. ІНФОЛОГІЧНЕ ПРОЕКТУВАННЯ .....	12
a. Моделі діючих сутностей .....	12
5. ВИСНОВКИ.....	13
6. СПИСОК ВИКОРИСТАНИХ ЛІТЕРАТУРНИХ ДЖЕРЕЛ.....	14

## ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

БД - база даних;

DI - dependency injection (впровадження залежностей);

REST API - Representational State Transfer (передача стану представлення);

U.S. – user story (критерій);

U.C. – use-case (сценарій виконання);

## ВСТУП

У студентів та викладачів багато часу для підготовки доповідей та лекцій саме фільтрація контенту та групування по змісту. Тепер все це можна довірити штучному інтелекту! Нейронна мережа може генерувати осмислені корисні тексти, і створювати унікальний контент на задану тематику. Вже зараз такі відомі проекти як Botnic.AI можуть генерувати сіквели до серії книг “Гаррі Поттер”. Але нейромережі можуть генерувати не тільки нові романи, тексти пісень, а ще й ресурси для навчання. Незважаючи на великий потенціал даної сфери, подібні додатки до сих пір викликають недовіру серед користувачів в мережі Інтернет; все через єдиний вагомий недолік: *нейронні мережі є непередбачуваними*. У сучасних мережах знаходяться мільярди компонентів-нейронів, кожен з яких так чи інакше впливає на результат обчислення, тому точно визначити алгоритм, за яким працює мережа майже неможливо. Але у сфери машинного навчання великий потенціал, саме тому вони мають місце у сучасних мікро-сервісних додатках. Принцип даного додатка є простим: опишіть тему лекції - отримайте вашу лекцію. І нічого зайвого! Штучний інтелект все зробить за вас!

## ОСНОВНА ЧАСТИНА

### АНАЛІЗ ПІДПРИЄМСТВА АВТОМАТИЗАЦІЇ

#### Передпроектне дослідження

Дано датасет з текстів на різні теми. Основною задачею є дослідження кореляції між парою (поточне слово/словосполучення, тема лекції) -> наступне слово/словосполучення. При знаходженні алгоритму з найбільшим коефіцієнтом кореляції, необхідно інтегрувати алгоритм з сервісом по генерації тексту відповідно до теми. Загальним алгоритмом сервісу стає: користувач вводить тему лекції -> користувач отримує текст лекції.

#### Мета

1. Автоматизація процесу аналізу тексту та перетворення його до логічного та релевантного вигляду;
2. Залучення основної аудиторії додатку: студентів та лекторів;

## ПОСТАНОВКА ЗАДАЧІ



## Визначення категорії користувачів

Основними користувачами даного сервісу є студенти 1-6 курсу, лектори та наукові діячі.

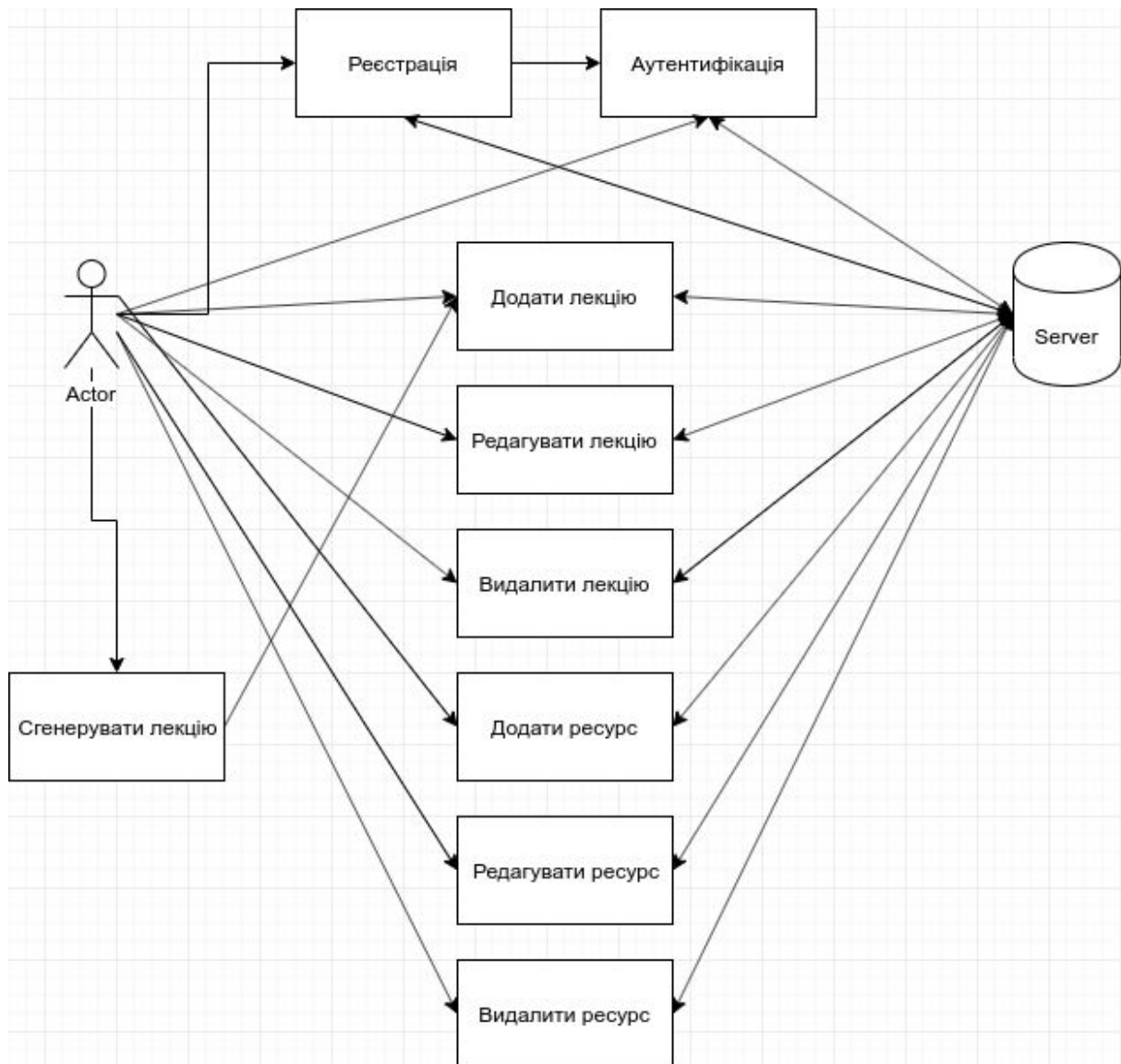
## Бізнес-правила

1. Користувач може зберегти лекцію лише за наявності заголовка та контенту лекції;
2. При генерації лекції вхідними даними є строка **не нульової довжини** що є темою лекції;
3. Для будь-якого нового користувача надана йому роль буде **Default user**;
4. Користувач має доступ до сервісу лише за наявності логіну та пароля;
5. Мінімальна довжина для пароля при реєстрації - 8 символів;
6. Максимальна довжина заголовку лекції: 250 символів;
7. Максимальна довжина контенту лекції: 5000 символів;
8. Користувач може мати лише одну єдину роль у певний час;
9. При додаванні нової лекції поля “Заголовок” та “Контент” повинні бути заповнені;
10. При додаванні нового ресурсу поля “URL” та “Опис” повинні бути заповнені;

## Визначення сутностей сервісу

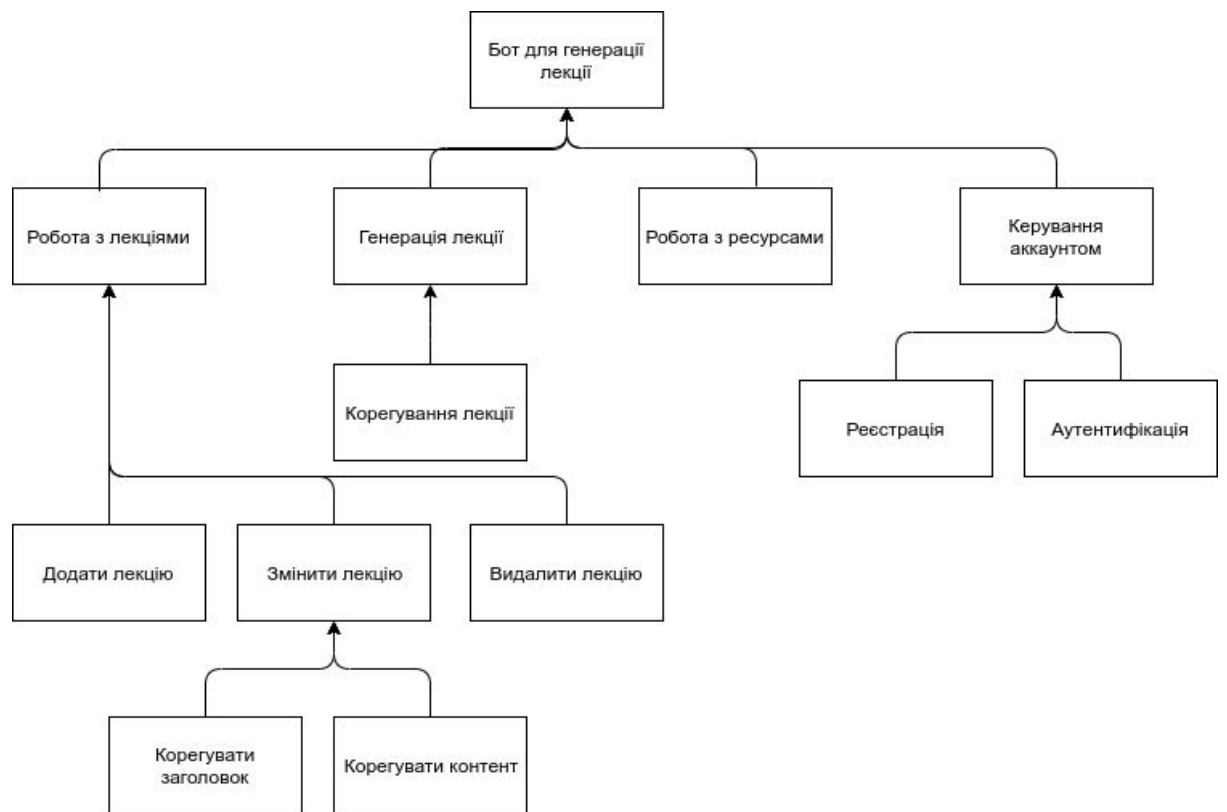
1. Роль користувача - ідентифікує права користувача на даному сервісі. За замовчуванням надається наступні ролі: адміністратор, модератор, преміум користувач та звичайний користувач. Атрибути: ідентифікатор, назва, пріоритет;
2. Користувач - власне набір sensitive-даних для користувача. Атрибути: логін, ідентифікатор ролі, хеш-пароллю, дата реєстрації;
3. Ресурс - посилання на текст. Атрибути: URL, назва, к-сть кліків;
4. Лекція - відформатований текст. Атрибути: ідентифікатор лекції, заголовок, контент, логін власника;

## Use-case



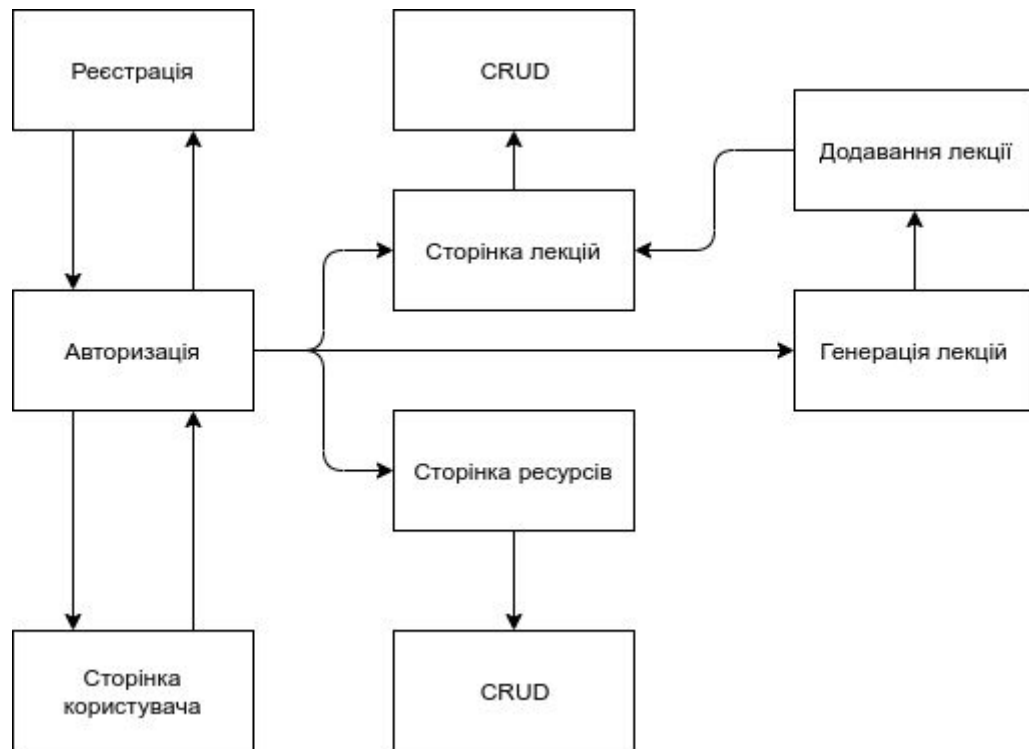
мал. 1 use-case diagram

## Component diagram



мал. 2 Component diagram

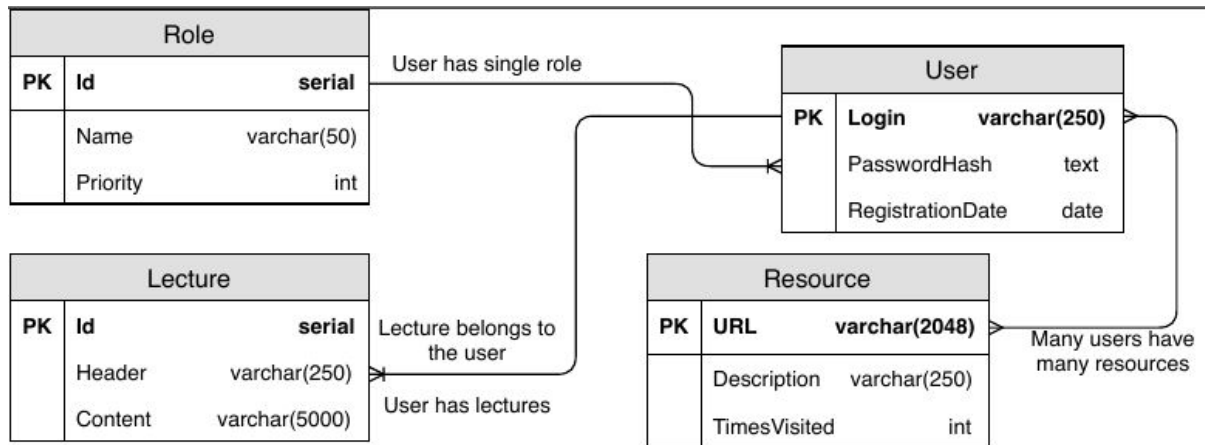
## UML



мал. 3 UML-diagram

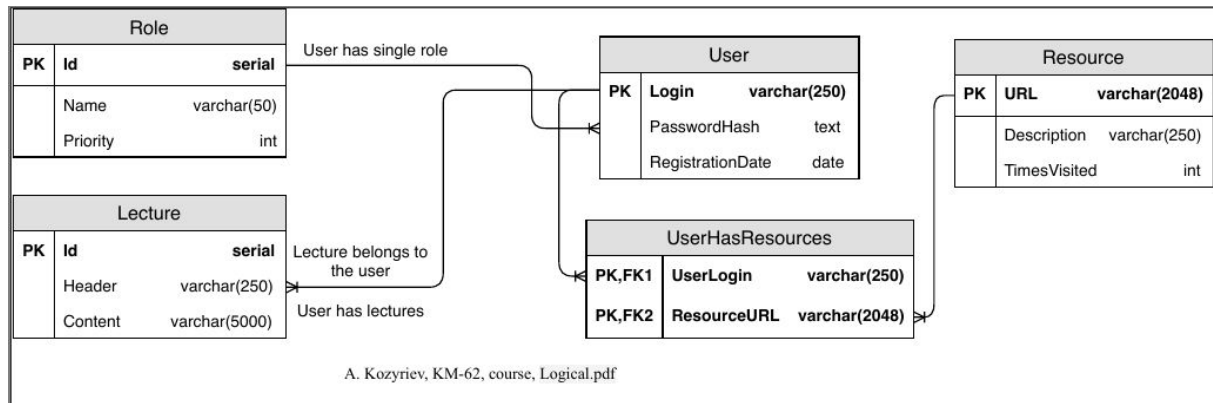
# ІНФОЛОГІЧНЕ ПРОЕКТУВАННЯ

## Моделі діючих сутностей



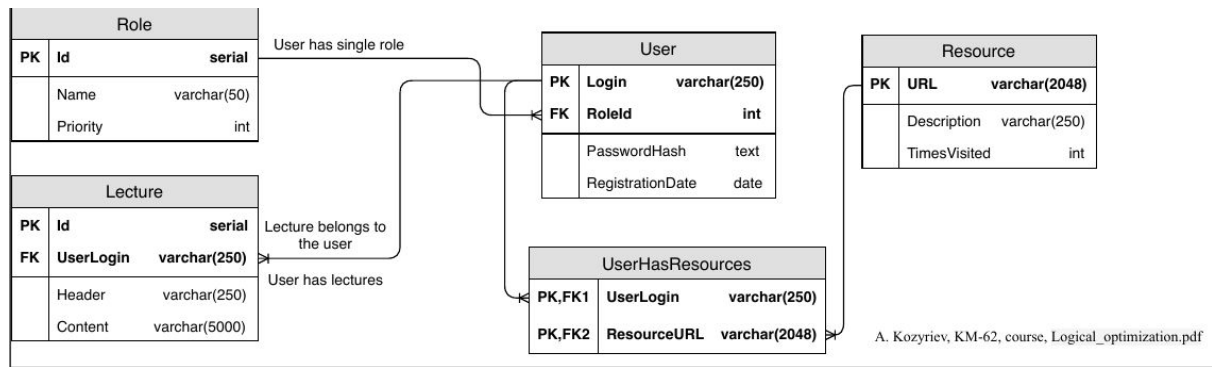
A. Kozyriev, KM-62, course, Conceptual.pdf

мал. 4 Conceptual ERD

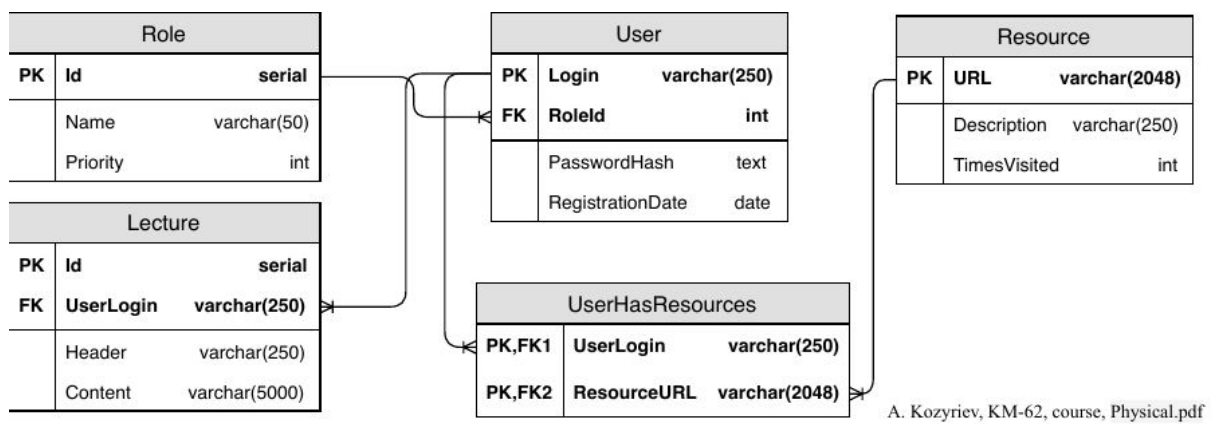


A. Kozyriev, KM-62, course, Logical.pdf

мал. 5 Logical ERD



мал. 6 Logical-Optimized ERD



мал. 7 Physical ERD

## ВИСНОВКИ

У ході даної роботи було створено додаток мікро сервісної архітектури, що генерує лекції за заданою темою. Додаток та його складові розміщено на серверах Heroku. Архітектура даного додатку є мікросервісною та архітектура виконання по REST API, що дає можливість масштабувати даний додаток без проблем. Було впроваджено механізм DI, що дало наступні переваги: тестування компонентів окремо, усі компоненти є слабо зв'язаними.

Стек використаних технологій:

1. Flask, SQL-alchemy, WTForms
2. Tensorflow 2.0
3. JQuery, Bootstrap
4. PostgreSQL 12



## СПИСОК ВИКОРИСТАНИХ ЛІТЕРАТУРНИХ ДЖЕРЕЛ

1. Neural Network Using Python and Numpy. Bernd Klein, Bodenseo;  
Design by Denise Mitchinson adapted for python-course.eu by Bernd Klein, 2011-2019,  
URL:[https://www.python-course.eu/neural\\_networks\\_with\\_python\\_numpy.php](https://www.python-course.eu/neural_networks_with_python_numpy.php)
2. WTForms Documentation, WTForms Team, documentation generated by Sphinx, 2010, URL: <https://wtforms.readthedocs.io/en/stable/>
3. Heroku Documentation, HEROKU Software IS A COMPANY 2019,  
URL:<https://devcenter.heroku.com/categories/reference>