

Розробка системи рекомендацій вакансій з використанням нейролінгвістичного програмування

ВАСИЛЬЄВА ТЕТЯНА, КН-4

Науковий керівник:
Ющенко Юрій Олексійович

Задача роботи

- Дослідити методи впровадження рекомендаційних систем
- Розробити систему рекомендацій вакансій за навичками користувача, що має назву Jobs by Skills.

Зміст роботи

Розділ перший.

Містить опис методів впровадження рекомендаційних систем, інформацію щодо технології NLP, та деяких показників подібності, які використовуються для дослідження в області рекомендацій.

Розділ другий.

Розповідається про методи, інструменти та технології, які використовувалися під час написання роботи: Фреймворк CRISP-DM, Spacy, Flask, Pandas, MongoDB.

Розділ третій.

Розглядається впровадження системи.

В кінці роботи наведені висновки та список використаної літератури.

Методи побудови рекомендаційних систем

- Рекомендаційна система - програма, яка намагається передбачити, які об'єкти будуть цікаві користувачеві, маючи певну інформацію про його профіль.
- Спільна фільтрація (collaborative filtering): ця техніка ґрунтується на оцінці елементу користувачем, і знаходженні подібності між різними користувачами, які дали подібні оцінки елементам.
- Фільтрація на основі вмісту (Content-based filtering): рекомендує користувачеві елементи, властивості яких схожі на елемент, до якого користувач раніше виявляв інтерес.
- Фільтрація на основі правил (Rule-based filtering): рекомендує на основі демографічного регіону користувача, історії покупок та інших атрибутів, які можна використовувати для профілювання користувача.
- Гібридна фільтрація (Hybrid Filtration): комбінація декількох технік фільтрації для покращення ефективності рекомендацій.

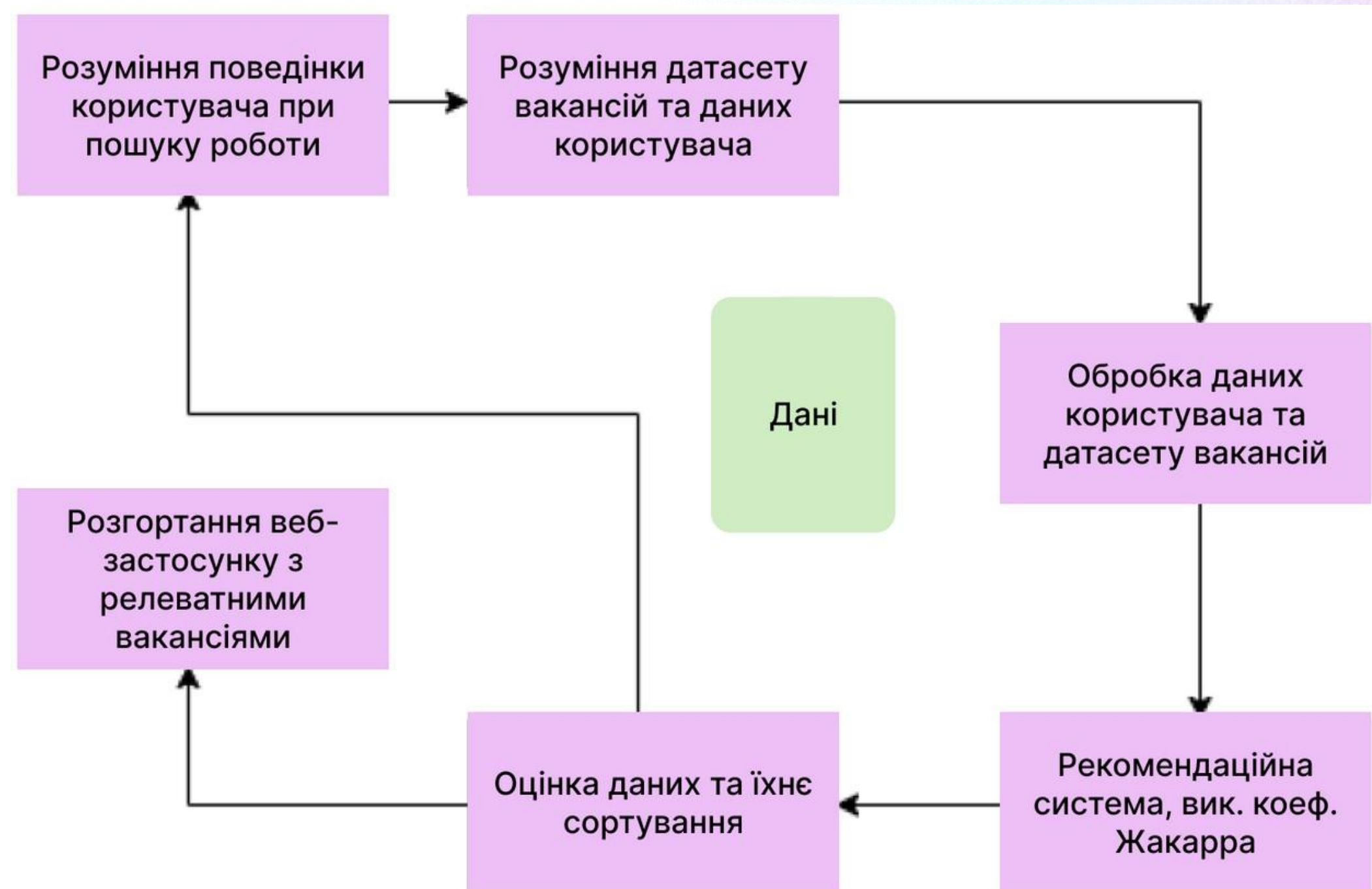
Нейролінгістичне програмування

- Обробка природної мови — це процес пошуку інформації з неструктурованих даних
- Коефіцієнт Жаккара — це метод порівняння елементів двох наборів даних, щоб визначити, які елементи є спільними, а які - різними. Це міра подібності для двох наборів даних із результатом від 0% до 100%. Два набори можна назвати схожими, коли результат близький до 100%.

$$\text{Індекс Жаккара} = \frac{\text{кількість елементів, спільних у двох наборах}}{\text{кількість елементів у двох наборах}}$$

Розробка системи - CRISP-DM

Для аналізу даних був використаний фреймворк CRISP-DM. Фреймворк CRISP-DM модифіковано до рівня, на якому він відповідає бізнес-вимогам розробки системи рекомендацій для шукача роботи.



Розробка системи - MongoDB

Для роботи системи необхідно було створити базу даних, використовуючи MongoDB. База даних використовується для зберігання обробленого датасету вакансій.

```
uri = "mongodb+srv://tanyavasylieva01:78Kf2f7o86vE82z6@cluste
client = MongoClient(uri, server_api=ServerApi('1'))

df = pd.read_csv("jobs_data_csv")

df['skills'] = df['skills'].apply(ast.literal_eval)
df['location'] = df['location'].apply(ast.literal_eval)

data = df.to_dict(orient="records")

db = client["jobs"]
db.records.insert_many(data)
```

Розробка системи - job data

Для отримання даних про вакансії використано датасет job_data_csv.csv

була виконана попередня його обробка та перетворення деяких даних

```
_id,experience,location,salary,search_type,skills,title,url
62ee2d1d4dcd52f533910d18,2-7 Yrs,"[""Chennai""]",Not disclosed,Technical Consultant,"[""Micros
62ee2d1d4dcd52f533910d19,2-7 Yrs,"[""Mumbai""]",Not disclosed,Technical Consultant,"[""OAF"","
62ee2d1d4dcd52f533910d1a,7-12 Yrs,"[""Chennai"",""Bangalore/Bengaluru""]",11,00,000 - 21,00,0
62ee2d1d4dcd52f533910d1b,2-7 Yrs,"[""Noida"",""Pune"",""Chennai"",""Bangalore/Bengaluru"",""De
",https://www.naukri.com/job-listings-oracle-cloud-technical-consultant-huron-eurasia-india-pr
62ee2d1d4dcd52f533910d1c,2-6 Yrs,"[""Hyderabad/Secunderabad"",""Bangalore/Bengaluru""]",4,00,
```

```
df = pd.read_csv("jobs_data_csv")

df['skills'] = df['skills'].apply(ast.literal_eval)
df['location'] = df['location'].apply(ast.literal_eval)

data = df.to_dict(orient="records")
```

Розробка системи - user data

Дані користувача про його навички отримуються безпосередньо шляхом введення інформації від користувача. Вони обробляються з використанням NLP, а саме функції токенізації.

```
def retrieve_user_skills(text):  
    nlp = spacy.load("en_core_web_sm")  
  
    doc = nlp(text)  
  
    type(doc)  
  
    dict = [token.text for token in doc]  
  
    return retrieve_vacancies(dict)
```

Розробка системи - вакансії

На малюнку наведена функція, що відповідає за отримання вакансій, релевантних навичкам користувача на основі коефіцієнту Жакарра. Спочатку отримуються всі релевантні вакансії, далі обраховується коефіцієнт ранжування та відбувається сортування вакансій за спаданням значення цього коефіцієнту.

```
def retrieve_vacancies(skills_list):
    length_skills_list = len(skills_list)
    num = mydb.find({"skills": {"$in": skills_list}})
    vacancies = []
    for i in num:
        vacancy_skills = i['skills']

        match = len([k for k, val in enumerate(vacancy_skills) if val in skills_list])

        total_length = len(vacancy_skills) + length_skills_list
        # ranking coefficient
        i['rank'] = match / total_length
        vacancies.append(i)

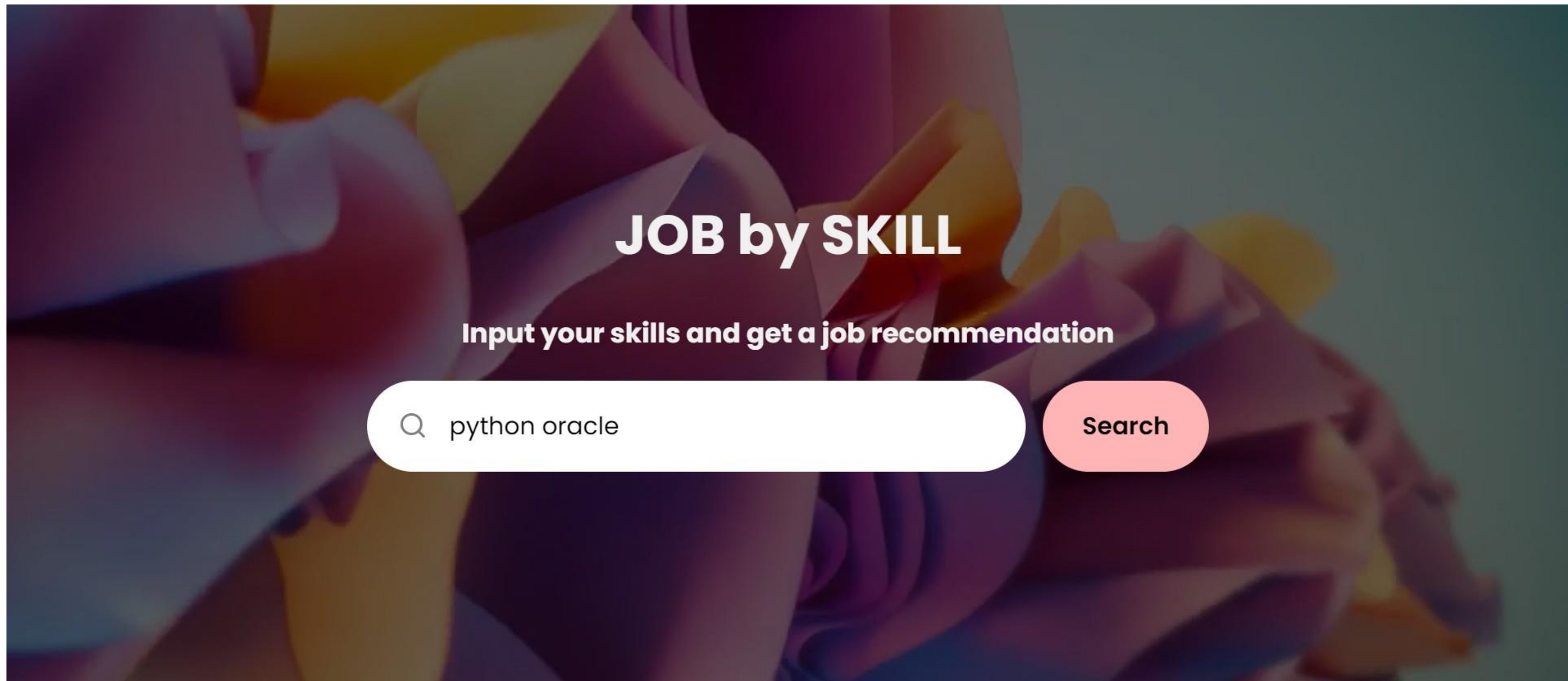
    return show_info(vacancies, skills_list, len(vacancies))
```

Розробка системи – відображення

Для відображення системи у веб-браузері, використовуються HTML, CSS та Flask.

```
@app.route('/')  
def start_page():  
    return render_template('start_page.html')  
  
@app.route('/', methods=['POST'])  
def post_skills():  
    text = request.form['text']  
  
    return retrieve_user_skills(text)
```

Результат



Початкова сторінка сервісу

Результат

The screenshot shows a web interface for job search results. At the top, there is a purple header with the text 'Jobs by Skills'. Below the header is a search bar with a magnifying glass icon and the text 'Enter your skills: python oracle'. The main content area has a light pink background with a grid pattern. It displays the text 'Your Skills : python oracle Total jobs found: 592'. Below this, there are two job listings. The first listing is for 'Lead Python Developer and Python Developer' with a score of 0.25. It includes details: 'Experience: 5-10 Yrs', 'Salary: 13,00,000 - 23,00,000 PA.', and 'Hyderabad/Secunderabad Bangalore/Bengaluru'. There are two skill tags: 'python' and 'javascript'. The second listing is partially visible, showing 'AWS Cloud Native Developer' with a score of 0.25.

Jobs by Skills

Enter your skills: python oracle

Your Skills : python oracle Total jobs found: 592

[Lead Python Developer and Python Developer](#) 0.25

Experience: 5-10 Yrs Salary: 13,00,000 - 23,00,000 PA. Hyderabad/Secunderabad Bangalore/Bengaluru

python javascript

[AWS Cloud Native Developer](#) 0.25

Сторінка з вакансіями

В И С Н О В К И

Було розроблено прототип системи рекомендацій вакансій за навичками користувача, що має назву *Jobs by Skills*, досліджено особливості побудови рекомендаційних систем, продемонстровано процес розробки, особливості архітектури сервісу та його роботи.

Можливі покращення отриманої системи:

- Додати більш повний аналіз тексту користувача, а саме можливість вводити навички у будь-якому форматі.
- Розбити отриманий список вакансій на сторінки та додати нумерацію.
- Додати фільтрацію вмісту.