

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Навчально-науковий центр заочної форми навчання

(повна назва)

Кафедра Інформаційних управляючих систем

(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА Пояснювальна записка

рівень вищої освіти другий (магістерський)

Дослідження методів формування рекомендацій вибору спеціальності для
абітурієнтів в ІС навчального закладу

(тема)

Виконав:

студент 2 курсу, групи ІУСТзм-22-1

Анастасія ЮРЧЕНКО

(Власне ім'я ПРІЗВИЩЕ)

Спеціальність 122 Комп'ютерні
науки

(код і повна назва спеціальності)

Тип програми освітньо-професійна

(освітньо-професійна або освітньо-наукова)

Освітня програма Інформаційні
управляючі системи та технології

(повна назва освітньої програми)

Керівник доц. каф ІУС Тетяна БІЛОВА

(Власне ім'я ПРІЗВИЩЕ)

Допускається до захисту

Зав. кафедри



(підпис)

Костянтин ПЕТРОВ

(Власне ім'я ПРІЗВИЩЕ)

2024 р.

Харківський національний університет радіоелектроніки

Навчально-науковий центр заочної форми навчання
 Кафедра Інформаційних управляючих систем
 Рівень вищої освіти другий (магістерський)
 Спеціальність 122 Комп'ютерні науки
 (код і повна назва)
 Тип програми освітньо-професійна
 (освітньо-професійна або освітньо-наукова)
 Освітня програма Інформаційні управляючі системи та технології
 (повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри

(підпис)

« 04 » грудня 20 23 р.

ЗАВДАННЯ

НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові Юрченко Анастасії Олександрівні
 (прізвище, ім'я, по батькові)

1. Тема роботи Дослідження методів формування рекомендацій вибору спеціальності для абітурієнтів в ІС навчального закладу

затверджена наказом університету від 01 грудня 2023 р. № 259Стз

2. Термін подання студентом роботи до екзаменаційної комісії 17 січня 2024 р.


3. Вихідні дані до роботи Науково-технічна література, публікації, інформація з інтернет-ресурсів

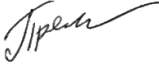
4. Перелік питань, що потрібно опрацювати в роботі Аналіз процесів профорієнтаційної роботи з абітурієнтами, огляд технологій, що використовуються для реалізації профорієнтаційної роботи з абітурієнтами, аналіз можливостей використання рекомендаційних систем для вибору спеціальності абітурієнтами, постановка задачі дослідження, метод колаборативної фільтрації для формування рекомендації вибору спеціальності у абітурієнтів, удосконалення методу для формування рекомендації вибору спеціальності у абітурієнтів, розробка інформаційної технології, перспективи подальших досліджень, обґрунтування вибору програмних засобів розробки, обґрунтування вибору платформи СУБД, реалізація інтерфейсу програми, експериментальна перевірка.

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Аналіз матеріалів з теми роботи	04.12.2023	Виконано
2	Формування рекомендації вибору спеціальності у абітурієнтів	16.12.2023	Виконано
3	Дослідження методу формування рекомендації вибору спеціальності у абітурієнтів	23.12.2023	Виконано
4	Інформаційна технологія для формування рекомендації вибору спеціальності у абітурієнтів	28.12.2023	Виконано
5	Реалізація інформаційної технології для формування рекомендації вибору спеціальності у абітурієнтів	04.01.2024	Виконано
6	Експериментальна перевірка	10.01.2024	Виконано
7	Підготовка пояснювальної записки та графічного матеріалу	14.01.2024	Виконано
8	Захист перед екзаменаційною комісією	19.01.2024	Виконано

Дата видачі завдання 04 грудня 2023 р.

Студент 
(підпис)

Керівник роботи  доц. каф ІУС Тетяна БІЛОВА
(підпис) (посада, власне ім'я, прізвище)

РЕФЕРАТ

Пояснювальна записка містить: 88 с., 25 рис., 7 табл., 1 дод., 37 джерела.

АБИТУРІЄНТ, ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ, КОЛАБОРАТИВНА ФІЛЬТРАЦІЯ, МЕТОД, ПРОФОРІЄНТАЦІЯ, РЕКОМЕНДАЦІЙНА СИСТЕМА.

Об'єкт дослідження є процес побудови рекомендацій вибору спеціальності абітурієнтам, котрі здобули повну загальну середню освіту і мають на меті вступити на навчання до закладів освіти.

Предмет дослідження – методи формування рекомендацій вибору спеціальності для абітурієнтів.

Мета роботи – аналіз методів формування рекомендацій вибору спеціальності і проектування системи рекомендацій для абітурієнтів, що вступають до закладів освіти.

Розроблено систему на основі метода колаборативної фільтрації та експериментально перевірено її роботу.

ABSTRACT

The explanatory note contains: 88 p., 25 fig., 7 tab., 1 ann., 37 sources.

APPLICANT, CAREER ORIENTATION, COLLABORATIVE
FILTERING, INFORMATION TECHNOLOGY, METHOD,
RECOMMENDATION SYSTEM,

The object of the research is the process of building recommendations for choosing a specialty for applicants who have completed general secondary education and intend to enroll in educational institutions.

The subject of the study is the methods of forming recommendations for the choice of specialty for applicants.

The purpose of the work is to analyze the methods of forming recommendations for choosing a specialty and designing a system of recommendations for applicants entering educational institutions.

A system based on the method of collaborative filtering was developed and its operation was tested experimentally.

ЗМІСТ

Скорочення та умовні позначення.....	8
Вступ.....	9
1 Формування рекомендації вибору спеціальності у абітурієнтів.....	11
1.1 Аналіз процесів профорієнтаційної роботи з абітурієнтами.....	11
1.2 Огляд технологій, що використовуються для реалізації профорієнтаційної роботи з абітурієнтами.....	18
1.3 Аналіз можливостей використання рекомендаційних систем для вибору спеціальності абітурієнтами.....	23
1.4 Постановка задачі дослідження.....	33
2 Дослідження методу формування рекомендації вибору спеціальності у абітурієнтів.....	35
2.1 Метод колаборативної фільтрації для формування рекомендації вибору спеціальності у абітурієнтів.....	35
2.2 Удосконалення методу для формування рекомендації вибору спеціальності у абітурієнтів.....	38
3 Інформаційна технологія для формування рекомендації вибору спеціальності у абітурієнтів.....	41
3.1 Розробка інформаційної технології.....	41
3.2 Перспективи подальших досліджень.....	50
4 Реалізація інформаційної технології для формування рекомендації вибору спеціальності у абітурієнтів.....	51
4.1 Обґрунтування вибору програмних засобів розробки.....	51
4.2 Обґрунтування вибору платформи СУБД.....	53
4.3 Реалізація інтерфейсу програми.....	56
4.4 Експериментальна перевірка.....	62

Висновки.....	68
Перелік джерел посилання.....	69
Додаток А Графічний матеріал.....	73

СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАКИ

РС - рекомендаційна система

ALS - alternating least squares

NCF - neural collaborative filtering

SVD - singular-value decomposition

БД - база даних

ЕС - експертна система

ПрО - предметна область

ПЗ – програмне забезпечення

СУБД - система управління базами даних

СППР – система підтримки прийняття рішень

ВСТУП

Вибір професії — один із головних у житті людини. По суті, він є вибором життєвого шляху, долі. Від нього багато залежить, наскільки людина зможе реалізувати себе в різних сферах життєдіяльності, якими будуть її власні набутки та віддача від неї суспільству. Тому вибір професії — точка, в якій схрещуються інтереси особистості та суспільства. Правильний вибір професії — це моральне задоволення, висока самооцінка. Водночас це й висока продуктивність праці, висока якість продукції. Вибір професії — точка, в якій сходяться інтереси особистості та суспільства, де можливе й необхідне поєднання особистих і загальних інтересів. Для кожної людини вибір професії — це вибір долі. І загальноосвітня школа має допомогти кожному вихованцеві обрати доцільну для нього професію.

У країнах із розвиненою ринковою економікою ідея успішної кар'єри буквально пронизує все суспільне життя. З одного боку, доведено, що профілізація, що базується на виборі професії, є одним із етапів проектування безперервної освіти як повноцінного та поступального процесу розвитку особистості. З іншого боку, соціологічні дослідження підтверджують, що правильний вибір професії вдвічі зменшує плинність кадрів, у півтора рази знижує вартість витрат на перекваліфікацію працівників, на 10-15% збільшує продуктивність праці.

Отже, суспільство розуміє, що, допомагаючи кожному громадянину у процесі професійного самовизначення, воно вирішує як особистісні завдання окремої людини, а й важливі соціально-економічні проблеми у масштабах держави. У умовах профорієнтаційна робота, спрямовану створення умов для раннього професійного самовизначення школярів, стає очевидним суспільним благом, досягнення якого у сфері суспільства з участю державних заходів впливу стає дедалі явною потребою.

З метою самовизначення абітурієнтів навчальні заклади, центри зайнятості і державні служби постійно проводять профорієнтаційну роботу, але в умовах війни це стало складно робити результативно за класичними методами. Профорієнтаційна допомога повинна задовольняти потреби школярів на різних рівнях. Необхідно проводити профорієнтаційну роботу різними методами, наприклад: тренінги, семінари, курси, майстер - класи, які надають груповий досвід або індивідуальні консультації фахівців.

Оскільки період воєнного стану може тривати невизначений час, профорієнтаційні заходи повинні бути гнучкими та адаптованими до змінних умов. Таким чином, профорієнтація у період воєнного стану є надзвичайно важливою темою, яка стосується багатьох людей. Вона дозволяє абітурієнтам зробити свідомий вибір професії, що відповідає їхнім здібностям та інтересам, а також забезпечує можливість пристосування до змінних умов життя на ринку праці.

1 ФОРМУВАННЯ РЕКОМЕНДАЦІЇ ВИБОРУ СПЕЦІАЛЬНОСТІ У АБІТУРІЄНТІВ

1.1 Аналіз процесів профорієнтаційної роботи з абітурієнтами

Профорієнтація, тобто орієнтація на професію, - це процес виявлення в людини схильностей до певного виду діяльності.

Як зазначено в Концепції державної системи професійної орієнтації населення [1], профорієнтаційну роботу в найповнішому обсязі здійснюють підрозділи Державної служби зайнятості. Однак вагоме значення для розвитку конкурентного ринку праці має проблема професійного самовизначення абітурієнтів навчальних закладів. Таким чином, одним із пріоритетних завдань сучасної системи освіти є організація роботи з учнівською молоддю щодо усвідомленого професійного самовизначення.

В основу профорієнтаційної діяльності покладено такі принципи:

- комплексний характер профорієнтаційних послуг: організація профорієнтаційної діяльності повинна здійснюватись в соціально-економічному, психолого-педагогічному, медико-фізіологічному та професійному напрямках для забезпечення різнобічної готовності молодої людини до вибору професії;
- узгодження інтересів особи та суспільства через ринок праці;
- діяльнісний підхід до визначення професійної придатності;
- рівні можливості отримання профорієнтаційних послуг громадянами України та громадянами іноземних держав, з якими укладено відповідні угоди, незалежно від місця роботи чи навчання, віку, статі, національності, релігійних переконань;
- доступність професійної та іншої інформації стосовно можливостей вибору чи зміни професії, форм навчання та працевлаштування;
- добровільність і безкоштовність отримання профорієнтаційних послуг усіма групами населення на гарантованому державному рівні;

– конфіденційний і рекомендований характер висновків профконсультацій і профдобору, додержання працівниками профорієнтаційних норм професійної етики;

– взаємозв'язок школи, сім'ї, виробництва і громадськості в здійсненні профорієнтаційних заходів;

– систематичність і наступність у профорієнтації [2].

У сучасних наукових і нормативно-методичних документах вирізняються різноманітні підходи щодо структурного складу профорієнтації:

– триетапна структура профорієнтації Д. Тхоржевського [3] містить профосвіту, профконсультацію, й працевлаштування;

– чотирикомпонентна (Б. Федоришина) [4] передбачає організаційно-функціональну, логіко-змістовну, особистісну й управлінську підсистеми;

– семикомпонентна (В. Васильков): профосвіта, профвиховання, профактивація, професійна психодіагностика, профконсультація, профвідбір, профадаптація [5].

Структура профорієнтаційної роботи приведена на рисунку 1.1.

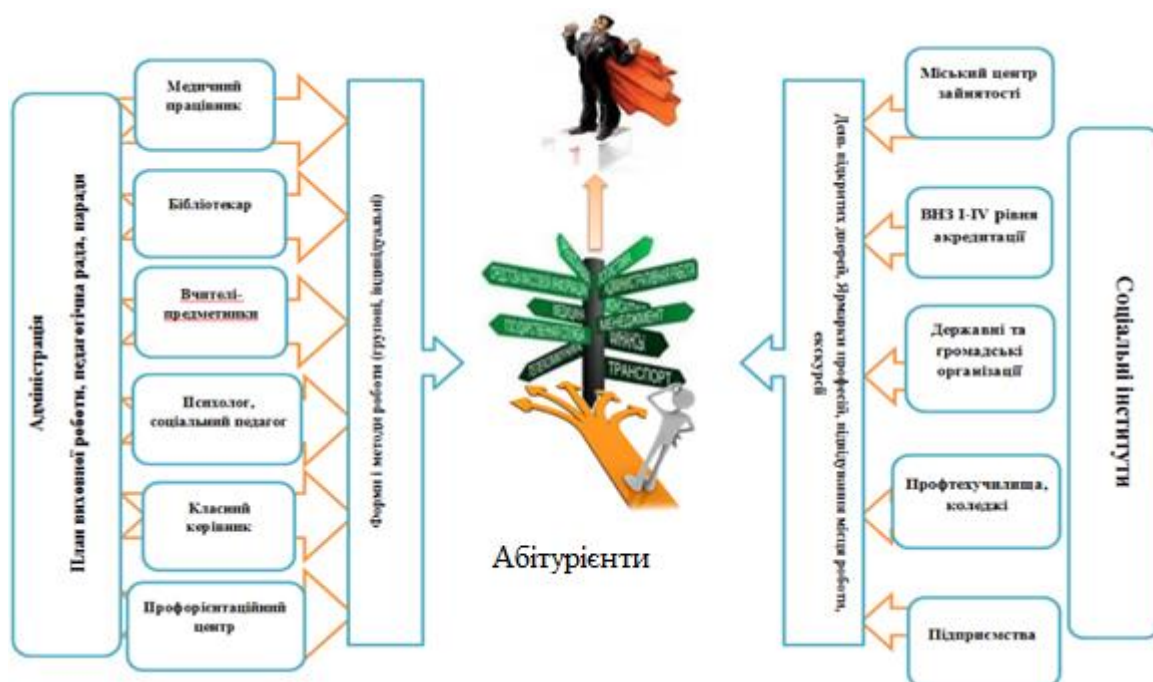


Рисунок 1.1 – Структура профорієнтаційної роботи

Перед тим, як обрати професію, необхідно тверезо оцінити свої сили і можливості за такими критеріями: хочу, можу, треба. Розподіл факторів оцінки своїх сил приведена на рисунку 1.2, оптимальна зона вибору в залежності від факторів зображена на рисунку 1.3.



Рисунок 1.2 – Оцінка своїх сил за критеріями



Рисунок 1.3 – Оптимальний вибір професії

Мотив – це те, що змушує людину діяти. Тоді якими можуть бути мотиви вибору професії? Виділяють зовнішні та внутрішні мотиви. Зовнішні мотиви зумовлені впливом навколишнього соціального середовища, а внутрішні мотиви пов'язані з вашим характером, здібностями, схильностями, звичками. Основні мотиви, що спонукають підлітків обрати ту чи іншу професію, - це престиж, високий рівень зарплати, кар'єрне зростання, інтерес до професії, умови роботи, можливості подальшої освіти.

На формування мотивів вибору професії впливають різноманітні фактори:

- сфера дружніх відносин;
- засоби масової інформації; художня та наукова література;
- педагогічний колектив: класний керівник, учитель, дирекція;
- сім'я, близькі, родичі та ін.

Критерії професійного самовизначення, тобто вибору професії, можуть бути різними. Їх різновиди приведені на рисунку 1.4.



Рисунок 1.4 – Критерії професійного самовизначення

Компоненти профорієнтаційної роботи:

- професійна інформація;
- професійна діагностика;
- професійна консультація;
- професійний відбір;
- професійна адаптація.

Схарактеризуємо їх докладніше [6].

Професійна інформація – це психолого-педагогічна система, яка реалізує суб'єктивні та об'єктивні умови вільного і свідомого професійного самовизначення, формуючи особистості позицію активної профорієнтації. Вона дає можливість абітурієнтам познайомитися з різними професіями, характером праці в кожній професії, вимогами, які ці професії ставлять до

людини, та місцями, де цими професіями можна оволодіти. З її допомогою абітурієнти можуть виявити власні здібності, психофізіологічні особливості та вміння, порівняти їх з вимогами професій, які їх цікавлять і до яких вони підходять, і прийняти правильне рішення про свою придатність до професійної діяльності в обраній професійній сфері.

Професійна діагностика – це система психологічного вивчення особистості з метою виявлення професійно важливих рис та якостей.

Професійна консультація – це надання науково обґрунтованої допомоги людям на основі досліджень щодо найбільш підходящих сфер та інструментів для професійної автономії. Види професійної консультації наведені на рисунку 1.5.



Рисунок 1.5 – Види професійної консультації

Профконсультації є кількох видів: довідково - організаційні (дають відповіді абітурієнтам та їх батькам на питання про можливості отримання певної професії в тому чи іншому навчальному закладі, про можливості оволодіння певною професією залежно від психолого-фізіологічних можливостей); медико-профілактичні (передбачають виявлення психолого-фізіологічних і медико-біологічних особливостей вихованців і їх відповідність бажаній професії); психолого-педагогічні (сприяють вибору професії відповідно до інтересів і здібностей вихованців).

Професійний відбір – система роботи, спрямована на надання допомоги учневі у визначенні й виборі конкретної професії на основі виявлення й оцінки його загальних і спеціальних здібностей, здатностей, інтересів, потреб і об'єктивних умов професійної підготовки і працевлаштування.

Професійна адаптація - процес пристосування людини до професійної діяльності, її умов; досягнення бажаної продуктивності праці й відповідності між професійними намірами, інтересами, якостями особистості та вимогами до діяльності.

Помилки в ході профорієнтації [7]:

- поділ професій на «престижні» й «непрестижні»;
- ототожнення навчального предмета з професією;
- перенесення ставлення до людини, представника професії, на саму професію;
- вибір професії під чийсь впливом;
- застарілі уявлення про характер праці у сфері матеріального виробництва.

В сучасних реаліях воєнного стану профорієнтаційна робота перемістилась в онлайн простір з залученням сучасних інформаційних систем. Найпростішою і мобільною формою стали онлайн тести для визначення професійної спрямованості.

Психологи класифікують професії на п'ять типів [8]. Якщо головними (провідними) об'єктами праці є рослини, тварини або мікроорганізми, то це

професії типу «людина - природа». Якщо основним (провідним) об'єктом праці є технічна система, матеріальний об'єкт, матеріал або енергія, то це професії типу «людина - техніка». Тип «людина - знакова система» характеризується такими основними об'єктами праці, як символи, числа, коди, природні та штучні мови. Професії типу «людина – художній образ» пов'язані зі створенням художніх образів, а в типі «людина – людина» об'єктом праці є людина, група, колектив або спільнота людей.

До типу «людина – знакова система» відносяться професії: бібліотекар, бібліограф, листоноша, патентознавець, історик, лінгвіст, фізик, фармацевт, хімік, економіст, картограф, інженер-програміст, математик-програміст, веб-дизайнер, SEO-спеціаліст, аудитор, контент-менеджер, копірайтер, медіапланер та інші.

До типу «людина – техніка» відносяться професії: машиніст, слюсар, електрик, радіоінженер, штукатур, столяр, швачка, пілот, космонавт, сервісний інженер та інші.

До типу «людина – природа» належать такі професії: агроном, садівник, зоотехнік, гідробіолог, еколог, генетик, гірний інженер, гідролог, технік - топограф, селекціоніст, зоостиліст та інші.

До типу «людина – художній образ» належать такі професії: артист, вокаліст, диригент, педагог, художник, ілюстратор, візажист, кліпмейкер та інші.

До типу «людина – людина» належать такі професії: вихователь, викладач, лікар, адвокат, гід-перекладач, організатор культурного дозвілля, підприємець, менеджер, журналіст, сімейний лікар, інтернет-коуч, івент-менеджер, PR-менеджер та інші.

За останні роки особливо популярними в освіті стали професії типу «людина – знакова система». Для спеціалістів даного типу професій потрібна наявність певного спектру навичок у абітурієнтів.

1.2 Огляд технологій, що використовуються для реалізації профорієнтаційної роботи з абітурієнтами

У сучасному світі в умовах перевантаженості користувачів інформаційних систем величезним кількістю різноманітних інформацією про товари, послуг, подій тощо, все більш важливу роль починають грати системи рекомендацій або рекомендаційні системи.

Фільтрування інформації – це процес моніторингу великих обсягів динамічно генерованих даних та виявлення лише того набору цих даних для користувача, який найбільш точно відповідає його потребам.

Загалом, системи рекомендацій являють собою допоміжні системи, які допомагають користувачам знаходити інформацію, продукти або послуги (такі як книги, фільми, музику, цифрові продукти, веб-сайти, телепрограми та ін.). Найчастіше це робиться за допомогою узагальнення та аналізу переваг інших користувачів (що означає розгляд різних аспектів їх поведінки), та атрибутів (характерних особливостей) поточного користувача [9].

З іншого боку, завдання системи рекомендацій - це завдання сфери машинного навчання. У машинному навчанні більшість завдань передбачає навчання функції, яка буде обробляти дані. Для створення такої функції необхідний набір аргументів і значення функції від цих аргументів. В ідеалі, чим більший набір даних, тим точнішою повинна бути функція, але на практиці ми стикаємося з такими проблемами, як шум даних та неправильно підібраний набір аргументів, що спричиняє проблеми перетренування (overfitting) та недотренування (underfitting). Ці проблеми до кінця не вирішені в рекомендаційних системах. На рисунку 1.6 можна побачити принцип функціонування рекомендаційної системи.

Системи рекомендацій перетворилися на фундаментальний інструмент для прийняття більш інформативних, ефективних та дієвих рішень практично

у всіх галузях сучасного життя, включаючи бізнес, фінанси, охорону здоров'я, транспорт, освіта, спілкування, розваги та ін. [10].



Рисунок 1.6 – Принцип роботи РС

Користувач взаємодіє з системою через її інтерфейс, яких поділяється на вхідний та вихідний. Вхідним інтерфейсом рекомендаційної системи називається сукупність точок взаємодії між користувачем і системою, через які система отримує персоналізовані дані користувача. Введемо поняття рейтингу. Рейтинг – якісна оцінка продукту користувачем, що, зазвичай, виражається числовим значенням. Наприклад, користувач може залишити оцінку відео, яке він переглянув, або залишити коментар під продуктом який він купив. Система отримує персональну інформацію від користувача двома способами: явним (explicit) і неявним (implicit). Вище наведені приклади явного рейтингу, адже користувач ціленаправлено залишає оцінку, або коментар. Прикладом неявного рейтингу може послугувати ситуація, коли користувач почав дивитися відео і закінчив перегляд за хвилину. Звідси система може зрозуміти, що відео користувачу не сподобалося, адже він завершив перегляд за хвилину [11]. Вихідним інтерфейсом рекомендаційної системи називається сукупність точок взаємодії між користувачем і системою,

через які користувач отримує рекомендації від системи. Існує поняття передбаченого рейтингу, що визначає з певною ймовірністю передбачену оцінку деякого продукту користувачем. Рекомендацією називають кінцевий набір предметів, який з певною ймовірністю буде до вподоби користувачу. Цей набір може бути відсортований у порядку спаду рейтингу.

В даний час існують різні підходи до генерування рекомендацій у системах рекомендацій, такі як генерування рекомендацій на основі контенту (Content-based recommender systems) [12], спільної фільтрації (Collaborative filtering recommender systems) [13], використання гібридного підходу (Hybrid recommender systems) [14], що об'єднує в собі підходи на основі контенту та спільної фільтрації та інші.

Суть content-based підходу (рисунок 1.7) в тому, що по історії дій користувача створити для нього вектор його вподобань у просторі предметів і видавати результати близькі до цього вектору.

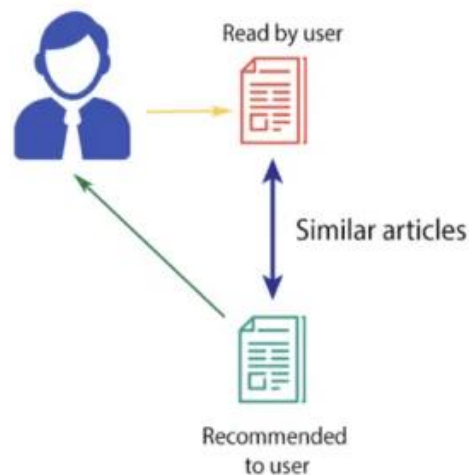


Рисунок 1.7 – Схема content-based підходу

У Collaborative filtering recommender systems (рисунок 1.8) наявна матриця оцінок користувачів. Ідея в тому щоб для кожного користувача

знайти найбільш схожих «сусідів» і заповнивши пропуски конкретного користувача, виважено усереднити рейтинги «сусідів».

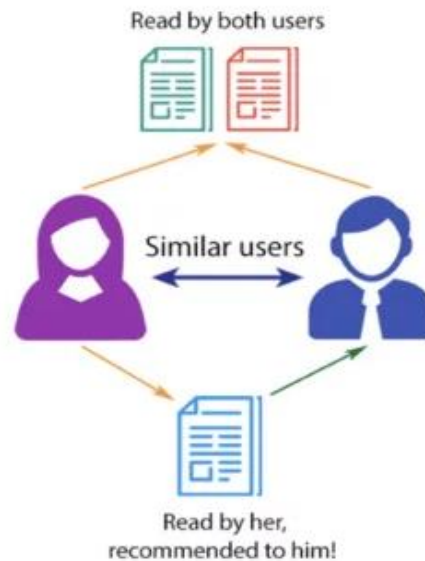


Рисунок 1.8 – Схема collaborative filtering підходу

Усі описані раніше типи мають певні недоліки. Комбінування кількох алгоритмів у межах однієї платформи дозволяє якщо не усунути їх повністю, то хоча б мінімізувати. В цьому і полягає суть Hybrid recommender systems.

Великі сервіси та інтернет-магазини використовують гібридні варіанти. Трохи там, дещо тут і виходить унікальна система. Універсальної інструкції та рекомендацій щодо реалізації такого інструменту немає. Все обмежується можливостями та фантазією розробників.

Є кілька поширених типів комбінування:

- реалізація окремо колаборативних та контентних алгоритмів та поєднання їх припущень;
- включення деяких контентних правил до колаборативної методики;
- включення деяких колаборативних правил до контентної методики;
- побудова загальної моделі, що включає правила обох методик.

Зазвичай ці варіанти беруть як основу і доповнюють за власним бажанням та за критеріями сфери діяльності. Як і у випадку з knowledge-based, основним недоліком гібридних систем є складність розробки.

Рекомендаційні системи працюють на двох рівнях [15]:

– глобальні оцінки; особливості та переваги, що не змінюються місяцями або роками; цікаві сторінки; залежність від характерних рис користувача: стать, місце проживання і т.п.;

– короткочасні тренди та швидкі зміни інтересів у часі.

Дані збирають «явним» та/або «неявним» способами. У першому випадку відвідувачу пропонують заповнювати анкети, проходити опитування тощо. визначення його переваг. Метод ефективний, лише користувачі не завжди погоджуються на заповнення анкет та проходження опитувань. Тому вигадують, як зацікавити їх цим (зазвичай пропонують якусь зиск в обмін).

Другий метод передбачає фіксування поведінки споживача на сайті або у додатку: які сторінки/розділи подивився, що додавав до кошика, які та де залишав коментарі тощо. За правильної організації збору даних та подальшої аналітики метод дає хороші результати. Але і тут є низка складнощів. По-перше, перегляд окремих сторінок чи карток товару не дає відомостей, чи сподобалося зрештою людині пропозицію. По-друге, навіть якщо покупка, складно припустити, що безпосередньо підштовхнуло людини до її вчиненню.

Як і у випадку з типами рекомендаційних систем окремо способи використовують рідко. Для отримання кращих результатів їх комбінують.

Грамотно налаштований збір інформації дозволяє зробити рекомендації релевантними. З їхньою допомогою скорочується час пошуку необхідних товарів чи послуг, і навіть підвищується можливість здійснення супутніх цільових дій. Наприклад, споживач хотів купити лише смартфон, але в результаті додатково придбав чохол для нього. Коли споживач витрачає мінімум часу для пошуку потрібних речей та отримує бажане, його лояльність до платформи підвищується.

Ще один важливий момент у роботі рекомендаційних систем, який варто враховувати, – дотримання конфіденційності. Вони можуть спрогнозувати такі результати та виявити такі закономірності, про які користувач навіть не замислювався або не хотів, щоб це стало комусь відомо.

Щодо сфер застосування рекомендаційних систем, то тут, по суті, жодних обмежень немає. Раніше ми вже неодноразово говорили, що все залежить від фантазії розробників. Їх можна використовувати у додатках, інтернет-магазинах, онлайн-кінотеатрах та інших сервісах, які пропонують якийсь контент, товари чи послуги.

Будь-яка компанія, навіть невеликий інтернет-магазин диванів із регіону, має використовувати рекомендаційні системи. Вони допоможуть у утриманні поточних споживачів, залученні нових та підвищенні прибутку.

1.3 Аналіз можливостей використання рекомендаційних систем для вибору спеціальності абітурієнтами

Рекомендаційні системи (РС) – це програмні інструменти та методи, які пропонують можливі списки пропозицій, стосовно саме тих товарів, що скоріш за все зацікавлять конкретного користувача. Списки пропозицій стосуються різних процесів прийняття рішень, стосовно того наприклад який фільм обрати для перегляду, яку музику обрати для прослуховування, чи який телефон придбати.

Зазвичай, персоналізовані рекомендації надаються користувачеві у вигляді упорядкованих списків предметів. При побудові рекомендацій, РС намагаються передбачити, які продукти чи послуги є найбільш підходящими, виходячи з уподобань і обмежень користувача.

Як би не було, фідбек користувачів є найважливішим видом інформації, який можна враховувати в рекомендаціях. Базові методи рекомендацій побудовані на методі найближчих сусідів. Для побудови рекомендації ми знаходимо користувачеві близьких за перевагою "сусідів" та усереднюємо їх оцінки. Або знаходимо об'єкти схожі на ті, які вже сподобалися користувачеві. Проблема в тому, що матриця колаборативних даних, як уже згадувалося, велика та розріджена, шукати сусідів незручно.

Тому постало завдання відобразити в єдине векторне місце і користувачів, і об'єкти, щоб мати можливість безпосередньо порівнювати їх схожість. Векторний простір, про який мова йде, зазвичай називають «прихованими факторами» (latent factors) або ембеддингами [16]. Шукається розкладання матриці взаємодій таке, щоб добуток двох матриць (користувачі-ембеддинги та об'єкти-ембеддинги) було близько до вихідної матриці взаємодій у тих позиціях, які були заповнені (відсоток заповнення матриці низький). Відповідно, скалярне твір вектора користувача вектор контенту дає передбачення ступеня інтересу користувача до цього контенту. Схема матричного розкладання приведена на рисунку 1.9.

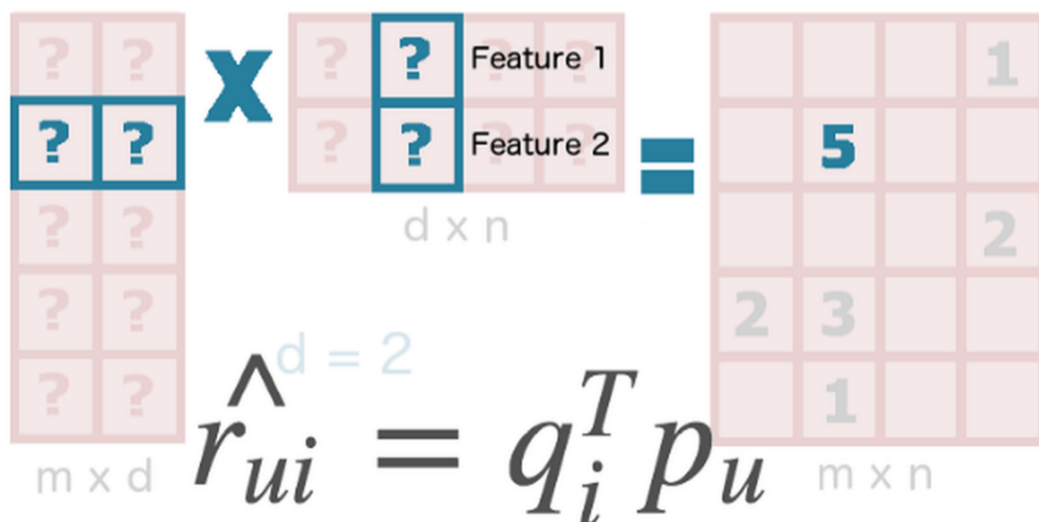


Рисунок 1.9 – Схема методу матричного розкладання

Сформулювати завдання можна як оптимізацію квадратичної помилки із регуляризацією:

$$\sum (R_{ui} - q_i * p_u)^2 + Regularize(q_i + p_u). \quad (1.1)$$

Але можна трохи уточнити завдання. Цьому фільму частіше ставлять оцінку десятки, а цей користувач, припустимо, дуже добрий і взагалі багатьом фільмам ставить десятки. Потрібно відокремити упередженість користувача та контенту від персональної переваги. Тому формулу часто уточнюють:

$$\sum (R_{ui} - q_i * p_u - b - w_i - w_u)^2 + Regularize(q_i + p_u + w_i + w_u). \quad (1.2)$$

Тут є три нових елементи: глобальна середня оцінка та зміщення щодо неї для конкретного користувача та об'єкта.

Alternating least squares (ALS) — метод отримання матриць, що нас цікавлять. Складність у тому, що потрібно оптимізувати одразу два набори параметрів:

$$L = \sum_{u,i \in S} (r_{ui} - \mathbf{x}_u^T \cdot \mathbf{y}_i)^2 + \lambda_x \sum_u \|\mathbf{x}_u\|^2 + \lambda_y \sum_u \|\mathbf{y}_i\|^2. \quad (1.3)$$

Алгоритм влаштований так. Фіксується один із наборів прихованих факторів, наприклад об'єктів (y_i). Обчислюється похідна функції втрат за іншим набором векторів (користувачів). Потім процедура повторюється для об'єктів векторів. І так чергування продовжується.

$$\frac{\partial L}{\partial \mathbf{x}_u} = -2 \sum_i (r_{ui} - \mathbf{x}_u^\top \cdot \mathbf{y}_i) \mathbf{y}_i^\top + 2\lambda_x \mathbf{x}_u^\top, \quad (1.4)$$

$$0 = -(\mathbf{r}_u - \mathbf{x}_u^\top Y^\top) Y + \lambda_x \mathbf{x}_u^\top, \quad (1.5)$$

$$\mathbf{x}_u^\top (Y^\top Y + \lambda_x I) = \mathbf{r}_u Y, \quad (1.6)$$

$$\mathbf{x}_u^\top = \mathbf{r}_u Y (Y^\top Y + \lambda_x I)^{-1}, \quad (1.7)$$

$$\frac{\partial L}{\partial \mathbf{y}_i} = -2 \sum_u (r_{iu} - \mathbf{y}_i^\top \cdot \mathbf{x}_u) \mathbf{x}_u^\top + 2\lambda_y \mathbf{y}_i^\top, \quad (1.8)$$

$$0 = -(\mathbf{r}_i - \mathbf{y}_i^\top X^\top) X + \lambda_y \mathbf{y}_i^\top, \quad (1.9)$$

$$\mathbf{y}_i^\top (X^\top X + \lambda_y I) = \mathbf{r}_i X, \quad (1.10)$$

$$\mathbf{y}_i^\top = \mathbf{r}_i X (X^\top X + \lambda_y I)^{-1}, \quad (1.11)$$

де $n \times m$ – розмір вихідної матриці;

Y – вектора ембеддингів контенту;

\mathbf{r}_u – рядок користувача u з матриці рейтингів з усіма рейтингами для всіх елементів (має розмірність $1 \times m$);

I – це поодинокі матриця, розмірність якої тут $k \times k$.

Так, званий холодний старт – проблема рекомендацій для нового контенту або користувача. Припустимо, ми оновлюємо рекомендації разів на три години. За ці три години може прийти новий юзер. Або вже присутні користувачі можуть зробити нові події, на які ми хочемо реагувати швидше (за три години активний користувач може подивитися багато роликів, а рекомендації залишаться колишніми)

Цей простий трюк зазвичай називають "fold-in". По суті, ми робимо один крок оптимізації ALS тільки для одного нового вектора користувача, фіксуючи матрицю прихованих факторів контенту.

Холодний старт може стосуватися і об'єктів, якщо до нього ще не було достатнього інтересу користувачів, щоб зробити прогноз. Таке можливо, якщо об'єкт тільки з'явився, або може бути винна кільцева проблема

непопулярності: поки контент не популярний, його не рекомендують, а значить, йому немає звідки взяти популярність для рекомендацій.

Отже, потрібно отримати вектор ембедингів контенту, за іншими даними. Можна використовувати додаткову інформацію. Наприклад, існує прийом, коли будують модель, що передбачає ембедінги за метаданими контенту. Зрозуміло, її доведеться переучувати при кожному оновленні основної моделі. Або якщо у нас є ембедінги отримані іншим чином (наприклад із самого контенту безпосередньо), їх також можна використовувати для такого передбачення (рисунок 1.10).

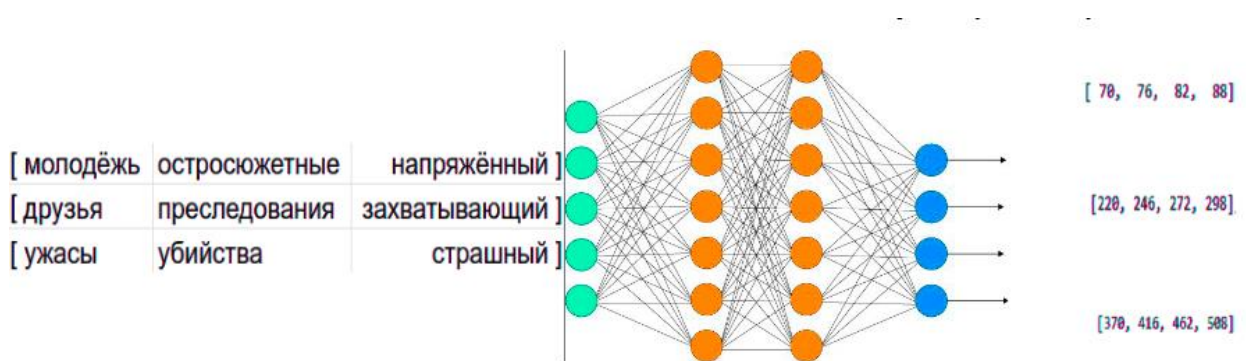


Рисунок 1.10 – Схема «холодного старту»

Холодний старт – багатоскладова проблема. Найкращий її варіант: абсолютно нові користувачі, про які практично немає жодних даних [17].

Базове рішення: показувати найпопулярніший контент. Можна розбити популярність на глобальну популярність та недавні тренди. Можна окремо вважати популярність країнами/майданчиками.

Критично важливо показати користувачу всю різноманітність цікавого контенту. Максимально різноманітно, але при цьому є максимально популярним. Це нетривіальна проблема, навіть якщо її вирішуватимуть у ручному режимі експерти. Якщо формалізувати завдання, то потрібен такий набір контенту (фіксованого розміру), щоб якнайбільше користувачів знайшли собі серед нього хоча б один цікавий елемент.

Ідея рішення досить проста: вектори ембедингів користувачів та об'єктів після матричного розкладання знаходяться в одному просторі. Що дозволяє нам говорити про схожість користувача з контентом. У цьому просторі об'єкти представлені векторами. Якщо ми візьмемо N таких векторів і подивимося на область, яку вони окреслюють, ми хочемо, щоб ця область була максимальна (щоб туди потрапило якнайбільше користувачів). Якщо говорити інтуїтивно і спрощено, то кожне число у векторі означає певну значущу ознаку об'єкта або набір ознак. А величина цієї ознаки – його виразність.

Вимоги можуть бути виконані шляхом пошуку підмножини стовпців (об'єктів), що максимізує обсяг паралелепіпеда, охопленого ними:

$$k \leftarrow \operatorname{argmax} \operatorname{Vol} S = \operatorname{argmax} |\det S|, \quad S = Q(:, k). \quad (1.12)$$

У випадку завдання NP-повна, але можна знайти субоптимальне рішення жадібним алгоритмом, з назвою *maximal-volume algorithm*. Суть у пошуку *dominant submatrix*, квадратної матриці.

Домінуюча властивість S означає, що всі стовпці $q_i \in \mathbb{R}^{L_0}$ поля Q можуть бути представлені через лінійну комбінацію стовпців S з коефіцієнтами, що не перевищують 1 по модулю. Хоча ця властивість не означає, що S має максимальний обсяг, воно гарантує, що S є локально оптимальним, що означає, що заміна будь-якого стовпця S стовпцем Q не збільшує обсяг.

Рекомендаційні системи, зазвичай, у своїй основі, базуються на тому, що вони включають в себе величезну кількість користувачів та товарів. Але користувачі не завжди оцінюють товари, навіть якщо придбали їх. В результаті дані в матриці «користувач-предмет» виходять дуже розріджені. Ця проблема особливо гостро стоїть для нещодавно створених рекомендаційних систем. Також ця проблема (розрідженість даних) загострює проблему холодного старту [18].

Зі збільшенням кількості користувачів в системі, з'являється проблема масштабованості. Наприклад, маючи десять мільйонів покупців $O(M)$ і мільйон предметів $O(N)$, алгоритм колаборативної фільтрації зі складністю рівній $O(N \cdot M)$ вже занадто складний для розрахунків. Також, багато систем повинні моментально реагувати на онлайн запити від всіх користувачів, незалежно від історії їх покупок і оцінок, що вимагає ще більшої масштабованості.

Neural collaborative filtering (NCF) спирається на колаборативні дані неявного фідбека. Але намагається на відміну від нього уникнути скалярного твору як міра схожості, оскільки NCF може вивчити складніші залежності. Основна ідея NCF у тому, щоб зробити цей захід більш складним і навчальним. Структура Neural collaborative filtering зображено на рисунку 1.11.

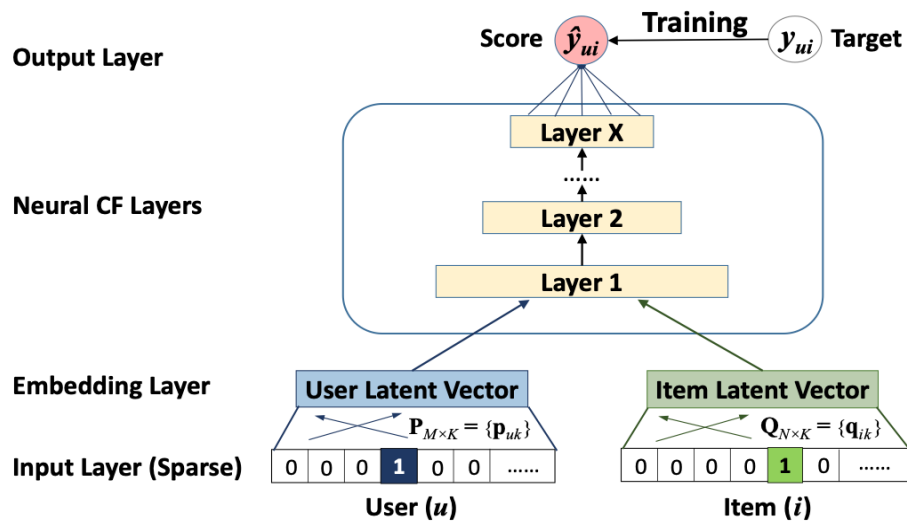


Рисунок 1.11 – Neural collaborative filtering

Вхідні дані можуть бути різними, але спочатку використовуються чисті колаборативні дані, що закодовані one-hot encoding. У ситуації холодного старту вхідними даними можуть бути метадані контенту і користувачів. За вхідним шаром знаходиться повнозв'язковий, який відповідає за перетворення розрідженого one-hot уявлення в щільні ембедінги. Після отримання

ембедінгів конкатинуються, а подальша архітектура носить ім'я *neural collaborative filtering* і відповідає за перетворення ембеддингів пари користувач-об'єкт безпосередньо в передбачення.

У цьому випадку стандартна середньоквадратична помилка може не відповідати виду реальних даних неявного фідбека, які представлені значеннями 0 і 1. Тому використовується імовірнісний підхід до навчання NCF. Якщо передбачене число розглядати як ймовірність значення 1 (тобто релевантності об'єкта користувачеві) і обмежити вихід мережі діапазоном $[0,1]$, можна використовувати \log loss:

$$\begin{aligned} L &= - \sum_{(u,i) \in \mathcal{Y}} \log \hat{y}_{ui} - \sum_{(u,j) \in \mathcal{Y}^-} \log(1 - \hat{y}_{uj}) \\ &= - \sum_{(u,i) \in \mathcal{Y} \cup \mathcal{Y}^-} y_{ui} \log \hat{y}_{ui} + (1 - y_{ui}) \log(1 - \hat{y}_{ui}), \end{aligned} \quad (1.13)$$

де \mathcal{Y} – безліч негативних прикладів, в якості яких береться деяке підмножина взаємодії, що не спостерігаються (пар користувач-об'єкт, за якими немає даних). Негативні приклади відбираються рівномірно, але можна використовувати, наприклад підходи, що враховують популярність, що може покращити продуктивність.

NCF виходить більш загальним варіантом класичного матричного розкладання, яке можна точно відтворити, якщо спростити шари, що йдуть за шаром ембеддингів.

Класична реалізація алгоритму колаборативної фільтрації ґрунтується на принципі k найближчих сусідів. На пальцях – для кожного користувача шукаємо найбільш схожих на нього (у термінах переваг) і доповнюємо інформацію про користувача відомими даними по його сусідах (рисунок 1.12).

На зображенні вище проілюстровано принцип роботи методу. У матриці переваг жовтим кольором виділено користувач, котрій ми хочемо визначити

оцінки по новим товарам (знаки питання). Синім кольором виділено три його найближчі сусіди [19].

	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8
u_1	?	4.0	4.0	2.0	1.0	2.0	?	?
u_2	3.0	?	?	?	5.0	1.0	?	?
u_3	3.0	?	?	3.0	2.0	2.0	?	3.0
u_4	4.0	?	?	2.0	1.0	1.0	2.0	4.0
u_5	1.0	1.0	?	?	?	?	?	1.0
u_6	?	1.0	?	?	1.0	1.0	?	1.0
u_a	?	?	4.0	3.0	?	1.0	?	5.0
r_a	3.5	4.0			1.3		2.0	

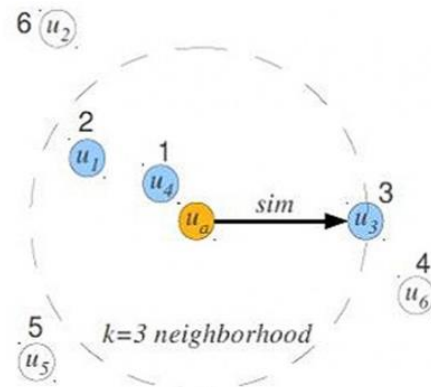


Рисунок 1.12 – Принцип роботи методу колаборативної фільтрації

"Схожість" – в даному випадку синонім "кореляції" інтересів і може вважатися безліччю способів (крім кореляції Пірсона, є ще косинусне відстань, є відстань Жаккара, відстань Хеммінга та ін.).

У класичній реалізації алгоритму є один явний мінус - він погано застосовується на практиці через квадратичну складність. Справді, як і будь-який метод найближчого сусіда, він вимагає розрахунку всіх попарних відстаней між користувачами (а користувачів можуть бути мільйони). Неважко порахувати, що складність розрахунку матриці відстаней буде $O(n^2m)$, де n – число користувачів, а m – число товарів.

Ця проблема частково може бути вирішена купівлею високопродуктивного заліза. Але якщо підходити з розумом, то краще ввести коригування в алгоритм:

- оновлювати відстані не при кожній покупці, а батчами (наприклад, щодня);

- не перераховувати матрицю відстаней повністю, а оновлювати її інкрементально;

– зробити вибір на користь ітеративних та наближених алгоритмів (наприклад ALS).

Околиця користувача у просторі переваг (його сусіди), яку аналізують для генерації нових рекомендацій, можна обирати по-різному. Можемо працювати взагалі з усіма користувачами системи, можемо задати якийсь поріг близькості, можемо вибрати кілька сусідів випадковим чином або брати найбільш схожих сусідів (це найбільш популярний підхід).

Оскільки всі користувачі оцінюють по-різному – хтось усім поспіль п'ятірки ставить, а від когось четвірки рідко дочекаєшся – перед розрахунком дані краще нормалізувати, тобто. привести до єдиної шкали, щоб алгоритм міг коректно порівнювати їх між собою.

Нормалізувати можна кількома способами:

- центруванням (mean-centering): з оцінок користувача просто віднімаємо його середню оцінку;
- стандартизацією (z-score): на додаток до центрування ділимо оцінку її на стандартне відхилення у користувача;
- подвійною стандартизацією: вперше нормуємо оцінками користувача, вдруге – оцінками товару.

"Схожість" або кореляцію переваг двох користувачів можна вважати різними способами. По суті, нам треба просто порівняти два вектори. Перелічимо кілька найпопулярніших.

Кореляція Пірсона - класичний коефіцієнт, який цілком можна застосувати і при порівнянні векторів:

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \quad (1.14)$$

Основний його недолік – коли перетин за оцінками низький, кореляція може бути високою випадково.

Кореляція Спірмана. Основна відмінність – коефіцієнт ранговий, тобто. працює не з абсолютними значеннями рейтингів, а з їхніми порядковими номерами. Загалом дає результат дуже близький до кореляції Пірсона:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}. \quad (1.15)$$

Косинусна відстань. Ще один класичний коефіцієнт. Якщо придивитися, косинус кута між стандартизованими векторами — це і є кореляція Пірсона, та сама формула:

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}. \quad (1.16)$$

Косинусне представлення використовується тому, що якщо два вектори співспрямовані (тобто кут між ними нульовий), то косинус кута між ними дорівнює одиниці. І навпаки, косинус кута між перпендикулярними векторами дорівнює нулю.

1.4 Постановка задачі дослідження

З кожним роком обсяг тої чи іншої інформації збільшується, тому вручну обробляти та освоювати її стає вкрай важко. Звідси впливають певні вимоги щодо створення різноманітних автоматизованих комп'ютерних засобів для зберігання, оброблення, аналізування інформації. Тому виникає потреба у розробленні (проектуванні) інформаційних систем, що дають змогу збирати усю інформацію, а після цього її аналізувати та добавляти до зібраної вище.

Також основною її характеристикою є адаптивність та прийняття рішень до будь-якого із середовищ користувача.

Об'єкт дослідження – процес побудови рекомендацій вибору спеціальності абітурієнтам, котрі здобули повну загальну середню освіту і мають на меті вступити на навчання до закладів освіти.

Предмет дослідження – методи формування рекомендацій вибору спеціальності для абітурієнтів.

Мета роботи – аналіз методів формування рекомендацій вибору спеціальності і проектування системи рекомендацій для абітурієнтів, що вступають до закладів освіти.

Математична постановка задачі дослідження:

Наявні:

- користувачі (users, $u \in U$);
- об'єкти (items, $i \in I$);
- події (events, $(r_{ui}, u, i, \dots) \in D$).

Необхідно:

- передбачити переваги $\hat{r}_{ui} = \text{Predict}(u, i, \dots) \approx r_{ui}$;
- персональні рекомендації $u \rightarrow (i_1, \dots, i_K) = \text{Recommend}_K(u, \dots)$;
- схожі об'єкти $i \rightarrow (i_1, \dots, i_M) = \text{Similar}_M(i)$.

Завдання роботи:

- аналіз предметної області профорієнтаційної роботи;
- аналіз методів рекомендаційних систем з точки зору використання в профорієнтаційній роботі;
- розробка рекомендаційної системи;
- експериментальне дослідження та аналіз отриманих даних.

2 ДОСЛІДЖЕННЯ МЕТОДУ ФОРМУВАННЯ РЕКОМЕНДАЦІЇ ВИБОРУ СПЕЦІАЛЬНОСТІ У АБІТУРІЄНТІВ

2.1 Метод колаборативної фільтрації для формування рекомендації вибору спеціальності у абітурієнтів

Системи рекомендацій на основі колаборативної фільтрації роблять прогнози на основі минулих реакцій користувача на елементи в системі. Це передбачення може бути оцінкою за шкалою або просто інформацією про використання елемента.

Основна ідея алгоритмів колаборативної фільтрації полягає в пропозиції нових елементів для конкретного користувача на основі попередніх переваг користувача або думки інших однодумців користувача. Сьогодні дослідники розробили низку алгоритмів колаборативної фільтрації, які можна розділити на три основні категорії:

- методи, засновані на аналізі наявних оцінок (Memory-based). Цей підхід ще називають методом найближчих сусідів: використання попередніх оцінок, зроблених клієнтом, і аналіз оцінок інших користувачів, які мають подібні переваги. Тоді рекомендації (прогноз) для цільового користувача формуються на підставі обчислення певної міри схожості по всіх накопичених даних;

- методи, засновані на побудові моделі даних (Model-based). У цьому випадку спочатку за сукупністю оцінок формується описова модель переваг користувачів, товарів і взаємозв'язку між ними, а потім формуються рекомендації на підставі отриманої моделі. Процес формування рекомендацій розбитий на два етапи: ресурсомістке навчання моделі в відкладеному режимі і досить просте обчислення рекомендацій на основі існуючої моделі в реальному часі. Ці алгоритми можуть базуватися на імовірнісному підході, кластерному аналізі, аналізі прихованих чинників;

– методи, засновані на об'єднанні попередніх алгоритмів, – гібридні методи.

Колаборативна фільтрація базується на матриці, де на одній осі відкладаються користувачі, а на іншій - елементи. Клітинки матриці містять інформацію про те, як конкретний користувач оцінив той чи інший елемент. Відсутність оцінки означає, що користувач ніяк не взаємодіяв з елементом до цього моменту. Мета системи - заповнити ці порожні клітинки [20].

Існує 2 підходи до побудови таких систем:

- на основі користувачів;
- на основі предметів.

Колаборативна фільтрація на основі користувачів працює наступним чином: серед усіх користувачів необхідно виділити n найбільш схожих на ключового та на основі їх оцінок зробити прогноз. Проілюструємо цей підхід таблицею 2.1.

Таблиця 2.1 – Приклад колаборативної фільтрації на основі користувачів

	Предмет 1	Предмет 2	Предмет 3	Предмет 4
Користувач 1	3	3	5	3
Користувач 2	1	5	2	5
Користувач 3			2	
Користувач 4	3	?	5	2

У таблиці 2.1 наведено матрицю оцінок між чотирма користувачами та чотирма предметами. У цьому випадку нам потрібно спрогнозувати оцінки користувача 1 для предмета 2. Оскільки оцінки вказують на те, що користувач 4 найбільш схожий на користувача 1, ми можемо передбачити, що бажане значення приблизно дорівнює 3.

Тепер постає питання, яким чином розрахувати схожість між користувачами або предметами, а потім зробити передбачення. Для розв'язання цієї задачі існує 2 групи алгоритмів:

- алгоритми засновані на пам'яті;
- алгоритми засновані на моделі.

Алгоритми на основі пам'яті працюють, обчислюючи певну міру на основі даних, наявних у системі (матриця оцінок). Рейтинг кожного користувача (або рейтинг кожного товару) можна представити у вигляді вектора, до якого можна застосувати міру схожості. Після того, як міра буде обчислена, її можна використовувати для обчислення прогнозів. Найчастіше використовуються косинус подібності (2.1), який є кутом між векторами оцінок, критерій кореляції Пірсона (2.2), скоригований косинус подібності (2.3) і коефіцієнт Жаккара (2.4) [21]:

$$\text{cosineSimilarity}(u_i, u_j) = \cos(u_i, u_j) = \frac{\sum_{k=1}^n u_{i,k} u_{j,k}}{\sqrt{\sum_{k=1}^n u_{i,k}^2} \sqrt{\sum_{k=1}^n u_{j,k}^2}}, \quad (2.1)$$

$$\text{PearsonSimilarity}(u_i, u_j) = \frac{\sum_{k=1}^n (u_{i,k} - \bar{u}_i)(u_{j,k} - \bar{u}_j)}{\sqrt{\sum_{k=1}^n (u_{i,k} - \bar{u}_i)^2} \sqrt{\sum_{k=1}^n (u_{j,k} - \bar{u}_j)^2}}, \quad (2.2)$$

$$\text{adjustedCosineSimilarity}(u_i, u_j) = \frac{\sum_{k=1}^n (u_{i,k} - \bar{u}_k)(u_{j,k} - \bar{u}_k)}{\sqrt{\sum_{k=1}^n (u_{i,k} - \bar{u}_k)^2} \sqrt{\sum_{k=1}^n (u_{j,k} - \bar{u}_k)^2}}, \quad (2.3)$$

$$\text{JaccardSimilarity}(u_i, u_j) = \frac{|u_i \cap u_j|}{|u_i \cup u_j|}, \quad (2.4)$$

де u_i, u_j – вектори оцінок користувачів i та j ;

n – кількість предметів;

\bar{u}_i, \bar{u}_j – середні оцінки користувачів i та j ;

\bar{u}_k – середня оцінка предмета k .

Далі необхідно розрахувати прогнозоване значення оцінки на основі отриманої міри схожості. Для цього спочатку необхідно вибрати найбільш схожих користувачів з усієї множини користувачів. Це можна зробити, наприклад, вибравши перших n користувачів з точки зору схожості або відфільтрувавши користувачів за порогом схожості. Для прогнозування оцінок використовуються агреговані функції, серед яких найбільш поширеними є

середнє значення (2.5) та середньозважене значення (2.6), відносно зважене середнє (2.7), серед інших [22]:

$$u_{i,k} = \frac{\sum_{u_j \in U'} u_{j,k}}{|U'|}, \quad (2.5)$$

$$u_{i,k} = \frac{\sum_{u_j \in U'} \text{similarity}(u_i, u_j) u_{j,k}}{\sum_{u_j \in U'} |\text{similarity}(u_i, u_j)|}, \quad (2.6)$$

$$u_{i,k} = \bar{u}_i + \frac{\sum_{u_j \in U'} \text{similarity}(u_i, u_j) (u_{j,k} - \bar{u}_{j,k})}{\sum_{u_j \in U'} |\text{similarity}(u_i, u_j)|}, \quad (2.7)$$

де $u_{i,k}$ - оцінка користувача і предмету k ;

U' – множина користувачів, які найбільше схожі на даного;

u_j – вектор оцінок користувача з U ;

\bar{u}_i – середня оцінка користувача і серед оцінених ним предметів.

Отримане у результаті значення $u_{i,k}$ і є шуканим прогнозом.

Наведена методологія колаборативної фільтрації може включати обмежену здатність рекомендацій для нових чи нетипових користувачів, відсутність врахування унікальних властивостей кожного користувача та неможливість зробити прогноз оцінки в разі відсутності історії оцінок об'єкта в кластері.

2.2 Удосконалення методу для формування рекомендації вибору спеціальності у абітурієнтів

У основі методу лежить предиктор, який являє собою середню різницю оцінок двох предметів (2.8). Після цього, предиктор застосовується для прогнозування оцінки. Очікуване значення оцінки розраховується за

формулою (2.9). Для більш точного прогнозування раціонально буде застосувати середньо зважене значення (2.10) [22]:

$$dev_{k,m} = \frac{\sum_{u_i \in U_{k,m}} (u_{j,k} - u_{j,m})}{|U_{k,m}|}, \quad (2.8)$$

$$u_{i,k} = \frac{\sum_{m \in S_i} (dev_{k,m} + u_{i,m})}{|S_i|}, \quad (2.9)$$

$$u_{i,k} = \frac{\sum_{m \in u_i} (dev_{k,m} + u_{i,m}) |U_{k,m}|}{\sum_{m \in u_i} |U_{k,m}|}, \quad (2.10)$$

де $d_{evk,m}$ – предиктор для предмета k на основі предмета m ;

$U_{k,m}$ – множина користувачів, які оцінили як предмет k , так і предмет m ;

u_i – оцінки користувача i ;

S_i – множина предметів, які були оцінені користувачем i та були оцінені будь-яким іншим користувачем з множини спільно з предметом k .

Спільна фільтрація на основі моделей використовує різні алгоритми інтелектуального аналізу даних і машинного навчання для прогнозування оцінок користувачів. Їх можна розділити на три групи. (наприклад, k -найближчих сусідів), алгоритми декомпозиції матриць (наприклад, SVD) та методи глибокого навчання (наприклад, багатошарові нейронні мережі) [23].

Матрична факторизація є одним з найпоширеніших підходів до побудови рекомендаційних систем на основі моделей. Факторизація - це операція, яка розкладає ціль на прості складові [24]. Спочатку необхідно визначити набір прихованих факторів предметної області. Вважається, що існує r факторів. Вони представляють характеристики об'єкта. Потім потрібен алгоритм для декомпозиції оціночної матриці X на дві матриці (матриця представлення користувача - U), та матрицю зв'язків об'єкта з прихованими факторами (матриця представлення об'єкта - V).

Найпоширенішими алгоритмами матричної факторизації є SVD, TruncatedSVD, PCA та NMF. Результатом є $X \approx U * V$ - рівняння 2.11:

$$\begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{pmatrix} \approx \begin{pmatrix} u_{11} & \cdots & u_{1r} \\ \vdots & \ddots & \vdots \\ u_{m1} & \cdots & u_{mr} \end{pmatrix} \begin{pmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & \ddots & \vdots \\ v_{r1} & \cdots & v_{rn} \end{pmatrix}. \quad (2.11)$$

Після цього, щоб отримати рекомендації – тобто заповнити пропуски у початковій матриці X – необхідно виконати множення U матриці на матрицю V , у результаті отримаємо матрицю $X' \approx X$, яка міститиме числа замість пропусків. Проте ці числа не будуть являти собою реальний прогноз оцінки – це лише певні значення, які тепер можна використовувати для рекомендацій. Наприклад, можна інтерполювати оцінки предметів у певний діапазон, а потім рекомендувати користувачам ті предмети, де отримане значення вище заданого порогу [24].

3 ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ДЛЯ ФОРМУВАННЯ РЕКОМЕНДАЦІЙ ВИБОРУ СПЕЦІАЛЬНОСТІ У АБІТУРІЄНТІВ

3.1 Розробка інформаційної технології

Проведений аналіз існуючих методик тестування показав, що кожен із тестів орієнтований на дослідження або особистісних якостей, або схильності до будь-якої професії. Для здійснення обґрунтованої процедури профорієнтації доцільно застосовувати комплекс методик [25].

Процес здійснення профорієнтації подано на наступній діаграмі (рисунок 3.1).

У мережі Інтернет представлено низку сайтів, що пропонують свої послуги у сфері професійної орієнтації:

– профорієнтація: тестування профорієнтації на цьому сайті відбувається в режимі он-лайн. Цей сайт орієнтований насамперед на школярів, які хочуть оцінити свої професійні уподобання. Ресурс пропонує абітурієнту тест «Опитувальник професійної готовності»;

– профГід: також є веб-сайт з розташованими на ньому наступними методиками тестування: карта інтересів, методика професійного самовизначення Дж. Голланда, тест Клімова;

– освіта.юа: на основі відповідей на питання ресурс визначає сфери інтересів, особисті та професійні особливості та пропонує список найбільш підходящих професій. Ресурс пропонує абітурієнту такі тести як карта інтересів, мотивація.

Кожен з представлених ресурсів має свої переваги та недоліки, одним з яких є те, що представлені на них тести не дозволяють вибрати серед напрямів підготовки, які є профілями одного напряму підготовки або знаходяться в одній укрупненій групі.

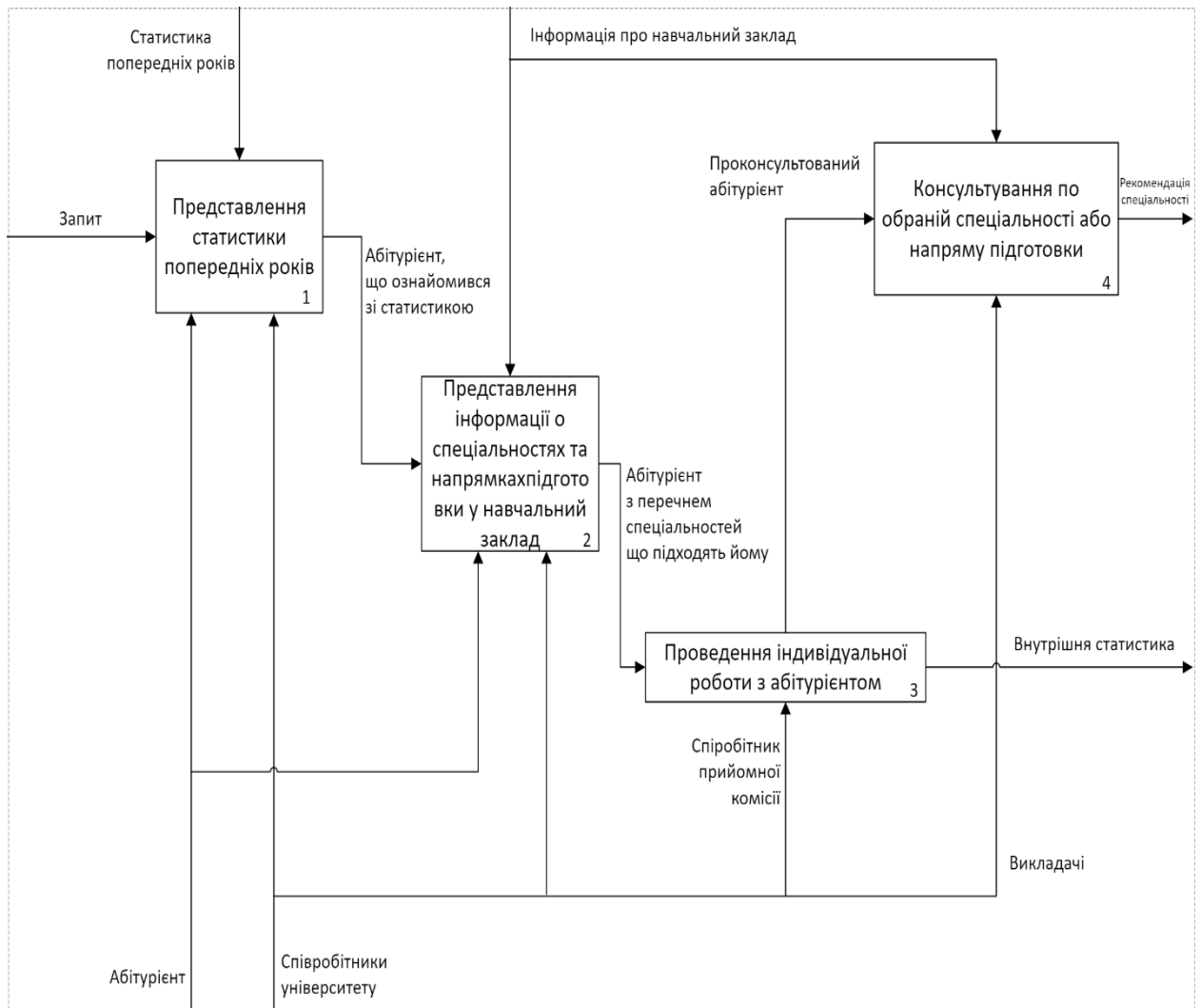


Рисунок 3.1 – Діаграма декомпозиції «Профорієнтація абітурієнтів»

Автоматизація процесу профорієнтації дає змогу абітурієнтам досліджувати свої схильності, особисті якості та характер та отримати сторонню оцінку та рекомендації щодо вибору напрямку підготовки чи спеціальності. У зв'язку з вищесказаним розробка інформаційної системи є актуальним завданням. Розроблена система була призначена для профорієнтації абітурієнтів, які планують вступати до навчального закладу.

Для реалізації поставленої мети інформаційна система відповідає наступним функціональним вимогам:

- централізація даних у БД;
- автоматизація процесу тестування та видачі результатів;

- зручність побудови графіків з результатами тестування та статистикою;
- автоматизація збирання внутрішньої статистики;
- розмежування ролей та доступу для користувачів системи;
- взаємодія інтерфейсів з базою даних системи.

З поширенням Internet з'явилася можливість централізованого зберігання та супроводу цих знань (експертних систем) та надання доступу до них через канали зв'язку. Більше того, у зв'язку з бурхливим розвитком мобільного зв'язку з'являється можливість доступу до таких експертних систем через дешеві та малогабаритні засоби, як мобільні телефони.

Популярним для визначення або підбору спеціальності, або напряду підготовки, що підходить абітурієнту, останнім часом стало використання експертних систем (ЕС).

ЕС видають поради, проводять аналіз, виконують класифікацію, дають консультації та ставлять діагноз. Вони орієнтовані рішення завдань, зазвичай потребують проведення експертизи людиною-фахівцем. На відміну від машинних програм, що використовує процедурний аналіз, ЕС вирішують завдання у вузькій предметній галузі (конкретній галузі експертизи) на основі дедуктивних міркувань. Такі системи часто виявляються здатними знайти розв'язання задач, які неструктуровані та погано визначені. Вони справляються з відсутністю структурованості шляхом залучення евристик, тобто правил, взятих "зі стелі", що може бути корисним у тих системах, коли брак необхідних знань або часу виключає можливість проведення повного аналізу.

У процесі дослідження було вирішено такі завдання: неформальний опис предметної області (ПрО), виділення когнітивних елементів у ПрО, перетворення ключових понять та відносин формальною мовою, програмна реалізація та перевірка працездатності розробленої автоматизованої інформаційної системи. В результаті цього дослідження було створено програму, що дозволяє користувачеві на основі обраних критеріїв (основних

понять) здійснити вибір відповідного напрямку підготовки або визначити схильність до конкретних видів діяльності.

Насамперед необхідно зрозуміти, як здійснюється вибір сфери професійної діяльності. Здобувач (абітурієнт), перш за все, має визначитися з цілями вибору і саме на їх основі мають бути обрані найважливіші критерії для оптимального підбору майбутнього напрямку діяльності [26].

Важливим кроком визначення критеріїв є виділення основних понять і зв'язків з-поміж них. Тобто, утворюється повний систематичний набір термінів з галузей професійних знань, пов'язаних із профорієнтацією та віковою психологією, а також будується мережа асоціацій, де зв'язки намічені та спрямовані, але не поіменовані.

Зв'язки, отримані першому кроці, дозволяють структурувати поняття і виявляти поняття вищого рівня узагальнення (метапоняття), і навіть деталізувати їх у нижчому рівні.

Оскільки було виявлено об'єкти, зв'язок між ними, а як і структуровані поняття вищому рівні, можна дійти невтішного висновку у тому, що за виборі напрямки підготовки (майбутньої сфери професійної діяльності) слід враховувати такі критерії вибору: сфера роботи, тип особистості, тип професії, тип інтелекту, мислення, здібності (а абітурієнтів - форма підготовки).

Концептуальна схема роботи системи представлена рисунком 3.2. На схемі представлені критерії вибору та його значення.

Необхідні для роботи з інформаційною системою значення критеріїв користувач отримує після проходження тестів профорієнтацій, представлених в системі.

Експертна система здійснює вибір відповідних користувачеві спеціальностей та напрямків підготовки в рамках навчального закладу на підставі особистих якостей, схильностей, які виявляються на етапі тестування.

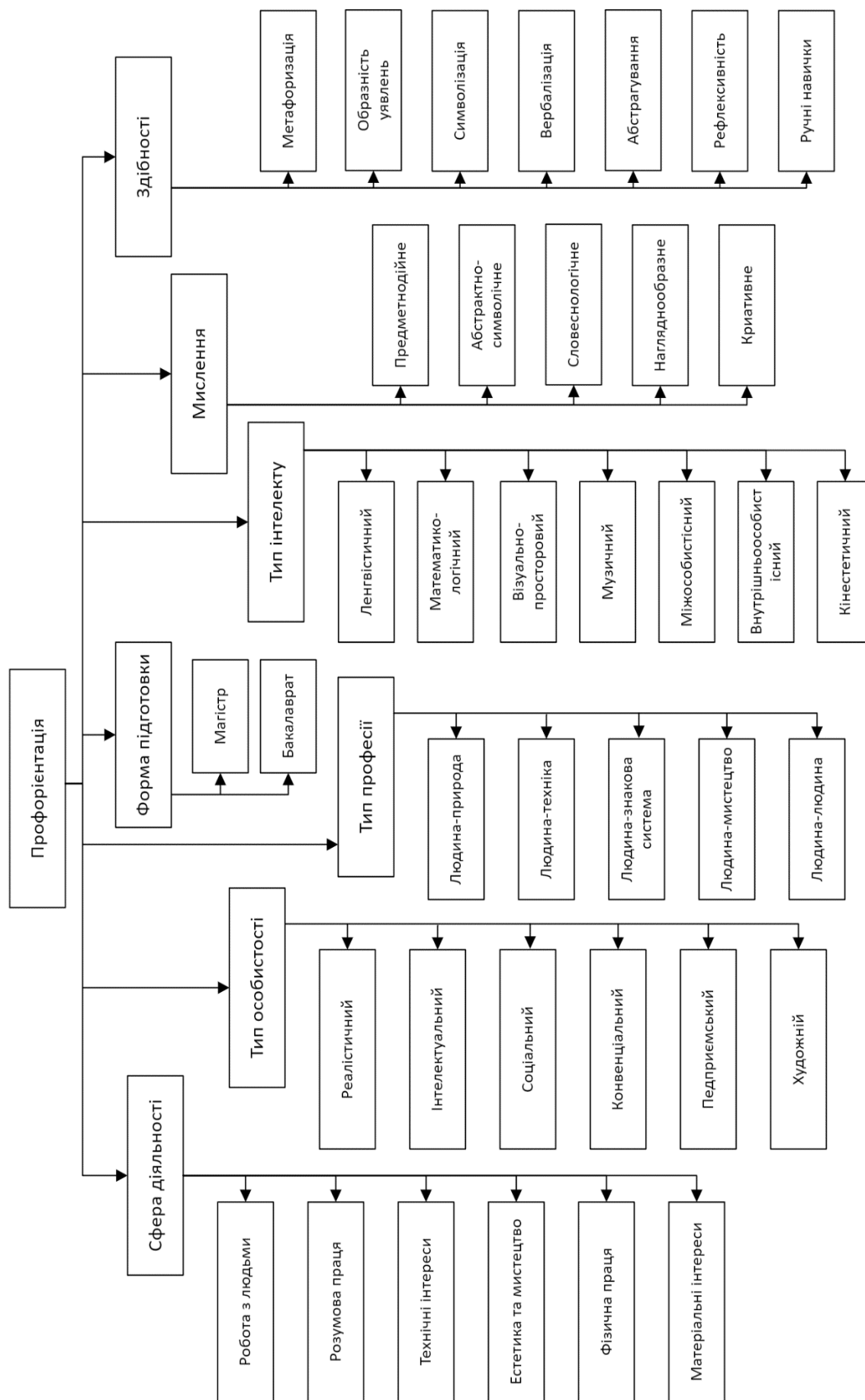


Рисунок 3.2 – Концептуальна схема експертної системи

Інформаційна система «Вибір абітурієнта» призначена для виконання кількох функцій, призначених для різних груп користувачів: для звичайних користувачів та адміністраторів системи, для розмежування їх функцій інтерфейс системи розділений, як показано на рисунку 3.3 [27].



Рисунок 3.3 – Розмежування інтерфейсу

Переваги системи полягають у такому:

- простота у користуванні;

- зрозумілий інтерфейс;
- висока швидкодія оброблення даних;
- система не містить реклами;
- безкоштовний сервіс;
- не потрібна реєстрація на сайті.

Недоліки системи такі, як:

- система працює тільки на сайті навчального закладу;
- потреба в адміністраторі, котрий час від часу вноситиме зміни, щодо самих рекомендацій.

Основний успішний сценарій:

- система починає обробляти введені рекомендаційні дані користувача при натисненні кнопки, цим самим формування запиту на оброблення;
- система формує саму рекомендацію, згідно з пройденим тестом і рекомендує бажаний результат користувачу.

Доповнення ключового сценарію (неправильний ідентифікатор):

- програмована система повідомляє користувача про помилку запиту і відміння оброблення даних щодо формування тестів;
- програмована система обробляє тести і надає кінцеву рекомендацію користувачу.

Діаграма варіантів використання, зображена на рисунку 3.4.

На діаграмі варіантів використання було використано два типи сутностей, такі, як: варіанти використання та дійові особи, між якими було встановлено такі типи відношень: асоціація, узагальнення та включення між варіантами використання.

Для включення в діаграму прецеденти повинні задовольняти такі критерії:

- прецедент повинен описувати, ЩО треба робити, а не ЯК;
- прецедент повинен описувати дії з погляду ВИКОНАВЦЯ;
- прецедент повинен повертати виконавцю деяке ПОВІДОМЛЕННЯ;

– послідовність дій всередині прецеденту повинна являти собою один НЕРОЗДІЛЬНИЙ ланцюжок.

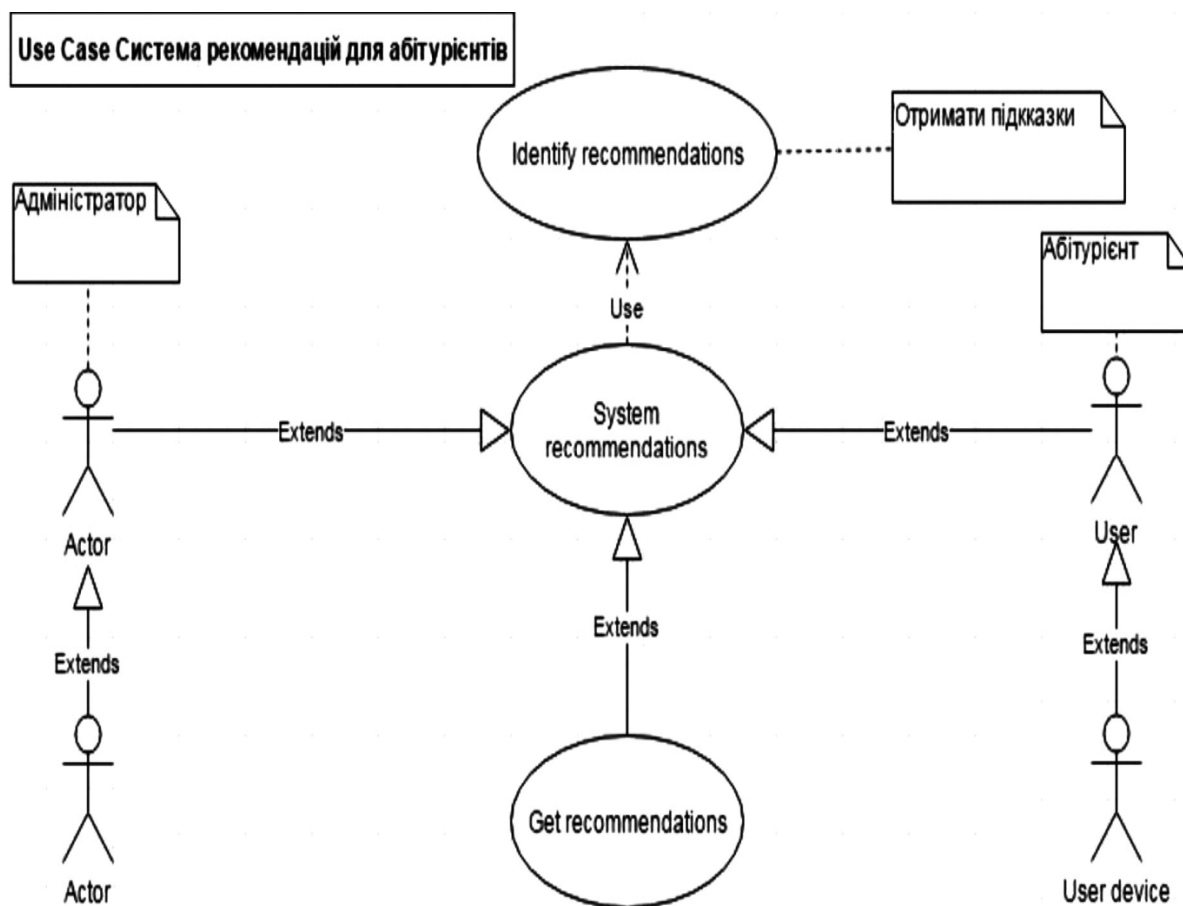


Рисунок 3.4 – Діаграма варіантів використання

Діаграми діяльності, що зображена на рисунку 3.5, відображає опис сценарію системи.

Застосування діаграм діяльності для реалізації варіантів використання приводить до глибшого розуміння суті завдання, а відтак надає засоби із покращення додатку, які важко було зауважити у початковій постановці завдання.

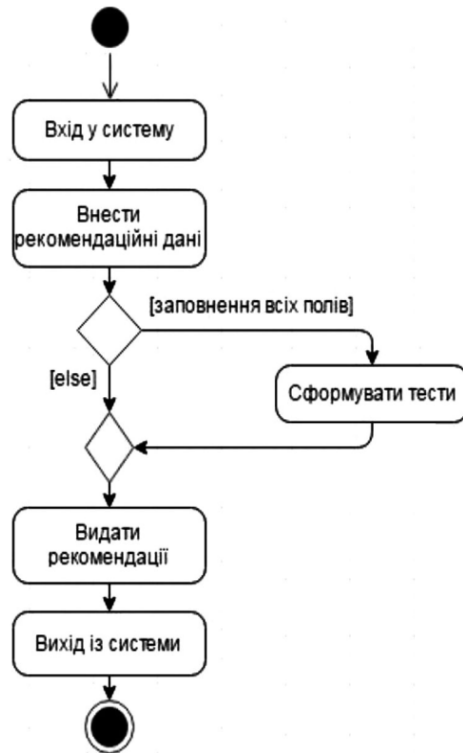


Рисунок 3.5 – Діаграма діяльності

Загальна структурна схема інформаційної системи представлена на рисунку 3.6 [28].

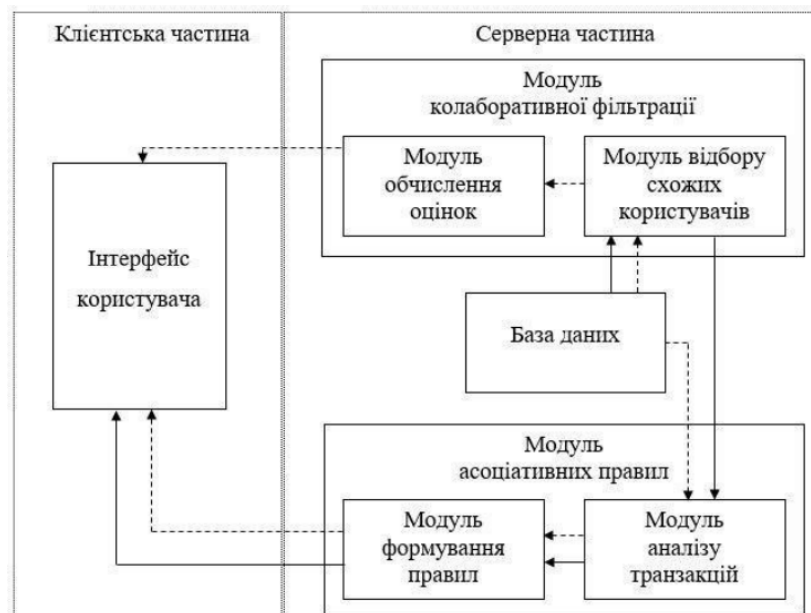


Рисунок 3.6 – Загальна структурна схема інформаційної системи

3.2 Перспективи подальших досліджень

Запропоновані підходи є перспективними кандидатами на подальші вдосконалення, а саме їх дослідження в інших сферах машинного навчання та великих даних, оскільки можуть принести вагомий внесок та прискорити різного роду системи та алгоритми, які ними використовуються.

Отримані багатообіцяючі результати є лише відправною точкою. Можна розглянути можливість розгортання цієї реалізації на багатовузловому кластері, щоб оцінити його масштабованість, продуктивність (зокрема час обчислень) і точність у розподіленому режимі.

В подальших дослідженнях пропонується зосередити увагу на розширенні інформаційної бази функціонування системи та збільшенні кількості функцій. Також доцільним є впровадження метрик оцінки якості рекомендацій, що можуть бути використані в алгоритмі самостійного додаткового навчання алгоритмів системи та при тонкому налаштуванні параметрів цих алгоритмів.

4 РЕАЛІЗАЦІЯ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ДЛЯ ФОРМУВАННЯ РЕКОМЕНДАЦІЇ ВИБОРУ СПЕЦІАЛЬНОСТІ У АБІТУРІЄНТІВ

4.1 Обґрунтування вибору програмних засобів розробки

Одним з важливих завдань при розробці програмного забезпечення є вдалий вибір мови програмування. Розглянемо переваги і недоліки популярних мов програмування з метою вибору оптимального варіанту для реалізації програми. В якості кандидатів для виконання даного завдання були вибрані наступні перелічені нижче мови програмування [29].

Microsoft Visual C ++. Мова програмування Visual C ++ з середовища розробки Visual Studio компанії Microsoft найбільш потужний засіб розробки системних програм. Однак, розробка програм на цій мові – трудомісткий і складний процес.

Microsoft Visual Basic. Об'єктно-орієнтована мова програмування, як правило, вбудована в додатки Microsoft Office. До переваг можна віднести простоту створення нескладних додатків, а також можливість редагування і створення компонент Microsoft Office.

Delphi. Заснована на Паскалі, об'єктно-орієнтоване середовище розробки компанії Borland. Має зручні засоби для розробки віконних додатків. Крім цього є стандартні компоненти для роботи і з базами даних і з веб-сервісами.

C # .NET. C # об'єктно-орієнтована мова, як і вся платформа .NET, яка орієнтована на написання компонент. перейнявши багато від своїх попередників – мов C ++, Java, Delphi, Модула і Smalltalk – C #, спираючись на практику їх використання, виключає деякі моделі, що зарекомендували себе як проблематичні при розробці програмних систем, наприклад, C # не підтримує множинне успадкування класів. В основі мови – легкість використання, переважна надпотужністю і швидкістю виконання. Збирач сміття, керуючи об'єктними посиланнями, автоматично звільняє пам'ять. Дає безпеку роботи з типами – один з найважливіших факторів уникнення помилок. C #

дає можливість реалізації графічних додатків. Windows. Forms – це набір різних керованих бібліотек, за допомогою яких можна виконати всі необхідні для віконної програми дії, починаючи від обміну повідомленнями з операційною системою для відстеження будь-яких подій клієнтського вікна, закінчуючи діалоговими системами, зв'язком з іншими комп'ютерами по мережі і багатьма іншими можливостями. Windows Forms є ще однією з двох технологій, яка використовується в Visual C # для створення інтелектуальних клієнтських програми на основі Windows, які виконуються в середовищі .NET Framework.

Java – об'єктно-орієнтована і крос-платформна мова програмування, яка дозволяє скоротити загальний час розробки і писати повторно використовуваний код. Платформонезалежність байт-коду забезпечується наявністю віртуальних java-машин для всіх основних платформ. У комплект поставки Java входять стандартні класи, які мають достатньою функціональністю для швидкої розробки додатків. Розвинені засоби безпеки дозволяють використовувати Java для розробки додатків, що працюють в Інтернеті. Недоліком Java є повільна швидкість роботи, обумовлена використанням JIT- компіляторів.

Після проведеного аналізу, було вирішено для реалізації нашого програмного продукту використати мову програмування C #. Це обумовлено тим, що C # має широкий інструментарій, безліч спеціальних бібліотек. Крім цього, мова має багатий інструментарій для побудови екранних форм [30].

В якості середовища розробки було обране середовище Microsoft Visual Studio. Середовище розробки Visual Studio продовжує традиції корпорації Microsoft в області надання ефективних інструментальних засобів для розробників складного ПЗ. Забезпечуючи середу розробки для всіх мов програмування, доповнену набором вікон з інтуїтивно зрозумілими інструментальними засобами, контекстної довідки, автоматизованими механізмами виконання різноманітних завдань розробки.

Visual Studio дозволяє в стислі терміни проводити професійну розробку програм різного призначення, в тому числі з підтримкою технології Windows Forms, а також веб-сайти, веб-додатки, веб-служби для всіх платформ, що підтримуються Windows, Windows Mobile, Windows CE, .NET Framework, Xbox, Windows Phone .NET Compact Framework і Silverlight.

Visual Studio включає в себе редактор вихідного коду з підтримкою технології IntelliSense і можливістю найпростішого рефакторінга коду. Вбудований відладчик може працювати як відладчик рівня вихідного коду, так і як відладчик машинного рівня. Visual Studio дозволяє створювати і підключати сторонні додатки (плагіни) для розширення функціональності практично на кожному рівні, включаючи додавання підтримки систем контролю версій вихідного коду (як, наприклад, Subversion і Visual SourceSafe).

4.2 Обґрунтування вибору платформи СУБД

Процес генерації фізичної структури бази даних з логічної моделі даних називається прямим проектуванням (Forward Engineering).

Процес генерації логічної моделі з фізичної бази даних називається оберненим проектуванням (Reverse Engineering). Програма IVExpert дозволяє швидко створити модель даних шляхом оберненого проектування наявної бази даних [31].

IVExpert може зробити обернене проектування існуючої бази даних, автоматично створюючи відповідну діаграму моделі даних. Після того як діаграма буде згенеровано, ви можете, використовуючи інструменти IVExpert, додавати нові об'єкти і перепроєктувати структуру бази даних.

Пряме проектування виконують для того, щоб з логічної моделі даних перейти до інструкцій мови SQL, за допомогою яких можна внести зміни в БД.

Обернене проектування виконують для:

- кращого розуміння структури БД;
- створення документації;
- подальшого внесення змін в структуру БД шляхом прямого проектування.

У складі програми ІВЕхpert міститься CASE-засіб Database Designer. Він дозволяє розробити модель даних – структурну схему БД. Також він може виконувати пряме і обернене проектування БД.

Для запуску Database Designer треба виконати команду Інструменти | Database Designer з меню програми ІВЕхpert.

Для виконання оберненого проектування БД потрібно виконати команду Designer | Reverse Engineer. На екрані з'явиться список зареєстрованих в програмі ІВЕхpert баз даних. Виберіть одну з них. У наступному вікні виберіть режим оновлення діаграми – Create new diagram (створити нову діаграму) чи Update current diagram (оновити поточну діаграму), після чого натисніть кнопку Start. На діаграмі з'явиться схема БД, представлену на рисунку 4.1.

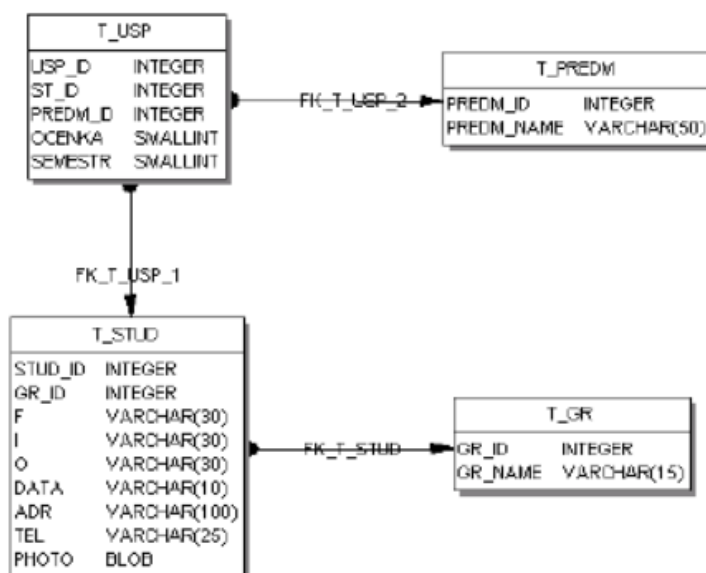


Рисунок 4.1 – Схема бази даних створена під час оберненого проектування

Для автоматизації проектування БД та розробки інформаційних систем використовуються програмно-технологічні засоби які мають назву CASE(Computer Aided Software Engineering). Програма IBConsole, яка входить до складу Interbase, має досить обмежені можливості. На думку багатьох адміністраторів БД, IBConsole є не дуже зручним інструментом адміністрування. Особливо чітко її недоліки проявляються при створенні та адмініструванні складних баз даних. Тому найчастіше для роботи з базами Interbase використовують продукти сторонніх розроблювачів, що забезпечують зручний графічний інтерфейс і додаткові засоби проектування та адміністрування БД. Одним з таких програмних продуктів є програма-оболонка IVExpert.

Програма – оболонка IVExpert призначена для створення, проектування й адміністрування БД, а також для редагування даних, що містяться в ній. До складу оболонки входять також SQL-редактор, візуальні засоби побудови SQL-запитів і засоби адміністрування. Нові версії програми IVExpert мають також CASE-засоби проектування БД [32].

Проаналізувавши всі переваги і недоліки можна зробити висновок, що IVExpert дає можливість виконувати складні операції проектування БД і при цьому не потребує знання мови SQL.

Фізична реалізація бази знань в середовищі IVExpert представлена рисунком 4.2.

#	CODE	NAME	INSTIYUTE	FACULTY
1	121	Інженерія програмного забезпечення	ХНУРЕ	Комп'ютерні науки
2	122	Комп'ютерні науки	ХНУРЕ	Комп'ютерні науки
3	123	Комп'ютерна інженерія	ХНУРЕ	Комп'ютерна інженерія та управління
4	124	Системний аналіз	ХНУРЕ	Інформаційно-аналітичних технологій та менеджменту
5	125	Кібербезпека	ХНУРЕ	Комп'ютерна інженерія та управління
6	126	Інформаційні системи та технології	ХНУРЕ	Інформаційні радіотехнології та технічний захист інфо

Рисунок 4.2 – Таблиця бази знань

База знань містить інформацію про спеціальності та напрями підготовки, за якими здійснюється набір, а також інформація про те, до якого типу професій, згідно з тестами, вони належать.

4.3 Реалізація інтерфейсу програми

Розглянемо приклад роботи користувача (абітурієнта) із системою. На головній формі користувач вирішує, з чого розпочати ознайомлення із системою. При виборі пункту «Інформація про університет» відкривається вікно, представлене на рисунку 4.3.

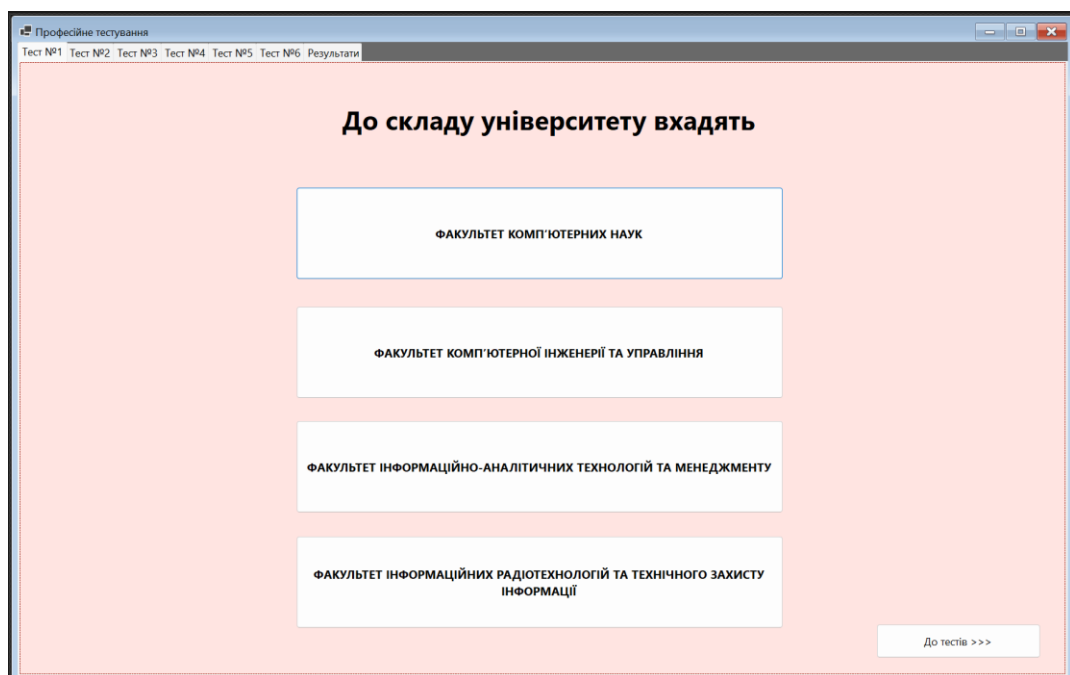


Рисунок 4.3 – Форма «Інформація про університет»

На формі представлено перелік інститутів та факультетів Білгородського державного національного дослідницького університету. При натисканні на кнопку з назвою інституту або факультету здійснюється перехід

на відповідну вкладку, де детально розписані всі спеціальності, напрями підготовки та магістерські програми, а також коротка інформація про терміни навчання та історію інституту/факультету.

Головний розділ системи – пункт «Тестування». Перейшовши цю форму, користувач починає тестування, як зображено на рисунку 4.4.

Професійне тестування
Тест №1 Тест №2 Тест №3 Тест №4 Тест №5 Тест №6 Результати

**Припустимо що після відповідного навчання ви можете виконувати будь-яку роботу.
Але якби вам довелося обирати тільки з двох можливостей,
щоб ви обрали?**

Вибір 1 <input checked="" type="radio"/> Доглядати за тваринами <input type="radio"/> Обслуговувати машини, прилади (стежити, регулювати)	Вибір 7 <input checked="" type="radio"/> Копіювати малюнки (або налаштувати музичні інструменти) <input type="radio"/> Керувати підйомним краном, трактором, тепловозом, тощо	Вибір 13 <input type="radio"/> Вивчати роботу гуртків художньої самодіяльності <input checked="" type="radio"/> Спостерігати, вивчати життя мікробів
Вибір 2 <input type="radio"/> Допомогати хворим <input checked="" type="radio"/> Складати таблиці, схеми, програми для обчислювальних машин	Вибір 8 <input checked="" type="radio"/> Пояснювати людям інформацію (біуро, екскурсії, тощо) <input type="radio"/> Оформлювати виставки, вітрини (або підготовка п'єс, концерт)	Вибір 14 <input checked="" type="radio"/> Обслуговувати, налагоджувати медичні прилади, апарати <input type="radio"/> Надавати людям мед. допомогу при пораненнях, опіках, тощо
Вибір 3 <input checked="" type="radio"/> Стежити за якістю книжкових ілюстрацій, плакатів, тощо <input type="radio"/> Стежити за станом, розвитком рослин	Вибір 9 <input type="radio"/> Ремонтувати речі, одяг, техніку, житло, тощо <input checked="" type="radio"/> Знаходити й виправляти помилки в текстах, таблицях, малюнках	Вибір 15 <input type="radio"/> Складати звіти про явища, події, об'єкти, які спостерігаються <input checked="" type="radio"/> Художньо описувати події, які спостерігають і які уявляють
Вибір 4 <input checked="" type="radio"/> Обробляти матеріали (дерево, тканини, метал, пластмасу тощо) <input type="radio"/> Доводити інформацію про товари до споживача, продавача	Вибір 10 <input checked="" type="radio"/> Лікувати тварин <input type="radio"/> Виконувати обчислення, розрахунки	Вибір 16 <input type="radio"/> Робити лабораторні аналізи в лікарні <input checked="" type="radio"/> Приймати хворих, розмовляти з ними, призначати лікування
Вибір 5 <input checked="" type="radio"/> Обговорювати науково-популярні книжки, статті <input type="radio"/> Обговорювати художні твори (п'єси, концерти)	Вибір 11 <input type="radio"/> Виводити нові види рослин <input checked="" type="radio"/> Конструювати, проектувати нові види промислових виробів	Вибір 17 <input type="radio"/> Фарбувати або розписувати стіни приміщень або вироби <input checked="" type="radio"/> Здійснювати монтаж або збирання машин, приладів
Вибір 6 <input type="radio"/> Вирощувати молодняк (тварини певної породи) <input checked="" type="radio"/> Тренувати товаришів (або дітей) виконувати певні дії	Вибір 12 <input type="radio"/> Аналізувати суперечки: пояснювати, карати, заохочувати <input checked="" type="radio"/> Аналізувати креслення, схеми, таблиці (перевіряти, уточнювати)	Вибір 18 <input checked="" type="radio"/> Організувати походи до театрів, музеїв, екскурсій, турпоходи <input type="radio"/> Грати на сцені, брати участь у концертах
Вибір 19 <input type="radio"/> Виготовлювати за кресленням вироби (машини, одяг будинки) <input checked="" type="radio"/> Займатися кресленням, копіювати креслення, карти	Вибір 20 <input type="radio"/> Вести боротьбу із хворобами рослин, зі шкідниками лісу, саду <input checked="" type="radio"/> Працювати на клавіш машинах (друкарській, складальній, тощо)	Скинути Результат

Рисунок 4.4 – Проходження тесту №3

Як можна побачити з малюнка, окрім тесту на формі розташовані дві кнопки: «Результат» та «Скидання». Натисканням на кнопку «Скинути» користувач може прибрати вибрані відповіді, використання кнопки «Результат» необхідне для переходу на вкладку «Результати», на якій централізовано відображаються результати пройденого тестування. Так, наприклад, тест №1 заснований на методиці Йовайші, тест №2 – це використання методики Голланду. Після проходження всіх тестів можна перейти на закладку Результати (рисунок 4.5).

Професійне тестування

Тест №1 Тест №2 Тест №3 Тест №4 Тест №5 Тест №6 Результати

Результати тесту №1

Робота з людьми	35%
Розумова праця	40%
Технічні інтереси	40%
Естетика та мистецтво	80%
Фізична праця	60%
Матеріальні інтереси	50%

Повернутись до десту

Результат

Результати тесту №2

Реалістичний	18%
Інтелектуальний	56%
Соціальний	36%
Конвенціональний	77%
Педприємський	35%
Художній	60%

Повернутись до десту

Результат

Результати тесту №3

Людина-природа	33%
Людина-техніка	50%
Людина-знакова система	100%
Людина-мистецтво	38%
Людина-людина	50%

Повернутись до десту

Результат

Результати тесту №4

Метафоризація	42%
Образність	28%
Символізація	84%
Вербалізація	42%
Абстрагування	14%
Рефлексивність	70%
Ручні навички	70%

Повернутись до десту

Результат

Результати тесту №5

Предметнодійне	72%
Абстрактно-символічне	24%
Словеснологічне	60%
Нагляднообразне	60%
Крилативне	48%

Повернутись до десту

Результат

Результати тесту №6

Ленгвістичний	75%
Математико-логічний	25%
Візуально-просторовий	50%
Музичний	25%
Міжособистісний	25%
Внутрішньоособистісний	75%
Кінестетичний	100%

Повернутись до десту

Результат

Підібрати спеціальність

Очистити результати

Рисунок 4.5 – Результат тестувань

Професійне тестування

Тест №1 Тест №2 Тест №3 Тест №4 Тест №5 Тест №6 Результати

На підставі пройдених тестів оберіть найбільш підходячі варіанти

1

- Робота з людьми
- Розумова праця
- Технічні інтереси
- Естетика та мистецтво
- Фізична праця
- Матеріальні інтереси

2

- Реалістичний
- Інтелектуальний
- Соціальний
- Конвенціональний
- Конвенціональний
- Художній

3

- Людина-природа
- Людина-техніка
- Людина-знакова система
- Людина-мистецтво
- Людина-людина

4

- Мегафоризація
- Образність
- Символізація
- Вербалізація
- Абстрагування
- Рефлексивність
- Ручні навички

Форма підготовки

- Магістр
- Бакалавр

126 - Інформаційні системи та технології

Скинути

Результат

Головна

Тести

Рисунок 4.6 – Рекомендація спеціальності

Для кожного тесту відводиться своя область з кнопками: «Повернутися до тесту» та «Результат», це зроблено для того, щоб уникнути помилок. Так як в системі представлені психологічні тести, то одним з важливих моментів є те, що на кожне питання тесту має бути відповідь, тому при натискання на кнопку програма перевіряє: чи всі запитання було дано відповіді, у разі негативного відповіді система видає повідомлення про помилку.

При правильному заповненні форм користувач може сміливо натискати кнопку «Підібрати спеціальність» та переходити на форму вибору спеціальності чи напрямки підготовки (рисунок 4.6).

Розглянемо можливості адміністрування системи відповідальною особою. Адміністратор після проходження процедури авторизації отримує доступ до операції адміністрування. Якщо дані не відповідають тим, що зберігаються в базі, буде видано повідомлення про помилку. В іншому випадку користувач перейде на форму редагування даних, яка представлена на рисунку 4.7.

ID	CODE	NAME	INSTIYUTE	FACULTY
1	121	Інженерія програмного забезпечення	ХНУРЕ	Комп'ютерні науки
2	122	Комп'ютерні науки	ХНУРЕ	Комп'ютерні науки
3	123	Комп'ютерна інженерія	ХНУРЕ	Комп'ютерна інженерія та управління
4	124	Системний аналіз	ХНУРЕ	Інформаційно-аналітичних технологій та менеджменту
5	125	Кібербезпека	ХНУРЕ	Комп'ютерна інженерія та управління
6	126	Інформаційні системи та технології	ХНУРЕ	Інформаційні радіотехнології та технічний захист інформації

Редагування

№ Код спеціальності Назва спеціальності Назва закладу

Назва факультету Назва кафедри

Значення тесту №1 Значення тесту №2 Значення тесту №3 Значення тесту №4 Значення тесту №5 Значення тесту №6 Форма навчання

Додати
Змінити
Видалити

Рисунок 4.7 – Вкладка редагування бази знань.

За допомогою вкладки, представленої рисунку 4.7, здійснюється редагування бази знань докладання. Наприклад, додавання нового запису. За допомогою представленої вкладки можна видалити записи з таблиці «Абітурієнти», а також формувати звіт.

Наголосимо на конкретних перевагах автоматизованої системи з елементами інтелектуального підбору для профорієнтаційної роботи [33].

Точність. В даний час фахівці з профорієнтаційної та кадрової роботи відмовляються від традиційних тестів (або сукупності окремих тестів) у зв'язку з їх недостатньою точністю і все більшу перевагу віддають системам штучного інтелекту, зокрема й ЕС.

Зміст параметрів. Традиційні психологічні тести (ММРІ, Люшер, Кеттел і т.п.) не розраховувалися на вирішення кадрових завдань, тому їх параметри швидше розкажуть тестувальнику про особистість здобувача (абітурієнта або фахівця), ніж про його схильність до професійної діяльності. Експертні системи спочатку проектуються на вирішення управлінських і кадрових завдань, тому й набір параметрів у яких відмінний від традиційно-психологічного, хоч і включає відповідні параметри обліку всіх чинників. Особливо це стосується прецедентних експертних систем - наявність прецедентів дозволяє заміряти, наприклад, стійкість у виконанні зобов'язань, ініціативність, вимогливість тощо. Також слід враховувати, що використання ЕС під час обстеження на професійну готовність враховується взаємовплив факторів. Крім того, для конкретної посади можлива розробка ЕС, що визначає успішність (або неуспішність) за умовним сценарієм, що відображає всі особливості цієї посади [34].

Автоматичне багатовимірне індивідуальне налаштування. Прецедентні експертні системи дозволяють користувачеві не лише додавати власні прецеденти, а й будувати профілі наявних конкретних посад згідно зі штатним розписом у багатовимірному режимі з використанням штучного інтелекту на підставі комп'ютерних обстежень своїх працівників. Потім система автоматично налаштовується на прецеденти підприємства-користувача та

максимально використовує це при прийнятті кадрових рішень. Тобто. це «індивідуальне пошиття» закладено в сам алгоритм просунутої прецедентної експертної системи.

Можливість прораховувати на багатовимірних моделях наслідки кадрових рішень. Щодо сучасної дійсності це означає, що при вирішенні будь-якої складної ситуації кадровик може в автоматичному режимі створити кілька моделей цієї ситуації, побачити швидкі та відстрочені плюси та мінуси кожного свого рішення, вибрати оптимальний варіант і потім впроваджувати його. Експериментів із живими людьми стає набагато менше. Звідси – менше конфліктів, менше порожніх «пробних» призначень на посаду тощо.

Крайня простота у роботі. Розроблені експертні системи можуть бути дуже складними всередині, але зазвичай прості у використанні. Настільки простіше, що з ними безпосередньо працює кадровик, менеджер з персоналу, керівник або будь-яка інша особа без залучення програмістів. Працюючи з прецедентними експертними системами конфліктів «кадровик-комп'ютер» немає навіть у тих кадрових фахівців, хто був практично незнайомий з комп'ютером.

4.4 Експериментальна перевірка

Кожного абітурієнта можна представити у вигляді вектора, який складатиметься з його оцінок проходження. Тоді ці вектори можна об'єднати у матрицю оцінок. При цьому, оскільки кожен студент має оцінку з усіх тестів, то розрідженість такої матриці буде дорівнювати 0. Виходячи з цього аналізу переваг та недоліків підходів до побудови рекомендаційних систем, наведених у розділі 2, можна зробити висновок, що колаборативна фільтрація найкраще підходить для СППР, що розробляється. Доцільно використати підхід «користувач-користувач». Він має високу точність за умови великого набору

даних та їх низької розрідженості – саме те, що наявне у нашому дослідженні. При цьому, також відсутня проблема холодного старту для користувачів, тобто будь-який абітурієнт зможе отримати рекомендацію.

Для вхідних даних «тести – оцінка за тест» (i, x) (тобто $u_i = x$, де u – вектор оцінок абітурієнтів):

- розрахувати схожість абітуо на даного абітурієнта на основі їх векторів оцінок;
- відсортувати абітурієнтів за їх схожістю на даного;
- для кожної спеціальності за вибором знайти n абітурієнтів, які найбільше схожі на даного;
- розрахувати середньо зважену оцінку кожної спеціальності серед усіх спеціальностей на основі оцінок абітурієнтів, яких було обрано на кроці 3;
- відсортувати спеціальності за вибором відповідно до їх середньо зважених оцінок, знайдених на кроці 4;
- рекомендувати абітурієнту m верхніх спеціальностей за вибором.

Для оцінювання схожості абітурієнтів будемо використовувати косинус подібності (2.2), для розрахунку прогнозованої оцінки – середньо зважене значення (2.6).

Для спрощення обрахунків та отримання більш коректних результатів, для знаходження коефіцієнту схожості оцінки будемо нормалізувати в інтервал $[0.01; 1.01]$ за формулою (4.1):

$$\text{norm}S(S) = \frac{S - S_{\min}}{S_{\max} - S_{\min}} + 0.01, \quad (4.1)$$

де S – вхідна оцінка;

S_{\min} – мінімально можлива оцінка (у нашому випадку 60);

S_{\max} – максимально можлива оцінка (у нашому випадку 100).

Нехай u_1, u_2 – вектори оцінок двох студентів, u'_1, u'_2 – вектори скоригованих оцінок. Тоді їх схожість можна подати у вигляді $\text{similarity}(u_1, u_2) = \cos(u'_1, u'_2)$.

Якщо $\cos(u'_1, u'_2) = 0$, це означає, що вектори перпендикулярні, тобто студенти абсолютно не схожі один на одного. Якщо $\cos(u'_1, u'_2) = 1$, то вектори або збігаються, або паралельні – студенти однакові за своїми оцінками.

Наприклад, $u_1 = (60, 100, 60)$, $u_2 = (68, 60, 100)$, після нормалізації отримуємо: $u'_1 = (0.01, 1.01, 0.01)$, $u'_2 = (0.21, 0.01, 1.01)$.

Отже результатом розрахунку буде: $\text{similarity}(u_1, u_2) = \cos((0.01, 1.01, 0.01), (0.21, 0.01, 1.01)) \approx 0.09$ – абітурієнти майже не схожі один на одного; $u_1 = (68, 76, 100)$, $u_2 = (76, 68, 76)$.

Після нормалізації отримуємо: $u'_1 = (0.21, 0.41, 1.01)$, $u'_2 = (0.41, 0.21, 0.41)$.

Отже результатом розрахунку буде: $\text{similarity}(u_1, u_2) = \cos((0.21, 0.41, 1.01), (0.41, 0.21, 0.41)) \approx 0.86$ – абітурієнти схожі; $u_1 = (92, 76, 76)$, $u_2 = (76, 76, 68)$.

Після нормалізації отримуємо: $u'_1 = (0.81, 0.41, 0.41)$, $u'_2 = (0.41, 0.21, 0.21)$,

Отже результатом розрахунку буде: $\text{similarity}(U_1, U_2) = \cos((0.81, 0.41, 0.41), (0.41, 0.21, 0.21)) \approx 1$ – абітурієнти однакові за своїми оцінками.

Розглянемо роботу алгоритму на прикладі. Значення n для кроку 3 алгоритму встановимо рівним 5, значення m для кроку 6 рівним 1.

Нехай маємо 5 студентів (u_1, \dots, u_5), 6 спеціальностей (j_1, \dots, j_6), оцінки абітурієнтів цих спеціальностей та інформацію про те, яку спеціальність обрав кожен із абітурієнтів та яку оцінку отримав з тестів. Побудуємо матрицю оцінок «абітурієнт – оцінка за тести» в таблиці 4.1.

Також маємо абітурієнта u_0 для якого необхідно зробити рекомендацію. Його оцінки наведено у таблиці 4.2.

Таблиця 4.1 – Матриця оцінок

	j_1	j_2	j_3	j_4	j_5	j_6
u_1	90	100	69	61	93	91
u_2	81	67	78	92	100	99
u_3	93	83	70	64	79	80
u_4	100	67	72	70	63	67
u_5	80	62	74	91	67	89

Таблиця 4.2 – Вхідний вектор оцінок

	j_1	j_2	j_3	j_4	j_5	j_6
u_0	75	80	63	71	95	56

Нормалізуємо оцінки спеціальності j_1, \dots, j_6 . Наприклад, нормалізація оцінки першого абітурієнта з першою спеціальністю:

$$\text{norm}S(S_{1,1}) = \frac{90 - 60}{100 - 60} + 0.01 = 0.76.$$

Результат нормалізації показано у таблиці 4.3

Таблиця 4.3 – Результат нормалізації оцінок з таблиць 2.1 та 2.2

	j_1	j_2	j_3	j_4	j_5	j_6
u_1	0,76	1,01	0,51	0,485	0,835	0,654
u_2	0,535	0,185	0,46	0,935	1,01	0,74
u_3	0,835	0,585	0,26	0,535	0,485	0,68
u_4	1,01	0,685	0,31	0,785	0,185	0,264
u_5	0,51	0,06	0,36	0,785	0,185	0,75
u_0	0,385	0,51	0,085	0,285	0,885	0,248

Знайдемо схожість ключового студента u_0 з іншими. Наприклад, його схожість на студента u_1 буде розрахована наступним чином:

$$\begin{aligned} \text{similarity}(U_1, U_0) &= \\ &= \cos(0.76, 1.01, 0.51, 0.485, 0.835), (0.385, 0.51, 0.085, 0.285, 0.885) = \\ &= \frac{0.76 * 0.385 + 1.01 * 0.51 + 0.51 * 0.085 + 0.485 * 0.285 + 0.835 * 0.885}{\sqrt{0.76^2 + 1.01^2 + 0.51^2 + 0.485^2 + 0.835^2} \sqrt{0.385^2 + 0.51^2 + 0.085^2 + 0.285^2 + 0.885^2}} \\ &= \frac{1.670404}{1.670404 * 1.131371} = 0.914492 . \end{aligned}$$

Дані про схожість ключового студента на інших наведено у таблиці 4.4.

Таблиця 4.4 – Схожість студента u_0 на інших

	u_1	u_2	u_3	u_4	u_5
u_0	0,914492	0,851011	0,847519	0,626895	0,558029

Розрахуємо середньо зважену оцінку для спеціальності на основі схожості студентів.

Середньо зважені оцінки для спеціальності наведено у таблиці 4.5.

$$u_{0,6} = \frac{91 * 0.914492 + 67 * 0.626895}{0.914492 + 0.626895} = \frac{124.132}{1.526338} = 81.239 .$$

Таблиця 4.5 – Середньо зважені оцінки спеціальності

	j_1	j_2	j_3	j_4	j_5	j_6
u_0	80,2648	75,318	68,264	81,329	74,6589	89,39105

Отже, можемо рекомендувати j_6 спеціальність для цього абітурієнта.

Об'єднаємо отримані результати експерименту у таблиці 4.6.

Таблиця 4.6 – Зведені результати тестування

Вектор оцінок	Рекомендована спеціальність	Середньо зважена оцінка
$u_1=(100; 100; 100; 100; 100; 100; 100; 100)$	Інформаційні системи та технології	~ 85,15
$u_2= (78; 87; 91; 84; 71; 90; 100; 89)$	Інженерія програмного забезпечення	~ 85
$u_3= (74; 91; 100; 79; 60; 78; 83; 84)$	Комп'ютерні науки	~ 82,91
$u_4= (60; 60; 60; 60; 60; 60; 60; 60)$	Інформаційні системи та технології	~ 85,15

Можна помітити, що результати для u_1 та u_4 ідентичні, попри те, що їх оцінки абсолютно різні. Але це правильне поведження системи при використанні косинусу схожості (бо вектори їх оцінок є колінеарними) і його можна пояснити логічно: ці абітурієнти мають схильності до вивчення спеціальності на одному рівні, без перекосів у яку-небудь область, тому і рекомендація для них однакова.

Проведений експеримент показав, що використання методу колаборативної фільтрації дозволило отримати доволі точний результат, але проблема використання методу полягає в складності обробки даних за рахунок їх об'ємності.

В перспективі слід розширити інформаційну базу функціонування системи та збільшенні кількості функцій, наприклад додати рекомендування ще декількох спеціальностей на основі результатів тестування. Також необхідно впровадити метрик оцінки якості рекомендацій.

ВИСНОВКИ

У ході роботи було проведено дослідження літератури та наявного досвіду в країні в сфері профорієнтаційної роботи серед абітурієнтів. Виявлені проблеми та особливості профорієнтаційної роботи в умовах війни. У результаті аналізу було запропоновано модель системи на основі методів колаборативної фільтрації.

В результаті проведеного дослідження також були вирішені такі завдання:

- розглянуті основні поняття теорії прийняття рішень;
- виконаний огляд розробок і проаналізовані підходи до реалізації існуючих програмних продуктів для проведення складних експертиз;
- на основі результатів дослідження розроблено оригінальний програмний продукт для підтримки прийняття рішень;
- виконано тестування програми на конкретній задачі багатокритеріального вибору.

На основі обраного підходу та виходячи з вимог до системи підтримки прийняття рішень було розроблено алгоритм її роботи. У якості мови програмування для реалізації було обрано C #, для зберігання даних про оцінки було використано СУБД ІВExpert, для розробки користувацького інтерфейсу обрано Visual Studio.

Було спроектовано базу даних та структуру класів додатку. На їх основі з використанням перелічених інструментів було розроблено додаток, який було успішно протестовано.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Про затвердження Концепції державної системи професійної орієнтації населення [Електронний ресурс] : Постанова КМ України № 842 від 17.09.2008 р. Режим доступу : <https://zakon.rada.gov.ua/laws/show/842-2008-%D0%BF#Text>.
2. Зінченко В.П., Янцур М.С. Теорія і практика розбудови системи професійної орієнтації в сучасних умовах // Оновлення змісту і методі психології освіти та професійної орієнтації, 1998. Вип. 4. С. 4-15.
3. Тхоржевський Д.О. Методика трудового і професійного навчання та викладання загальнотехнічних дисциплін, 1992. Вища школа. 332 с.
4. Федоришин Б. О. Психолого-педагогічні основи професійної орієнтації: Автореф. дис. докт. пед. наук: 13. 00. 04. / ІППО АПН України. К., 1996. 49 с.
5. Професійна діагностика / Упорядник Т. Гончаренко. К.: Ред. загальнопед. газ., 2004. 120 с.
6. Москалець В. П. Психологія особистості: Навчальний посібник. / В. П. Москалець, 2013. К.: Видавництво «Центр учбової літератури». 262с.
7. Вороніна Г., Гуцан Л. Проблеми професійної орієнтації старшокласників: англійський досвід та українські реалії. Педагогічні науки: Теорія, історія, інноваційні технології. 2019. № 5. С. 24-35.
8. Зайцева О. М., Мичка Д. В. Проектні технології в організації та управлінні професійною орієнтацією учнів у закладах загальної середньої освіти. Вісник Глухівського національного педагогічного університету імені Олександра Довженка. Сер.: Педагогічні науки. 2019. Вип. 3. С. 113-121.
9. Recommendation Systems – How Companies are Making Money URL:<https://sigmoidal.io/recommender-systems-recommendation-engine/> (Last accessed: 29.12.2023).
10. Daniar Asanov (n.d.) Algorithms and Methods in Recommender

Systems, Berlin, Germany: Berlin Institute of Technology. 2011. P. 1-7.

11. Gleb Beliakov, Tomasa Calvo and Simon James. Aggregation of preferences in recommender systems / Gleb Beliakov and Simon James // School of Information Technology, Deakin University. 2008. P. 705-735.

12. M. Balabanovic Fab: Content-Based, Collaborative Recommendation / M.Balabanovic, Y. Shoham // Comm. ACM. 1997. Vol. 40, No3. P.66-72.

13. Collaborative Filtering for Implicit Feedback Datasets URL: <http://yifanhu.net/PUB/cf.pdf> (Last accessed: 01.01.2024).

14. Hybrid Recommender Systems: Survey and Experiments. [Электронный ресурс] - Режим доступа до ресурсу: <https://pdfs.semanticscholar.org/5880/b9bc3f75f4649b8ec819c3f983a14fca9927.pdf>.

15. Retailrocket recommender system dataset [Электронниий ресурс]. Режим доступа до ресурсу: <https://www.kaggle.com/retailrocket/ecommerce-dataset#events.csv>.

16. McLeod D. Collaborative Filtering for Information Recommendation Systems [Электронный ресурс]. D. McLeod, A. Y. Chen. Research Reports. 2009. Режим доступа до ресурсу: http://research.create.usc.edu/cgi/viewcontent.cgi?article=1101&context=nonpublished_reports.

17. A.I. Schein Methods and Metrics for Cold-Start Recommendations / A.I. Schein, A.Popescul, L.H. Ungar, D.M. Pennock. Proc. 25th Ann. Int'l ACM SIGIR Conf. 2002. P. 253-260.

18. Dunning T. Practical Machine Learning: Innovations in Recommendation / T. Dunning, E. Friedman, A. Felfernig. O'Reilly Media, Inc., 2014. 56 с.

19. Aggarwal C. Recommender Systems. Charu Aggarwal., 2016. 498 с.

20. J. Basilico and T. Hofmann. Unifying collaborative and content-based filtering In Proceedings of the 21th International Conference on Machine Learning, 2004. P. 9-16.

21. B. Sarwar Item-Based Collaborative Filtering Recommendation Algorithms / B.Sarwar, G. Karypis, J. Konstan, J. Riedl // Proc. 10th Int'l WWW

Conf. 2001. P. 285-295.

22. A. Nakamura Collaborative Filtering Using Weighted Majority Prediction Algorithms / A. Nakamura, N. Abe. Proc. 15th Int'l Conf. Machine Learning. 1998. P.395-403.

23. A Singularly Valuable Decomposition: The SVD of a Matrix
 URL:[https://datajobs.com/data-science-repo/SVD-\[Dan-Kalman\].pdf](https://datajobs.com/data-science-repo/SVD-[Dan-Kalman].pdf) (Last accessed: 24.12.2023).

24. Метрики качества ранжирования [Электронный ресурс] // HABR. 2016. URL: <https://m.habr.com/ru/company/econtenta/blog/303458/>. (Last accessed: 28.12.2023).

25. Блейхер В.М., Бурлачук Л.Ф. Психологічна діагностика інтелекта і особистості/ В.М. Блейхер , Л.Ф. Бурлачук . СПб.: Петербург, 2003. 351 с.

26. Zajceva T.V., Igrunova S.V., Putivceva N.P., Pusnaja O.P., Nesterova E.V. 2011. The use of a semiotic approach to knowledge representation to build a model of logical structure of educational material, the Series History. Political science. Economy. Informatics. Belgorod Nauchnye vedomosti BelGU, 19(108): 143-149.

27. Katrenko, A. V., Pasichnyk, V. V., & Pasko, V. P. (2009). Teoriia pryiniattia rishen: pidruch. dlia stud. vyshiv. Kyiv: BHV, p. 448.

28. Шапуго, В. Д. (1996). Керування проєктами: підруч. дlia vishiv. Sankt-Peterburh: Dvatry, p. 388.

29. Дудзяний І.М. Об'єктно-орієнтоване моделювання програмних систем. Львів: Видавничий центр ЛНУ імені Івана Франка, 2007. 108 с

30. Голуб Б.М. С#. Концепція та синтаксис. Львів: Видавничий центр ЛНУ імені Івана Франка, 2006. 136 с.

31. Пасічник В.В. Організація баз даних та знань: підручник для ВНЗ/ В.В. Пасічник, В.А. Резніченко. К.: Видавнича група BHV, 2006. 384с.

32. Литвин В.В. Методи та засоби інженерії даних та знань: навч. посіб. для студ. вищ. навч. закл. / В. В. Литвин // М-во освіти і науки, молоді та спорту України. Л.: Магнолія 2006, 2012. 248 с.

33. Fel'dshtejn D.I. 2012. Age and pedagogical psychology: Selected psychological works. M., MPSI, 427.

34. Zeer Je.F., Pavlova A.M., Sadovnikova N.O. 2006. Proforientacii: Theory and practice: proc.manual.for the higher school. M., Akademicheskij proekt; Ekaterinburg, Delovaja kniga, 192.

35. Методичні вказівки щодо розробки та оформлення кваліфікаційної роботи (для студентів усіх форм навчання другого (магістерського) рівня програми "Інформаційні управляючі системи та технології) / Упоряд.:Петров К.Е., Левикін В.М., Чалий С.Ф., Євланов М.В., Саєнко В.І., Міхнов Д.К., Міхнова А.В., Чала О.В. Харків: ХНУРЕ,2021. 30с.

36. ДСТУ 3008:2015. Інформація та документація. Звіти у сфері науки і техніки. Структура і правила оформлювання. Чинний від 22.06.2015. Київ: ДП «УкрНДНЦ», 2016. 31 с.

37. ДСТУ 8302:2015. Інформація та документація. Бібліографічні посилання. Загальні положення та правила складання. Чинний від 04.03.2016. Київ: ДП «УкрНДНЦ», 2016. 20 с.