

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет _____ *Комп'ютерних наук* _____
(повна назва)

Кафедра _____ *Системотехніки* _____
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

другий (магістерський)
(освітньо-кваліфікаційний рівень)

Розробка та дослідження аналітичних методів компоненту CRM системи

на прикладі аптечної мережі

(тема роботи)

Виконав:

Здобувач групи СПРМ-20-1
Щебликіна Д. Ю.
(прізвище, ініціали)

Спеціальність: 122 – Комп'ютерні науки
(код і повна назва спеціальності)

Тип програми: Освітньо-професійна
(освітньо-професійна або освітньо-наукова)

Освітня програма: Системне проектування
(повна назва освітньої програми)

Керівник проф. Ситніков Д. Е.
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри СТ

(підпис)

проф. Гребеннік І.В.

(прізвище, ініціали)

2021 р.

Харківський національний університет радіоелектроніки

Факультет _____ Комп'ютерних наук _____

Кафедра _____ Системотехніки _____

Рівень вищої освіти _____ другий (магістрський) _____

Спеціальність _____ 122 – Комп'ютерні науки _____
(код і повна назва)

Тип програми: _____ Освітньо-професійна _____
(освітньо-професійна або освітньо-наукова)

Освітня програма _____ ООП _____ Системне проектування _____
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____ СТ _____

_____ проф. Гребеннік І.В. _____

" _____ " _____ 2021 р.

ЗАВДАННЯ

НА КВАЛІФІКАЦІЙНУ РОБОТУ

Студентові _____ Щебликіній Дар'ї Юріївні _____
(прізвище, ім'я, по батькові)

Тема роботи Розробка та дослідження аналітичних методів компоненту CRM системи на прикладі аптечної мережі

затверджена наказом по університету від " 8 " листопада 2021р. № 1663 Ст

2. Термін подання студентом роботи 17 грудня 2021 р.

Вихідні дані до роботи (проекту) Розробка та дослідження аналітичних методів компоненту CRM системи на прикладі аптечної мережі. Система повинна являти собою web-додаток з інтерфейсом доступу до бази даних. Перелік використовуваних програмних засобів: ОС Microsoft Windows 10, IntelliJ IDEA Ultimate 2021.2.1, Tomcat Apache 9.0, Erwin Data Modeler, AllFusion Process Modeler, Rational Rose, MySQL server, MySQL Workbench , Технічне забезпечення: IBM-сумісний ПК з MPI Core 2 Duo та вище

3. Зміст пояснювальної записки (перелік питань, що потрібно розробити) 3.1 Вступ, 3.2 Аналіз предметної області, 3.3 CRM системи в інформаційних технологіях, 3.4 CRM система та її компоненти, 3.5 Аналіз існуючих систем аптечних мереж, 3.6 Доцільність введення покращень, 3.7 Огляд методів та технологій, що використовуються в предметній області, 3.8 Методи кластеризації даних, 3.9 Рекомендаційні системи та їх різновиди, 3.10 Виявлення закономірностей на основі асоціативних правил, 3.11 Постановка задачі, 3.12 Обмеження предметної області, 3.13 Постановка задачі дослідження та її формалізація, 3.14 Етапи проведення досліджень, 3.15 Розробка методу, 3.16 Аргументація обраних методів, 3.17 Розроблений метод, 3.18 Варіанти оптимізації запропонованого алгоритму, 3.18 Програмна реалізація та інтеграція методу, 3.19 Вибір технологій для проектування та програмування, 3.20 Уточнення функціональних вимог, 3.21 Розширення діаграми варіантів використання, 3.22 Розширення логічної та фізичної моделювання даних аптечної мережі, 3.23 Інтеграція методу до системи, Висновки, Перелік джерел посилання, Додатки.

4. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень, плакатів)

4.1 Етапи аналітичного компоненту CRM системи (1 аркуш формату A4), 4.2 Алгоритм агрегативної ієрархічної кластеризації (1 аркуш формату A4), 4.3 Алгоритм методу кластеризації k-

середніх (1 аркуш формату А4), 4.4 Алгоритм Apriori (1 аркуш формату А4), 4.5 Алгоритм другого етапу розробленого методу (1 аркуш формату А4), 4.6 Контекстна діаграма системи «Продажу препаратів» (1 аркуш формату А4), 4.7 Декомпозиція контекстної діаграми системи «Продажу препаратів» (1 аркуш формату А4), 4.8 Декомпозиція цільової функції «Адміністрування» (1 аркуш формату А4), 4.9 Контекстна DFD діаграма «Продажу препаратів» (1 аркуш формату А4), 4.10 Декомпозиція контекстної DFD діаграми «Продажу препаратів» (1 аркуш формату А4), 4.11 Декомпозиція роботи DFD діаграми «Адміністрування» (1 аркуш формату А4), 4.12 Діаграма використання системи E-commerce «Аптечної мережі» (1 аркуш формату А4), 4.13 Результат логічного моделювання (1 аркуш формату А4), 4.14 Результат фізичного моделювання (1 аркуш формату А4).....

6. Консультанти з роботи із зазначенням розділів роботи, що їх стосуються

Найменування Розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		(підпис)	(дата)
			

КАЛЕНДАРНИЙ ПЛАН

Пор. №	Назва етапів атестаційної роботи	Термін виконання етапів роботи	Примітка
1.	Аналіз завдання, літератури та аналогів з теми атестаційної роботи	14.09.21 – 20.09.21	
2.	Огляд існуючих підходів, технологій та методів, що використовуються в CRM системах	21.09.21 – 30.09.21	
3.	Постановка задачі розробки та дослідження	01.10.21 – 13.10.21	
4.	Порівняння існуючих розглянутих методів у рамках досліджуваної предметної області	14.10.21 – 27.10.21	
5.	Розробка методу для аналітичного компоненту CRM системи	28.10.21 – 06.11.21	
6.	Розширення спроектованої системи з урахування введення методу в використання	07.11.21 – 21.11.21	
7.	Програмування інтегрованих компонентів	20.11.21 – 30.11.21	
8.	Оформлення пояснювальної записки та програмної документації	01.12.21 – 05.12.21	
9.	Оформлення графічної частини та презентаційних матеріалів комп'ютерного захисту	06.12.21 – 10.12.21	
10.	Представлення на рецензування	15.12.2021	
11.	Представлення атестаційної роботи до ЕК	10.12.21 – 17.12.21	

Дата видачі завдання _____

Студент _____ Щебликіна Д. Ю.
(підпис)

Керівник роботи _____ проф Ситніков Д. Е.
(підпис)

РЕФЕРАТ

Пояснювальна записка до атестаційної роботи містить 74 аркушів, 20 рисунків, 2 таблиці, 25 джерел, 3 додатки. Графічна частина містить 14 плакатів.

КОМПОНЕНТИ CRM-СИСТЕМИ, МЕТОДИ ІНТЕЛЕКТУЛЬНОЇ ОБРОБКИ ДАНИХ, КЛАСТЕРИЗАЦІЯ, РЕКОМЕНДАЦІЙНІ СИСТЕМИ, АСОЦІАТИВНІ ПРАВИЛА

Об'єктом дослідження є аналітичні методи компоненту CRM системи на прикладі аптечної мережі.

Предметом дослідження виступають методи інтелектуального аналізу даних для заохочення клієнтів.

Метою роботи є дослідження існуючого переліку методів інтелектуального аналізу даних для заохочення клієнтів та розробка власного методу на їх основі в рамках визначеної предметної області.

Методами розробки являються системний підхід та об'єктно-орієнтований аналіз.

Результатами роботи є розроблений та інтегрований метод до системи електронної комерції «Аптечної мережі», що дозволяє на базі збережених у системі даних створювати рекомендації для кожної продукції й виокремлювати таргетовані пропозиції для клієнтів застосунку, для подальшого використання у рекламних розсилках.

ABSTRACT

Explanatory note contains 74 sheets, 20 figures, 2 tables, 3 annexes, 25 sources; graphic material - 14 sheets.

COMPONENTS OF CRM-SYSTEM, METHODS OF INTELLECTUAL DATA PROCESSING, CLUSTERIZATION, RECOMMENDATION SYSTEMS, ASSOCIATIVE RULES

The object of research is the analytical methods of the CRM system component on the example of a pharmacy network.

The subject of the study is methods of data mining to encourage customers.

The aim of the work is to study the existing list of methods of data mining to encourage customers and develop a method based on them within a certain subject area.

Development methods are systems approach and object-oriented analysis.

The results are a developed and integrated method to the e-commerce system "Pharmacy Network", which allows on the basis of data stored in the system to create recommendations for each product () and identify targeted proposals for customers, for further use in advertising.

ЗМІСТ

	ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ	7
	ВСТУП.....	8
1	АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ.....	10
1.1	CRM системи в інформаційних технологіях.....	10
1.1.1	CRM система та її компоненти	12
1.1.2	Аналіз існуючих систем аптечних мереж. Доцільність введення покращень.....	14
1.1.3	Обґрунтування необхідності розробки методу	20
2	ОГЛЯД МЕТОДІВ ТА ТЕХНОЛОГІЙ, ЩО ВИКОРИСТОВУЮТЬСЯ В ПРЕДМЕТНІЙ ОБЛАСТІ.....	22
2.1	Методи кластеризації даних.....	23
2.2	Рекомендаційні системи та їх різновиди.....	28
2.3	Виявлення закономірностей на основі асоціативних правил.....	34
3	ПОСТАНОВКА ЗАДАЧІ.....	38
3.1	Обмеження предметної області.....	38
3.2	Постановка задачі дослідження та її формалізація.....	38
3.3	Етапи проведення досліджень.....	41
4	РОЗРОБКА МЕТОДУ	43
4.1	Аргументація обраних методів.....	43
4.2	Розроблений метод.....	45
4.3	Варіанти оптимізації запропонованого алгоритму.....	47
5	ПРОГРАМНА РЕАЛІЗАЦІЯ ТА ІНТЕГРАЦІЯ МЕТОДУ.....	49
5.1	Вибір технологій для проектування та програмування.....	49
5.2	Уточнення функціональних вимог.....	52
5.3	Розширення діаграми варіантів використання.....	58

3		
5.	Розширення логічної та фізичної моделювання даних системи e-	60
4	commerce «Аптечної мережі».....	
5.	Інтеграція методу до системи.....	68
5		
	ВИСНОВКИ.....	71
	ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ.....	72
	Додаток А.....	75
	Додаток Б.....	91

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І
ТЕРМІНІВ

CRM – Customer Relationship Management

DDL – Data Definition Language

DFD – Data Flow Diagram

FK – Foreign Key

HQL – Hibernate Query Language

IDEF0 – Identity Ecosystem Framework 0

JDBC – [Java Database Connectivity](#)

JSP – Java Server Pages

JSTL – JSP Standard Tag Library

MVC – [Model View Controller](#)

PK – Primary Key

SQL – Structured Query Language

СУБД – Система Управління Базами Даних

ВСТУП

Інформаційні технології все міцніше пов'язані з процесами сьогодення завдяки доступному Інтернету та простим у користуванні браузерам.

Такі зміни найяскравіше відобразились у бізнесі, адже відтепер немає необхідності орендувати приміщення, обставляти їх, наймати працівників – необхідно тільки впровадити систему електронної комерції.

Подібні системи у повній мірі містять функціонал для комфортної закупівлі: пошуку товарів або послуг, їх оформлення, сплати, введення документообігу. Система також виконує й неочевидні для користувача функції – збір та накопичення даних.

Дані є одним з найцінніших ресурсів, бо являються основою для створення інформації та знань в пов'язаних з ними областях. Наприклад, знанням може виступати чіткий профіль користувача, або прогнозування найбільш прибуткових напрямлень за динамікою продажів, тощо. Коректне застосування отриманого знання приносить прибуток його власнику.

Разом з перевагами та зменшенням вкладу для власників бізнесу з'являється суттєвий недолік – необхідність ще більше утримувати клієнтів, котрі, в свою чергу, простіше можуть змінювати місце придбання товару або послуги.

Так, зумовлюється широке використання CRM систем, їх інтеграція вводиться, насамперед, для управління взаємодіями з клієнтами та збільшення їх задоволення від користування системою.

Розглянемо можливість відтворення процесів підтримки на прикладі аптечної мережі. Продаж ліків є поширеною та конкурентною галуззю для бізнесу, тому важливо прикласти максимум зусиль для утримання клієнта в конкретній аптечній мережі.

Одним з найпростіших методів утримання клієнтів є рекламна розсилка. Особливо ефективно вона відпрацьовує, коли товари попередньо відбираюся у

відповідності до вподобань клієнта. В даній предметній області виявлення цільової рекомендації представляє досить складний процес, адже товари у кошику не будуть постійно однотипними й сформувати профіль користувача досить складно.

Іншим варіантом утримання клієнта є підказки безпосередньо під час створення замовлення. Наприклад, доречно нагадати покупцю попікнутися про перевезення препарату, якщо він цього потребує за інструкцією. Менш тривіальним прикладом є нагадування придбати пробіотик при покупці антибіотиків, тощо.

Під час професійної практики ставиться за мету розглянути методи, що використовуються у CRM системах, обрати найбільш підходящий до предметної області або створити власний метод.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 CRM системи в інформаційних технологіях

Інформаційні технології є сукупністю передачі та зберігання інформації, розробки, встановлення, впровадженням і керуванням комп'ютерними системами в компаніях, університетах та інших різноманітних організаціях.

Інформаційні технології в сукупності з банками даних та комплексами програмно-технічних засобів формують інформаційні системи.

Основні класифікації наведено у класифікації наведено у таблиці 1.1.

Таблиця 1.1 – Класифікація інформаційних технологій

Назва	Складові
предметні області використання	Фінансова
	Виробнича
	Банківська
	Маркетингова
	кадрова діяльність
характер обробки даних	інформаційно-довідкові
	інформаційно-пошукові
	АСУ
	СППР
види процесів управління	управління технологічними процесами
	організаційно-технічними процесами
	інтегрованими процесами
	корпоративними процесами
	науково-дослідними процесами
	навчальними процесами
спосіб охоплення завдань	електронна обробка даних
	експертні системи
	робота з текстовим редактором
	СУБД
	ППР
	автоматизація функцій управління

Найбільш цікавими інформаційними системами для розгляду є системи підтримки виробничого циклу. До них відносяться наступні системи:

1. MRP – система для планування вимог до матеріалів;
2. ERP – система комплексного планування ресурсів підприємства;
3. CRM – система управління відносинами з клієнтами;
4. SCM – система управління поставками.

Перелічені вище системи добре справлялися з поставленими задачами у свій час, але сьогодення потребує включення в обіг Інтернету як платформи для ведення бізнесу.

Електронний бізнес представляє обмін товарами та послугами між підприємствами, групами та окремими особами через застосування сучасних інформаційних та комунікаційних технологій.

Впровадження системи електронного бізнесу дозволяє компанії знизити витрати та збільшити прибуток за рахунок:

5. оптимізації внутрішніх та зовнішніх інформаційних потоків;
6. прискорення бізнес-процесів;
7. здобуття нових каналів маркетингу та дистрибуції;
8. підвищення якості обслуговування клієнтів та ефективності роботи з дистриб'юторами.

Різновидом електронного бізнесу є електронна комерція – реалізований електронним способом процес, що надає можливість здійснювати купівлю та продаж товарів або послуг через Інтернет, а також передачу грошей та даних для завершення продажу.

Перевагами ведення електронної комерції виступають: відкрите відношення до клієнтів, взаємодія, що направлена на встановлення довгострокових відносин (CRM), скорочення циклу маркетингу та продажу, поява можливості пост-підтримки продукту (стосується технічного забезпечення) тощо.

Недоліками електронної комерції є необхідність придбання спеціалізованих програмних засобів, ризик втрати критично важливої для бізнесу інформації через провайдера.

1.2 CRM система та її компоненти

Зі збільшенням доступності персональних пристроїв будь-яка людина може легко отримати всю необхідну інформацію щодо переваг та недоліків тієї чи іншої компанії. Саме тому для бізнесу зростає необхідність зберегти та примножити лояльність існуючих та потенціальних клієнтів.

В цьому питанні на допомогу приходить інтегрування CRM систем – систем управління взаємовідносинами з клієнтами. Очікуваними результатами є збільшення прибутку та зменшення витрат:

9. приріст нових клієнтів;
10. крос-продаж товарів;
11. пізнаваність бренду;
12. збільшення продуктивності;
13. прискорення ознайомлення клієнтів з новими пропозиціями.

CRM система поділяється на компоненти: оперативний, аналітичний та колабораційний [1].

Оперативний компонент CRM системи проектується задля підвищення лояльності клієнтів під час безпосередньої роботи з ними. Такий результат забезпечується за рахунок реєстрації вхідного трафіку, збереження в базах даних інформації про клієнтів, шуканих запитів та замовлень, автоматизованих документообігів та нагадуваннях про заплановані дзвінки, листи, зустрічі тощо.

В аналітичному компоненті CRM системи ставиться за мету не фіксація та збереження, а дослідження наявних даних та виявлення закономірностей. Дані системи мають функції для сегментації клієнтської бази, визначення цінності клієнта, його рентабельності, аналізу ефективності маркетингових інструментів та прогнозування.

Колабораційний компонент CRM системи використовується для отримання зворотного зв'язку. Спираючись на отриману інформацію, компанія може вносити корективи в асортимент, змінювати цінову політику, покращувати загальний супровід клієнта.

Розглядаючи web-сайт або застосунок для ведення роздрібної торгівлі на прикладі аптечної мережі, визначаємо занеобхідним подальший розгляд саме аналітичного компоненту CRM системи.

Аналітичний компонент дозволяє розглядати клієнта в чотири етапи (рис. 1.1): ідентифікація, залучення, утримання та розвиток [2, 3]. Утворена послідовність досліджень дозволяє краще взаємодіяти з клієнтами системи.

Процес ідентифікації полягає в сегментації та аналізі цільових клієнтів. Сегментація передбачає поділ всіх клієнтів системи за схожими характеристиками, які визначаються під час аналізу цільових клієнтів – їх якісних та кількісних даних.

Залучення відбувається за сегментом, до якого віднесено клієнта. Даний етап складається з вибору каналу через який будуть надсилатися повідомлення та їх зміст для кожного сегменту. Вибір каналів зв'язку залежить від їх характеристик, основними з яких є вартість та профіль користувачів. Наприклад, телебачення та радіо є гарними каналами для поширення бренду, але не дозволяють орієнтуватися на певний сегмент населення. Друковані видання мають точну цільову аудиторію, що дозволяє працювати з потенційно зацікавленими людьми. Найпоширенішим рішенням є використання прямого маркетингу – поштова розсилка, надання акційних купонів, тощо.

Розвиток клієнта передбачається в збільшенні кількості замовлень, їх вартості та загальної прибутковості клієнта для компанії. Процес реалізується за рахунок виявлених закономірностей під час аналізу кошика ринку – введення крос-продажів.

На етапі утримання клієнтів ставиться за мету фіксація задоволеності клієнтів на належному рівні, що знижує вірогідність їх переходу до конкурентів. Задоволеність визначається як співвідношення очікувань до

сприйняття. Сприйняття базується на задоволенні від пропонованої вартості та накопиченні якісного досвіду за час співпраці з компанією. Означене реалізується шляхом введення стратегій, що базуються на індивідуальних пропозиціях та програмах лояльності.

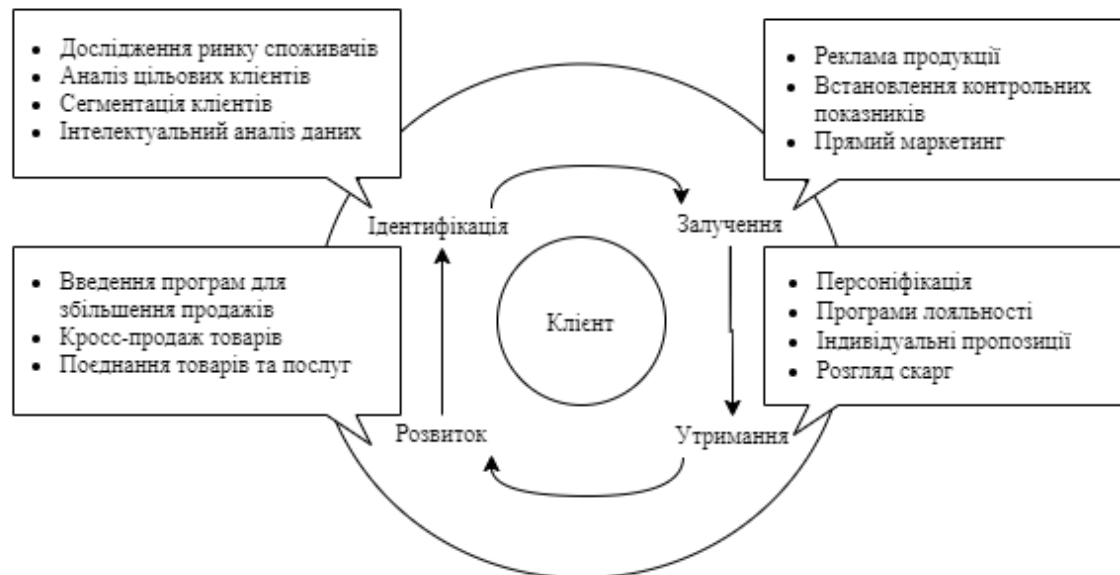


Рисунок 1.1 – Етапи аналітичного компоненту CRM системи

1.3 Аналіз існуючих систем аптечних мереж

В сучасному світі все більше і більше компаній паралельно з фізичними точками продажу створюють платформи для продажу в Інтернеті – електронні комерції. Таким чином, бізнес отримує клієнтів не тільки завдяки зручному розташуванню фізичних точок, аптеки у даному випадку, але й завдяки наданню інформації через сайт. Дійсно, лікарські засоби – це один з тих товарів, які з точки зору споживачів найбільш підходять до покупки он-лайн: їх не потрібно міряти (на відміну від одягу), відсутня необхідність розуміння їх точного розміру (на відміну від предметів інтер'єру), їх не потрібно перевіряти/випробовувати (на відміну від побутової техніки) тощо. Можливості подібного роду покращують враження клієнта від покупки та взаємодії з

мережею, адже уся інформація щодо товару, його кількості, місцезнаходження та вартості знаходиться у відкритому доступі.

З точки зору бізнесу сайти надають можливість для кращого розуміння та заохочення клієнтів – реалізується шляхом збору та аналізу даних, тобто інтегрування CRM-систем.

Слабкі та сильні сторони вже існуючих систем розглянемо на прикладі аптечних мереж «Аптека низьких цін» (АНЦ) та «Аптека 911».

Основними критеріями розгляду виступають дані, які накопичують системи, загальні та індивідуальні рекламні пропозиції:

- 14.дані необхідні для реєстрації в системі;
- 15.додаткові параметри, що збирає система;
- 16.рекламні пропозиції, які розміщені на сайті;
- 17.розсилка (та її змістовність) в залежності від замовлень за період існування кабінету;
- 18.додаткові функції;
- 19.програми лояльності;
- 20.рекомендації, щодо аналогів та супутніх товарів.

Інтерфейс «Аптеки низьких цін» виконаний в жовто-білій кольоровій палітрі, що не є випадковим (рис 1.2). Дослідниками при вивченні аспектів психології кольору та складанні колірної товарної матриці було доведено роль кольорів та їх вплив на поведінку споживачів, оскільки вони звертаються до почуття людини. Так, білий – це товари функціональної потреби, а жовтий – емоційні. За різними інформаційними джерелами жовтий колір (разом із червоним, синім та зеленим) відносять до «фокусних маркерів», що сприймає людство; його також співставляють із кольором молодості та оптимізму, а також тим, який довше за інших лишається у пам'яті.

На головній сторінці відображено усі акції: деякі з них розміщені у каруселі, для перегляду інших необхідно прокрутити сторінку сайту.

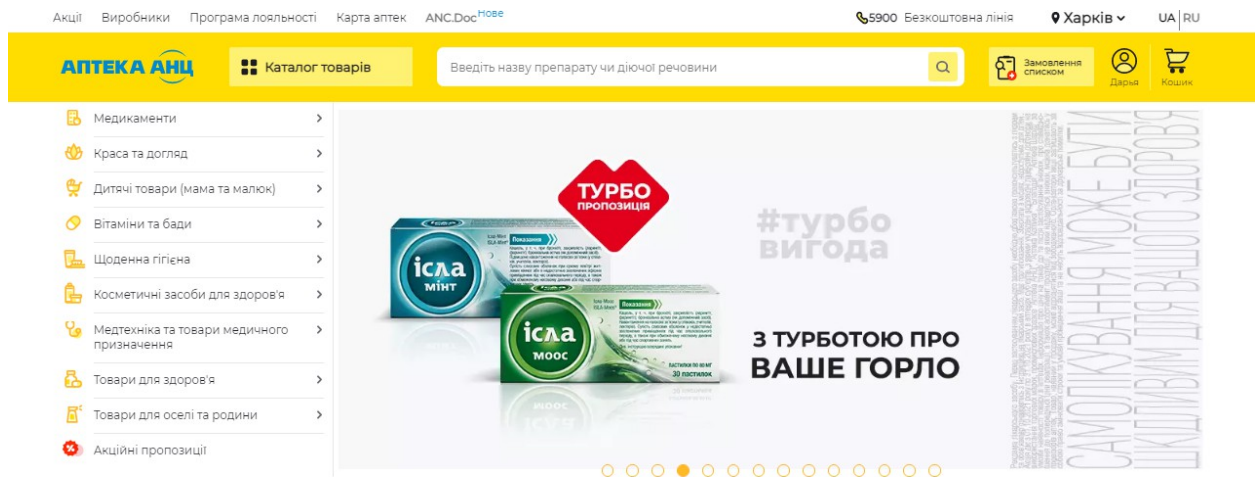


Рисунок 1.2 – Інтерфейс «Аптеки низьких цін»

Для реєстрації в системі необхідно вказати мобільний телефон та пароль. В налаштуванні особистого профілю є можливість редагувати введенні дані (рис 1.3).

 The screenshot shows the 'Персональные данные' (Personal Data) profile editing page. It is divided into three sections: 'Персональные данные', 'Контакты', and 'Дополнительная информация'. The 'Персональные данные' section contains fields for 'Имя' (Name), 'Фамилия' (Surname), 'Отчество' (Patronymic), 'День рождения' (Date of Birth), 'Пол' (Gender), and 'Язык' (Language). The 'Контакты' section contains fields for 'Телефон' (Phone) and 'Электронная почта' (Email). The 'Дополнительная информация' section contains checkboxes for 'У меня есть автомобиль' (I have a car) and 'У меня есть ребенок' (I have a child).

Персональные данные		
Персональные данные		
Имя	Фамилия	Отчество
Дарья	Не указано	Не указано
День рождения	Пол	Язык
Не указано	Жіноча	Не указано
Контакты		
Телефон	Электронная почта	
	daria.shcheblykina@nure.ua	
	Необходимо подтвердить электронную почту для дальнейшего использования сайта	
Дополнительная информация		
☐ У меня есть автомобиль	☐ У меня есть ребенок	

Рисунок 1.3 – Сторінка редагування профілю «АНЦ»

Також є додаткові поля для введення прізвища, ім'я та по-батькові, статі, мови, дати народження, електронної скриньки. Окремо розташовані спеціальні

чек бокси для відмітки наявності автомобіля та дітей. Жодної агітації, щодо заповнення даних на сайті немає, крім підписки на розсилку.

Функціонал сайту надає можливість створювати замовлення як з реєстрацією, так і без неї. Обов'язковими полями для заповнення є мобільний телефон та ім'я, поле з адресою поштової скриньки є необов'язковим.

Пошук здійснюється за діючою речовиною або назвою лікарського засобу (ЛЗ). На сторінці препарату наведені опис, ціни, інструкція, аналоги та замінники, кнопка додання до кошика. Нижче розташована карта аптечних закладів із зазначенням наявної кількості засобу в кожному з них. Існує можливість одразу обрати аптеку для замовлення.

Додатковою функцією є пошук списком: користувач вводить необхідний перелік ЛЗ, вказує кількість упаковок або блістерів й переглядає весь перелік аптек, в котрих необхідна кількість товарів є у наявності.

Присутня програма лояльності «Лайк», за котрою один відсоток від покупки йде на рахунок в еквіваленті один відсоток – одна гривня. При наступних покупках в аптеці можна використовувати накопичені кошти для зменшення поточного чеку.

Розсилка представляє собою шаблони в різних стилях, що заповнюються акційними товарами різних категорій та призначень або реклама бренду.

Сайт «Аптеки 911» виконаний у біло-червоних відтінках (рис. 1.4). Червоний колір стимулює мозок, він спроможний викликати відчуття потреби та дії. Як і на сайті «Аптеки низьких цін» є пов'язаність кольору із білим, на фоні якого по-перше будь-який колір виглядає яскравіше, а по-друге це єдиний колір спокою та чистоти, якому властива завершеність і який демонструє остаточне рішення.

Головна сторінка сайту представлена акційними товарами.

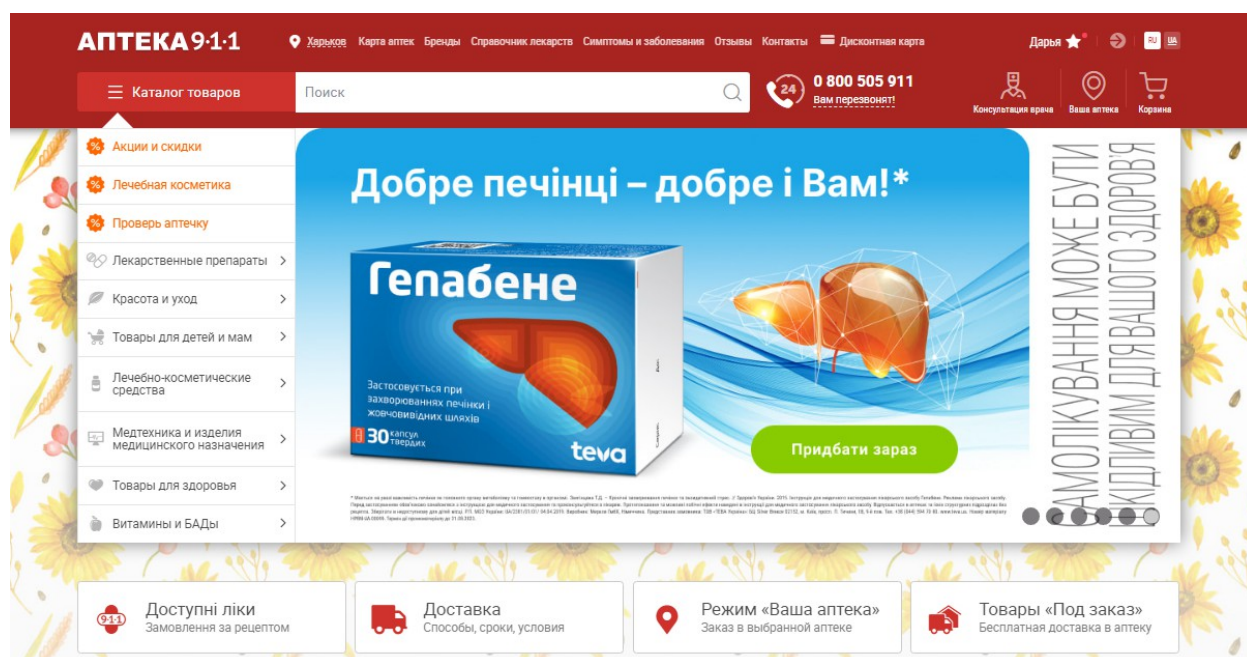


Рисунок 1.4 – Интерфейс «Аптеки 911»

Для реєстрації в системі необхідно вказати ім'я, прізвище, мобільний телефон, поштову скриньку та пароль. В особистому кабінеті ці дані можна редагувати й вносити додаткову інформацію – по-батькові, стать, дату народження та місто (рис 1.5).

Рисунок 1.5 – Сторінка редагування профілю «Аптеки 911»

На сайті можна створювати замовлення без наявності особистого кабінету. Обов'язковими полями для заповнення є прізвище, ім'я, мобільний телефон, поле з адресою поштової скриньки є необов'язковим.

Пошук ЛЗ та їх замовлення здійснюються так само, як і в попередній аптеці. Особливістю є наявність додаткового розділу «Рекомендуємо переглянути». До нього заносяться ЛЗ, які поєднані одними категоріями та ранжовані за популярністю, котра базується на фактичній кількості замовлень кожної позиції клієнтами сайту за останні 30 днів.

На сайті існує багато додаткових функцій: нагадування щодо транспортування ЛЗ у відповідних до інструкції умовах із запропонуванням придбання термобоксу, можливість заздалегідь встановити аптеку для замовлення, залишати та переглядати відгуки інших користувачів, створювати замовлення за фотографією рецепта тощо.

Рекламні розсилки представлять собою шаблони з промокодами у вигляді посилок та акційними пропозиціями різного плану – від реклами ЛЗ до сезонних знижок на певні категорії товарів. Окремими розсилками відправляють пропозиції бренду.

1.4 Обґрунтування необхідності розробки методу

На основі розглянутих аптечних мереж визначено ряд бізнес-функцій для різних ролей системи.

З боку клієнтів взаємодія з аптечною мережею здійснюється через реалізацію базових функцій електронної комерції, таких як:

- 21.реєстрація та авторизація;
- 22.перегляд каталогів з фармацевтичною продукцією;
- 23.пошук та фільтрація необхідних товарів;
- 24.додання товару до кошика;
- 25.формування замовлення – його реєстрація у системі;
- 26.визначення місця видачі замовлення або адреси доставки.

З боку фармацевтів система забезпечує отримання та відпуск замовлень – зміна їх поточного статусу.

З боку адміністраторів система пропонує наступний функціонал:

27. додання нової продукції, категорій;

28. внесення коректив до інформації вже представленої на сайті (кількості, вартості, наявності лікарських препаратів);

29. формування акційних пропозицій за представленими замовниками переліками.

Приведений функціонал задовольняє потреби клієнта на всіх етапах покупки в аптечній мережі, але з точки зору бізнесу не заохочує в повній мірі продовжувати купувати саме в ній.

До переваг вже існуючих систем можна віднести наступне:

30. простий та інтуїтивно зрозумілий інтерфейс: якщо сайтом захоче скористатися літня людина в неї не виникне проблем;

31. фільтрація при пошуку ЛЗ: надання можливостей обирати дозування, ЛЗ, які відпускаються без рецепту, форму відпуску, кількість пігулок чи ампул в упаковці;

32. оформлення замовлення без реєстрації;

33. короткий перелік даних для реєстрації: не займає багато часу та не створює дискомфорт клієнтові через надання системі великої кількості особистої інформації;

34. наявні програми лояльності та заохочення промокодами: знижки та системи накопичення збільшують лояльність клієнта.

До недоліків можна віднести:

35. необов'язковість надання поштової скриньки при реєстрації: таким чином бізнес втрачає можливість впливу на клієнта щодо здійснення подальших покупок;

36. ігнорування сегментації клієнтів: при розсилці промоакцій ігнорується попередній аналіз замовлень клієнта, таким чином не враховуються його потреби та категорії вже здійснених замовлень;

37.система не враховує ЛЗ, що часто комбінуються в одному замовленні: втрачається можливість збільшити середній чек або надати корисні рекомендації безпосередньо під час формування замовлення.

Розглядаючи існуючі системи з точки зору зібрання та аналізу даних та розуміючи, що система представляє собою «чорний ящик», приходимо до припущення, що зібрані дані не обробляються належним чином, а результат обробки в повній мірі не використовується для утримання та розвитку клієнтів.

2 ОГЛЯД МЕТОДІВ ТА ТЕХНОЛОГІЙ, ЩО ВИКОРИСТОВУЮТЬСЯ В ПРЕДМЕТНІЙ ОБЛАСТІ

Виявлення знань з накопичених системою даних (KDD) – процес виявлення нетривіальних, нових, потенційно корисних та зрозумілих кінцевому користувачеві закономірностей – є основою аналітичної підсистеми [4]. Отримані знання надають змогу відтворювати більш цілеспрямовані пропозиції, формувати вектори розвитку, виявляти приховані взаємозв'язки.

Процес виявлення знань складається з чітко визначених етапів:

- створення вибірки;
- попередня обробка даних;
- перетворення даних;
- інтелектуальний аналіз даних;
- інтерпретація отриманих результатів.

На етапі створення робочої вибірки вирішуються питання визначення необхідних атрибутів та проміжку часу для аналізу в рамках поставленої задачі.

Етап попередньої обробки даних включає в себе виявлення шумів та аномалій, введення стратегії взаємодії з пропущеними даними. Питання заповнення пропущених даних вирішується видаленням ознак, у випадку, коли кількість пропущених даних занадто велика, або заповненням пропусків певною константою.

Перетворення даних є необхідним для представлення даних у вигляді, який вимагає метод інтелектуальної обробки даних. В залежності від наявної вибірки може включати в себе нормалізацію, стандартизацію, агрегацію та дискретизацію тощо.

Інтелектуальний аналіз даних складається з вибору декількох найбільш придатних методів, їх оцінки та застосування найкращого з них до попередньо сформованої вибірки.

Останнім етапом є інтерпретація отриманих результатів, під час котрої оцінюється корисність здобутих знань та сфери їх застосування.

Інтелектуальна обробка даних за класифікацією поділяються на методи:

- навчання з вчителем (класифікація, регресія, ранжирування);
- навчання без вчителя (кластеризація, зменшення розмірності даних);
- навчання з частковим залученням вчителя та його частий випадок – навчання з підкріпленням.

Ураховуючи специфіку поставленої задачі та предметну область, приходимо до висновку, що оптимальними до розгляду методами є ті, що не потребують залучення експертів або використання сторонніх допоміжних систем.

2.1 Методи кластеризації даних

Кластерний аналіз використовується для класифікації великої кількості даних й ставить за мету отримання відомостей щодо досліджуваної вибірки. Одержані результати можуть бути інтерпретовані для детального опису сформованих кластерів та їх ознак.

В маркетингу кластеризація використовується досить широко: для групування клієнтів за їх поведінкою, дослідження конкурентів, визначення ніші для позиціонування продукції тощо. Сегментація клієнтів – групування клієнтів в однорідні кластери – виконується для отримання максимально повного уявлення щодо їх загальної поведінки та факторів впливу. На основі цих кластерів для кожного сегменту створюється індивідуальна політика.

Методи кластеризації даних поділяють на дві великі підгрупи – ієрархічні та неієрархічні [5].

Ієрархічні методи бувають агломеративні та дивізімні методи. Агломеративні методи характеризуються послідовним об'єднанням ознак, кожна з яких на початку роботи алгоритму являє собою окремий кластер. Дивізімні методи працюють навпаки: спочатку роботи алгоритму всі ознаки

належать одному кластеру, котрий на кожній ітерації розгалужується на менші кластери. З точки зору обчислюваної ефективності краще використовувати агломеративні методи [6].

Алгоритм агломератної ієрархічної кластеризації наведено на рис. 2.1.

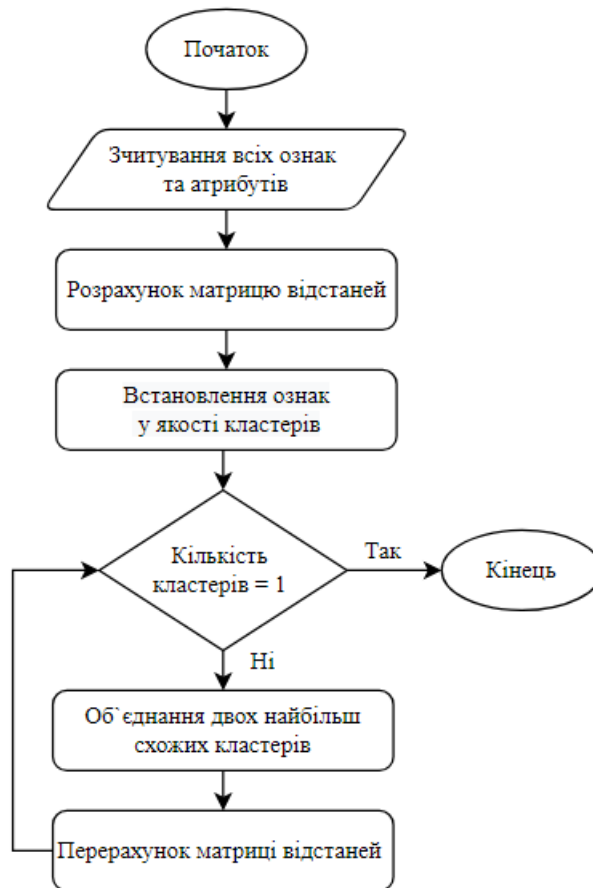


Рисунок 2.1 – Алгоритм агломеративної ієрархічної кластеризації

Методи знаходження відстані між кластерами поділяється на: одиночний (single), повний (complete), середній (average) і центроїдний (centroid) зв'язки:

– одиночний зв'язок розраховується як найменша відстань між найближчими ознаками різних кластерів за формулою (2.1). Метод добре підходить для виявлення викидів, адже вони виступають у вигляді поодиноких кластерів з однією ознакою. Недоліками постають форма кластерів у вигляді ланцюгів та скошений вигляд дендограм;

$$; \quad (2.1)$$

– метод повного зв'язку розраховується як відстань між двома найбільш віддаленими ознаками різних кластерів (2.2). Метод гарантує формування зв'язаних кластерів, приблизно однакової форми, але є дуже чутливим до викидів;

; (2.2)

– метод середнього зв'язку базується на розрахунку відстані між усіма точками двох кластерів (2.3). Добре справляється з викидами, але потребує значної кількості додаткових розрахунків;

; (2.3)

– центроїдний метод полягає в розрахунку центроїдів кластерів – точок, що є найближчими до всіх ознак в межах одного кластера, та знаходження мінімальної відстані між ними (2.4);

(2.4)

Ієрархічні методи мають один суттєвий недолік – підходять для обробки тільки невеликої вибірки – їх алгоритмічна складність становить .

Неієрархічні методи кластеризації базуються на задачі оптимізації, тобто групування вихідної вибірки у кластери є рішенням екстремальної задачі. Найпоширенішим та тривіальним методом є k-середніх.

Алгоритм методу наведений на рис. 2.2.



Рисунок 2.3 – Алгоритм методу кластеризації k-середніх

Метод k-середніх вважається швидким кластерним алгоритмом. На відміну від ієрархічних методів, котрі не вимагають додаткових відомостей окрім вибірки, потребує гіпотезу з найбільш вірогідною кількістю кластерів. Головна ідея полягає в мінімалізації відстані між елементами кластера та максимізації відстані між самими кластерами.

Метод кластеризації k-середніх для розбиття ознак використовує одну з наступних метрик [7]:

– евклідова відстань (2.5) є найбільш поширеною серед всіх метрик й являє собою геометричне відстань в багатомірному просторі;

$$; \quad (2.5)$$

– квадрат евклідової відстані (2.6) використовується для надання більшої ваги найбільш віддаленим один від одної ознакам;

$$; \quad (2.6)$$

– манхеттенська відстань (2.7) розраховується як середня різниця по координатам, призводить до таких самих результатів, як і евклідова відстань, за винятком того, що вплив вибросів зменшується – відстань не зводиться у ступінь;

$$; \quad (2.7)$$

– відстань Чебишева (2.8) використовується для виокремлення двох ознак у різні кластери, якщо вони розрізняються за якоюсь однією координатою.

$$(2.8)$$

Як висновок, неієрархічний кластерний метод k-середніх підходить для обробки великих вибірок – алгоритмічна складність становить $O(n^2k)$, де n – це кількість ознак у вибірці, k – кількість кластерів, t – кількість ітерацій алгоритму. До недоліків методу відноситься необхідність задавати початкову кількість кластерів та чутливість алгоритму до вибору початкових наближень центроїдів – в залежності від значення центроїду змінюється навіть на однаковій вибірці. Останній недолік виправляється у модифікованому алгоритмі k-середніх – k-середніх++.

2.2 Рекомендаційні системи та їх різновиди

Обсяг збираних даних, прагнення економити час користувачів та збільшити їх задоволення від взаємодії з сервісом породжує попит на нові і ефективні рекомендаційні системи.

Основна мета розвитку рекомендаційних систем полягає в зменшенні інформаційного навантаження та витрат на обробку даних. Аналіз інтересів та поведінки користувача, задля визначення його вподобань, є вигідним як для самих користувачів, так і для постачальників послуг. Користувачі цінують свій вільний час й бажають витратити його з користю – рекомендаційні системи самі виконують підбір можливих цікавих товарів чи послуг та пропонує їх. Організації використовують рекомендаційні системи для максимізації свого прибутку і мінімізації можливого ризику. Одними з найпопулярніших рекомендаційних систем сучасності є система рекомендації фільмів Netflix та системи персоналізованої вибірки новин від Google.

За класифікацією рекомендаційні системи розрізняють: на основі колабораційної фільтрації, на основі контенту, на основі моделі знань та гібридні (рис. 2.3).



Р

исунок 2.3 – Класифікація рекомендаційних систем

Розглянемо рекомендаційні бази на основі фільтрації та контенту [8, 9].

Рекомендаційні системи на основі колабораційної фільтрації виносять пропозиції користувачеві на базі того, що обирали користувачі зі схожими інтересами в минулому.

Рекомендаційні системи використовують матрицю вподобань, стовпці котрої представляють користувачів, а рядки – товарів. Елемент матриці містять в собі рейтинг товару наданий користувачем.

Рейтинги поділяються з типом на:

- числові – користувач використовує певну шкалу для визначення вподобання, наприклад від 1 до 5;
- впорядковані – користувач залишає свій відгук у вигляді словесного формату, наприклад «повністю згоден», «згоден частково», «дотримуюсь нейтралітету», тощо;
- двійкові – містять інформацію у вигляді 0 та 1, наприклад, чи подобається цей товар користувачеві або чи був товар доданий до замовлення. Даний рейтинг часто використовується для неявного збору інформації, тобто клієнт не заповнює ніякі анкети чи відгуки, а сама система фіксує його інтерес на основі взаємодій;
- абсолютні – містять інформацію про кількість взаємодії з певним продуктом, наприклад, відзначають кількість прослуховувань аудіозапису чи переглядів статті.

Рекомендаційні системи даного типу поділяються на: орієнтовані на клієнта (user-based) та орієнтовані на товар (item-based) системи. В орієнтованій на клієнта системі результатом буде перелік товарів, котрі можуть сподобатися користувачеві найбільше. В орієнтованій на товар системі результатом буде прогнозування рейтингу.

Процес визначення схожості між клієнтами системи відбувається за рахунок визначення міри кореляції між ними. Для визначення міри використовують наступні методи: коефіцієнт кореляції Пірсона, коефіцієнт кореляції Спірмена, евклідову відстань, манхетенську відстань, косинусову міру та коефіцієнт кореляції Танімото.

Коефіцієнт кореляції Пірсона визначається як лінійна кореляція між двома векторами рейтингів (2.9) й приймає значення в діапазоні $[-1, 1]$. Можна використовувати тільки при нормальному розподілі.

$$\bar{r}_{ij} = \frac{r_{ij} + r_{ji}}{2} \quad (2.9)$$

де $\bar{r}_{ij}$ – середнє арифметичне векторів та n_{ij} – кількість товарів, що співпали у векторах оцінок.

Косинусова міра (2.10) розраховується як косинус кута між векторами рейтингів користувачів – користувачі вважають схожими, якщо кут між їх векторами малий. Цю міру доречно використовувати при рейтингу у вигляді симетричної шкали, наприклад, від -5 до 5.

$$\cos(\theta_{ij}) = \frac{\sum_{k=1}^n r_{ik} r_{jk}}{\sqrt{\sum_{k=1}^n r_{ik}^2 \sum_{k=1}^n r_{jk}^2}} \quad (2.10)$$

Коефіцієнт кореляції Тонімото (2.11) використовується у випадку роботи з двійковим рейтингом та розраховується як відношення кількості товарів, які оцінили обидва користувача n_{ij} , до загальної кількості товарів n_i оцінених хоча б одним з них.

$$t_{ij} = \frac{n_{ij}}{\min(n_i, n_j)} \quad (2.11)$$

Перед тим, як почати прогнозування рейтингу товару всі попередньо відібрані користувачі зі спільними інтересами сортуються за ступенем кореляції та відбирається константна кількість для подальшого розгляду.

Прогнозування рейтингу виконується за формулою (2.12). Рекомендується для кожного користувача з суміжними інтересами помножити проставлений рейтинг на міру його схожості – у такому випадку міра схожості виступає ваговим коефіцієнтом.

$$\hat{r}_{ij} = \frac{\sum_{k=1}^n r_{ik} r_{jk}}{\sum_{k=1}^n r_{ik}^2} \quad (2.12)$$

Для створення персональної рекомендації слід розрахувати рейтинги для усіх неоцінених товарів та рекомендувати ті, що з найвищим рейтингом.

Перевагами рекомендаційних систем на основі колабораційної фільтрації є:

- теоретично висока точність;
- відсутність наявності профілю користувача та атрибутів товару;
- легкість додавання нових даних;
- проста реалізація.

До недоліків відносяться:

- проблема холодного старту: нові користувачі не можуть отримувати рекомендації, адже система нічого про них не знає; так само й з новими товарами – їх ще не купували, тому їх не буде рекомендовано;
- оцінювані товари мають бути досить близькими за тематикою, інакше отримані оцінки не матимуть жодного відношення до реального товару;
- необхідно мати дані щодо вподобань користувачів;
- продуктивність роботи алгоритму падає у випадку існування розрідженої матриці вподобань;
- залежність точності прогнозування рейтингу від кількості накопичених рейтингів.

Рекомендаційні системи на основі контенту аналізують ознаки товарів системи, щоб визначити ті, які представляють особливий інтерес для користувача.

В залежності від предметної області, система може збирати інформацію щодо користувача у явному та неявному вигляді. В першому випадку користувач сам надає системі інформацію, щодо свого соціально-демографічного становища, особистих характеристик, інтересів тощо. Наприклад, під час реєстрації у книжковому магазині пропонується обрати найбільш цікаві жанри. В другому випадку інтереси визначаються під час взаємодії з системою – фіксуються пошукові запити, переглянуті товари, вміст товарів у замовленні, відгуки.

Товари системи містять ознаки у вигляді опису, призначення, цінової категорії, бренду та інших характеристик. Для надання рекомендацій

використовуються вибірка з атрибутів кожного товару з бази даних або додаткові тегові позначення. Ці ознаки для рекомендаційної системи представляються у вигляді вектора ключових слів у багатовимірному просторі. За допомогою цих векторів система надає оцінку, що визначає ступінь потенційної зацікавленості користувача у товарі.

В рекомендаційних системах даного типу використовуються наступні технології [10, 11]:

KNN метод класифікації. Розраховує зацікавленість користувача до товару, шляхом порівняння усіх відомих користувачеві товарів до шуканого. Для роботи методу необхідно визначати кількість ксусідів та метрики для розрахунку відстані між товарами (об'єктами) – евклідова (2.5), манхеттенська (2.7) та косинусної подібності (2.13). Визначена кількість найближчих сусідів й буде рекомендуватися;

(2.13)

TF-IDF метод, що є провідним у текстовому дослідженні даних. Доречно застосовувати у випадках, коли товари системи містять багато текстового опису. Наприклад, сюжет фільму, вподобані рецензії, статті, анотація до книги, тощо. Основна логіка базується на положенні, що не всі елементи однаково важливі: союзні слова мають мале інформативне навантаження для дослідження. Тому при визначенні числа співпадаючих елементів у двох векторах все виміри слід попередньо зважити за їх значимістю. Так, TF за формулою (2.14) характеризує співвідношення кількості конкретного слова до загального набору слів, й чим вище його значення, тим вагомніше слово. IDF за формулою (2.15) характеризує інверсію частотності, з якою конкретне слово використовується в тексті. Саме ця метрика знижує вагомість часто вживаних слів.

, (2.14)

, (2.15)

де – загальна кількість текстових документів, – кількість документів з загального переліку, котра містить у собі слова, $()$.

Також даний метод може використовуватися й з звичайними атрибутами: TF характеризує міру значимості атрибута для користувача, а IDF – міра рідкості атрибута.

Переваги рекомендаційної системи на основі контенту :

- для роботи не потребує даних багатьох користувачів;
- можливість створювати рекомендації для користувачів з унікальним набором інтересів;
- відсутня проблема холодного старту для нових товарів, адже використовуючи їх атрибути легко знайти схожі вже серед існуючих;
- результати роботи системи легко інтерпретувати.

До недоліків відноситься наступне:

- надання рекомендацій новому користувачеві, який не виконав достатньої кількості взаємодії з системою, не може бути точним – проблема холодного старту для користувачів;
- схожі групи товарів призводять до того, що рекомендуються одні й ті самі позиції – постійна рекомендація підмножини;
- обмежена інформація щодо товару призводить до складності в їх ідентифікації та угруповання, наслідком означеного є неточні рекомендації.

2.3 Виявлення закономірностей на основі асоціативних правил

Асоціативні правила є передовим способом для визначення залежностей у ринковому кошику. Показовим прикладом є розміщення продуктів у магазинах – так, покупець не тільки економить власний час, а й отримує нагадування щодо релевантних товарів.

Асоціативні правила не виділяють особисті вподобання клієнта, як робиться в описаних вище методах, а знаходить зв'язки між переліком товарів у

рамках однієї покупки – транзакції [12, 13, 14, 15]. Кожне правило складається з попереднього елемента і наступного, обидва з котрих можуть представляти собою деяку множину. Зазначено, що мається на увазі спільне виникнення, а не причинний зв'язок між елементами.

Отримані правила поділяються на три категорії: корисні, тривіальні та незрозумілі. Корисні правила містять тільки якісну інформацію, котру не можна визначити припущенням. Тривіальні правила вже є відомими для бізнесу, тому не приносять жодних відкриттів. Незрозумілі правила не мають ніякого сенсу.

Для визначення тільки корисних залежностей між товарами системи використовують наступний перелік показників:

– підтримка – визначає як часто зустрічається множина товарів відносно усіх транзакцій (2.16):

$$\text{support}(X) = \frac{\text{count}(X)}{n} \quad (2.16)$$

де визначає кількість транзакцій, де два товари з'явилися разом.

– достовірність – визначає як часто правило працює відносно появи кожного його товару в окремому порядку або підмножині (2.17). Іншими словами умовна ймовірність придбання наступного товару з огляду на попередній придбаний;

$$\text{confidence}(X \rightarrow Y) = \frac{\text{support}(X \cup Y)}{\text{support}(X)} \quad (2.17)$$

– підйом – визначає у якій ступені товари залежать один від одного – відношення підтримки відносно підтримки кожного з них в окремому порядку (2.18). Чим більше отриманий показник за 1, тим сильніше розглянуте правило, в протилежному випадку правило є слабким і його не слід включати до результуючого переліку.

$$\text{lift}(X \rightarrow Y) = \frac{\text{confidence}(X \rightarrow Y)}{\text{support}(Y)} \quad (2.18)$$

Одним з найуживаніших методів генерації асоціативних правил є Apriori – алгоритм наведено на рисунку 2.4.

Використання алгоритму потребує попереднього визначення експертами значення мінімальної підтримки, мінімальної достовірності та підйому.

Алгоритм поділяється на чотири етапи:

- формалізація даних – кожна транзакція переформовується у двійкову послідовність й містить усі товари системи. Якщо елемент міститься у конкретній транзакції, то його значення в переліку дорівнює 1, в протилежному випадку 0;

- пошук одноелементних наборів, з підтримкою вище за мінімальну задану;

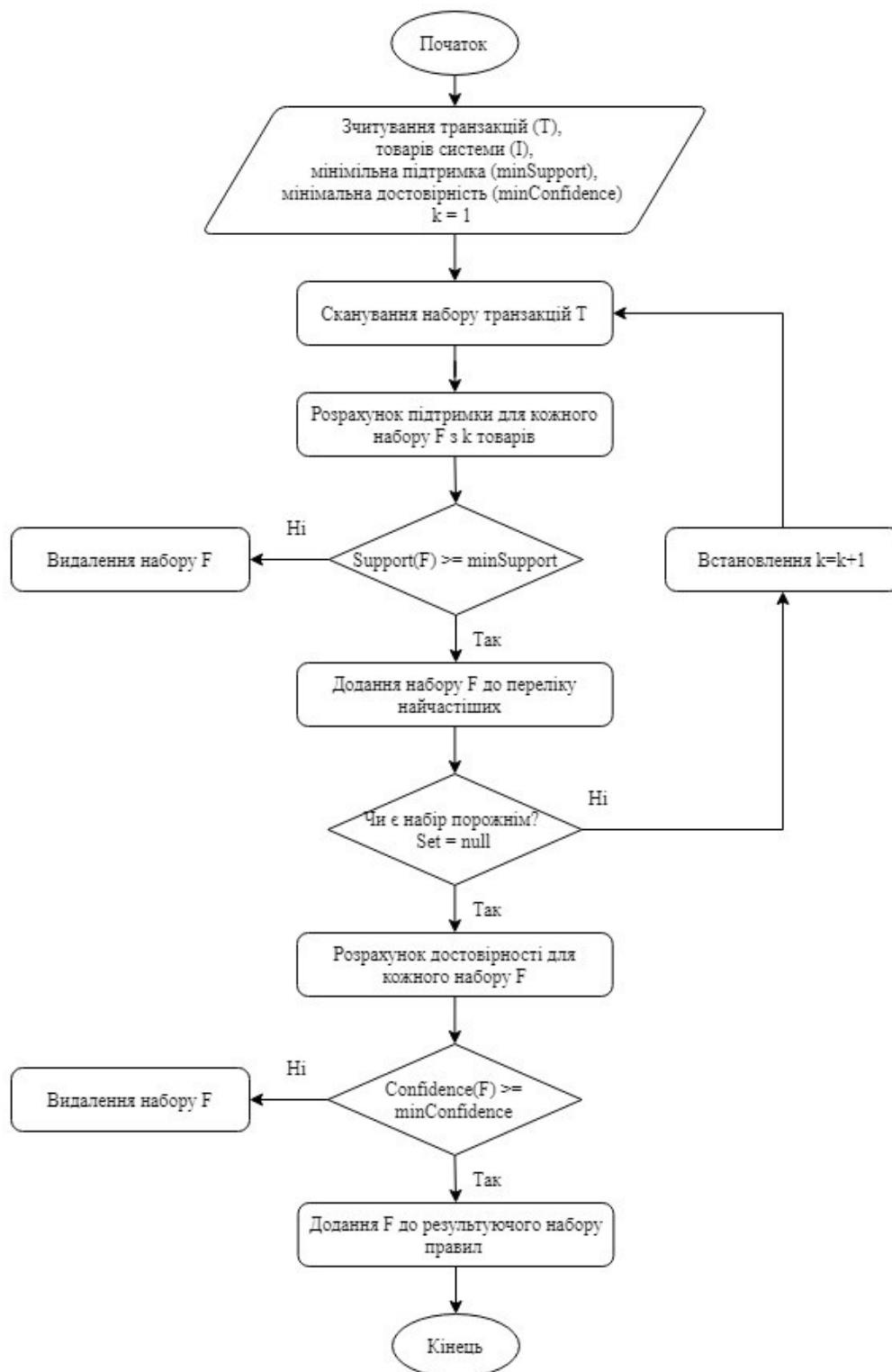


Рисунок 2.4 – Алгоритм Apriori

– пошук наборів, що часто трапляються серед транзакцій. На цьому етапі спрацьовує одна з основних властивостей алгоритму – антимонотоність. За даною властивістю набір може трапитися тільки тоді, коли часто зустрічаються усі товари;

– формування переліку правил – до кінцевого набору правил додаються тільки ті, чиї показники перевищують порогові значення достовірності й підйому.

До переваг наведеного методу генерації асоціативних правил Apriori відноситься його проста реалізація, спроможність опрацьовувати великі набори даних та придатність до розпаралелювання. Недоліками виступає кількість необхідних сканувань.

3 ПОСТАНОВКА ЗАДАЧІ

3.1 Призначення розроблюваного методу

Покращення взаємодії з клієнтами аптечної мережі є необхідним для створення компонентів CRM системи, особливої уваги в котрій слід приділити дослідженню використовуваних аналітичних методів та їх модифікації.

3.2 Обмеження предметної області з боку законодавства

Основним обмеженням для ведення оптової та роздрібною торгівлі ЛЗ є 6 розділ «Реалізація лікарських засобів» закону України «Про лікарські засоби» [16] та «Перелік лікарських засобів, заборонених для рекламування» за наказом МОЗ України [17].

Таким чином, реклама ЛЗ, які не можна придбати без рецепту не є корисною з логічної точки зору – клієнт не може просто так купити дану продукцію. З іншого боку, дані ЛЗ та ті, що заборонені для рекламування, мають подаватися на вхід розроблюваного методу, бо є основою для визначення частих сумісних пар товарів у кошику клієнта для надання односторонньої рекомендації.

3.3 Постановка задачі дослідження та її формалізація

Специфікою системи електронної комерції «Аптечної мережі» є наявність препаратів (або фармацевтичної продукції не лікувального призначення), що є супутніми до інших або часто купуються сумісно. Наприклад, при придбанні лікарських препаратів з антибактеріальною дією (антибіотики), раціонально нагадати щодо придбання пробіотичного лікарського препарату з метою запобігання негативного впливу на кишкову мікрофлору (мікробіом) людини.

Або, після подолання гострої респіраторної інфекції (ГРІ) щодо позачергової зміни зубної щітки, поряд з рекомендацією щодо її придбання разом з антисептиками для ротової порожнини.

В системі електронної комерції клієнти формують замовлення з товарів (препаратів), представлених на сторінках браузера. Позначимо ці товари через множину (3.1) з зазначеною потужністю та підмножиною – товарів для реклами :

(3.1)

Клієнтів системи представимо множиною (3.2) з потужністю та підмножиною (3.3) – клієнти, що вже робили замовлення у системі. Припускається, що користувач міг зареєструватися у системі, але на момент запуску методу для генерації рекомендацій та особистих пропозицій, ще не встиг сформувати замовлення:

(3.2)

(3.3)

В рамках одного замовлення клієнт може придбати декілька товарів (препаратів). Здійснення одного замовлення еквівалентно здійсненню однієї транзакції, яку позначимо через T (3.4):

(3.4)

На основі наявної інформації необхідно згенерувати рекомендації щодо конкретних найменувань товарів і в подальшому для формування персональних пропозицій. Дану задачу вирішено розв’язувати у два етапи. На першому з них визначаються залежності між товарами системи – генеруються асоціативні правила, на другому, з використанням результатів першого етапу, створюються персональні пропозиції.

Вхідними даними для першого етапу виступають:

38.D множина з транзакцій (3.5) :

$$; \quad (3.5)$$

39.множина товарів (3.1);

40.визначені значення мінімальної підтримки кожної транзакції та мінімальної достовірності. Ці значення попередньо розраховуються експертом та використовуються для відсіювання некорисних та тривіальних залежностей між товарами.

Введемо множину часто вживаних разом товарів F (3.6) та множину , що визначає всі транзакції, в котрих міститься F (3.7):

$$, \quad (3.6)$$

$$(3.7)$$

Задля виявлення корисних залежностей для кожної множини часто вживаних разом об'єктів F розраховується значення підтримки та достовірності на основі кількості транзакцій, де ця множина з'являється.

Фактичне значення підтримки розраховується за формулою (3.8) як співвідношення кількості транзакцій, в котрі входить множина , до загальної кількості транзакцій:

$$(3.8)$$

Фактичне значення достовірності визначається за формулою (3.9) як вірогідність наявності в транзакції множини, якщо вже наявна множина :

(3.9)

Результатом першого етапу методу є множина L , що містить у собі всі часті набори об'єктів (3.10). За термінологією методу генерації асоціативних правил кожен підмножина множина L вважається правилом:

(3.10)

Вхідними даними для першого етапу виступають:

- 41.множина визначених правил (3.10);
- 42.множина рекламних пропозицій (3.1);
- 43.множина користувачів (3.3).

Результатом роботи другого етапу є множина найбільш придатних до рекомендації товарів певному користувачеві (3.11). Алгоритм формування цієї множини слід розробити.

(3.11)

3.4 Визначення основних етапів

Розробка методу та його подальша інтеграція потребує попередніх досліджень таких як:

- 44.виявлення найбільш підхожих методів для реалізації заохочення клієнтів аптечної мережі;
- 45.формування методу з врахуваннями законодавчих обмежень та вимог системи;
- 46.визначення основних моментів інтегрування розробленого методу в систему аптечної мережі та технологій для його реалізації.

4 РОЗРОБКА МЕТОДУ

4.1 Аргументація обраних методів

За результатами аналізу предметної області було виявлено низький рівень підтримки клієнта система під час формування замовлення та не таргетованість рекламних розсилок. Ставиться за мету розробити метод для вирішення даних задач.

Розглянемо кожен з наведених у попередньому розділі методів для предметної галузі аптечної мережі.

Кластеризацію доречно використовувати для виокремлення схожих групи ознак. В цьому випадку – клієнтів системи. Важливим постає питання атрибутів клієнтів, за якими виконується кластеризація. Часто орієнтуються на вікову приналежність клієнтів та зміст їх замовлень за певний період – таким чином, дослідження даних кластерів дає можливість формувати розсилки для певних груп з суміжними інтересами. У випадку аптечних закладів даний приклад сегментації є некоректним за декількох причин:

- неоднозначність вікового критерію. Як видно з дослідження предметної області, мало систем просять від користувачів обов'язкового надання цієї інформації, адже важко виокремити ЛЗ для певної вікової категорії, а ліки для дитячого контингенту та похилих людей, зазвичай, купують клієнти середнього віку;

- перетин кластерів. В роздрібній торгівлі ЛЗ не є доцільним вважати, що люди будуть купувати продукцію виняткового однієї або декількох категорій. Наприклад, у зоомагазинах доречно рекомендувати до придбання іграшки для малих тварин, якщо попередньо здійснювались покупки у цій категорії.

З точки зору надання рекомендацій під час формування замовлення, даний метод буде видавати виключно тривіальні пропозиції з тієї самої категорії, що не дуже підходить в даному випадку.

Використання рекомендаційних систем на основі колабораційного фільтрування дозволяє частково персоніфікувати клієнта системи – обираються схожі клієнти та виокремлюються для рекомендацій найбільш релевантні за рейтингом пропозиції.

Для аптечних мереж даний підхід працюватиме досить неоднозначно, адже тематика пропонованої продукції не є близькою, а наявність сильно розрідженої матриці вподобань гарантовано. Подібні рекомендаційні системи доречно використовувати в вузькопрофільних системах.

Рекомендаційні системи на основі контенту в даному випадку можна використовувати, але результатом будуть схожі ЛЗ за фармакологічною дією, активним фармацевтичним інгредієнтом (діючим компонентом) та призначенням.

Асоціативні правила найкраще всього підходять для даної предметної області:

- отримані комбінації продукції можна використовувати для рекомендацій під час формування замовлення;
- пропозиції не будуть тривіальними, тому є шанс дійсно зацікавити користувача;
- зміни в асоціативних правилах не відбуваються часто, тобто можна їх генерувати не кожного разу, а раз на певний період.

Розглянуті методи можуть бути використані у комбінації: так, пропонується згенерувати асоційовані правила та використати їх для виокремлення потенційно найбільш цікавих для клієнтів рекламних пропозицій, використовуючи матрицю вподобань.

4.2 Розроблений метод

Розроблений метод є комбінацією методу генерації асоціативних правил на прикладі методу Apriori, його показника підтримки та матриці переваг. Сам алгоритм складається з двох етапів – генерація асоціативних правил та формування особистих рекомендацій на їх основі.

До першого етапу на вхід подаються наступні дані:

- множина замовлень (транзакцій) (3.5);
- множина усіх товарів системи (3.1);
- параметр мінімальної підтримки;
- параметр мінімальної вірогідності.

Алгоритм Apriori складається з наступних восьми кроків, алгоритм наведено на рисунку 2.4.

На першому кроці для всіх множин (3.6) з товарами розраховується підтримка у транзакціях (3.8).

На другому кроці видаляються всі множини, чия підтримка нижче за мінімальну вказану.

На третьому кроці формуються нові множини товарів з товарів.

На четвертому кроці розраховується підтримка для кожного отриманої множини.

На п'ятому кроці перевіряється чи більша підтримка за мінімальну вказану, якщо так, то множина товарів додається до переліку множин, що часто зустрічаються. Крок повторюється для всіх множин, з мінімальною підтримкою;

На шостому кроці перевіряється наявність зміни кількості частих множин за ітерацію: якщо перелік на попередньому кроці не поповнився новими наборами, то переходимо до кроку 7. В протилежному випадку – повертаємося до 3 кроку;

На сьомому кроці для всіх отриманих множин розраховується вірогідність;

На восьмому кроці перевіряється чи більша вірогідність за мінімальну вказану : якщо так, то множина додається до множини правил алгоритму (3.10). В протилежному випадку – видаляється.

На вхід до другого етапу подаються наступні дані:

- отримані асоціативні правила (3.10);
- значення підтримки для кожного правила , визначається на попередньому етапі разом з асоціативними правилами;
- множина замовлень системи D (3.5) – з врахування ідентифікатора клієнта (3.3);
- перелік товарів для реклами ;

Другий етап містить наступні 4 кроки й зображений на рисунку 4.1:

На першому кроці формується матриця вподобань (4.1): по вертикалі розташовуються ідентифікатори товарів системи, по горизонталі – ідентифікатори клієнтів системи. Елемент матриці може приймати значення 0 або 1, та означає наявність певного товару i в кошику конкретного користувача .

(4.1)

На другому кроці для кожного товару з товарів для рекламування обирається перелік правил , в котрих вони присутні (4.2):

(4.2)

На третьому кроці для кожного клієнта з множини створюється окрема множина персоніфікованих товарів (4.3) та їх підтримки : якщо, клієнт вже купував пропонований товар раніше, то він додається до множини з підтримкою 1, якщо ні – перевіряється, чи додавалися товари з множини правил даного товару – якщо так, то вони додаються до персоніфікованої множини пропозицій з підтримкою правила.

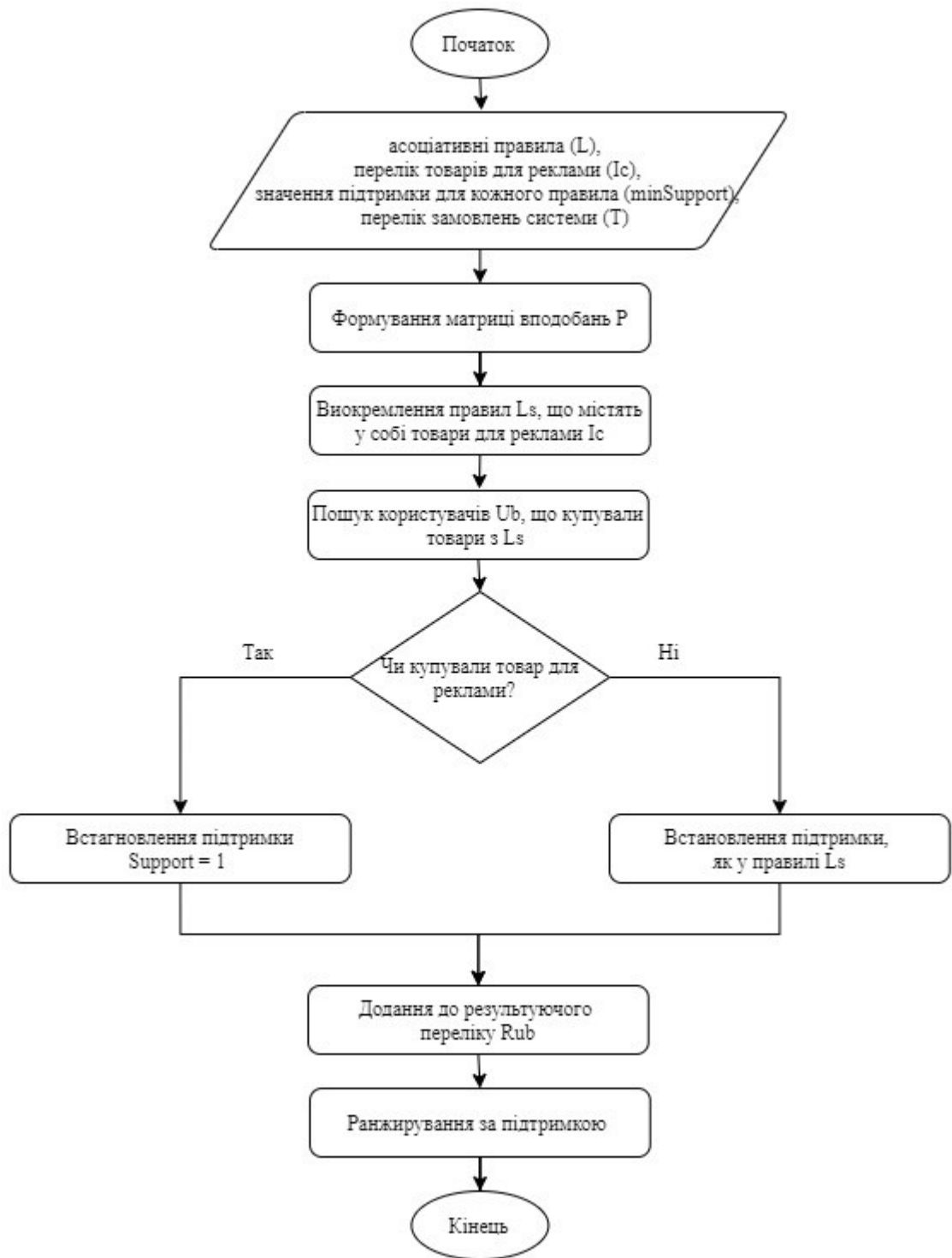


Рисунок 4.1 – Алгоритм другого етапу розробленого методу

(4.3)

На четвертому кроці отримані множини ранжуються за підтримкою правила t й обирається константна кількість товарів для рекомендації у персональній рекламі.

4.3 Варіанти оптимізації пропонованого алгоритму

Розрахунок асоціативних правил можна оптимізувати шляхом введення булевого параметра для зазначення необхідності регенерації асоціативних правил. Таким чином, адміністратор системи матиме змогу швидко розіслати рекламні пропозиції.

Окремо слід розглянути доречність введення оптимізацій самого алгоритму асоціативних правил, таких як хеш-таблиці та часткового розрахунок (partning).

Іншим шляхом оптимізації є опрацювання замовлень лише за конкретний, чітко визначений період. Немає сенсу розглядати набір товарів у замовленні, якщо воно було створено більше року тому. Так само, слід виключити з розсилки користувачів, що останній час взаємодіяли з системою досить давно.

5 ПРОГРАМНА РЕАЛІЗАЦІЯ ТА ІНТЕГРАЦІЯ МЕТОДУ

5.1 Вибір технологій для проектування та програмування

Розроблений метод має бути інтегровано у якості аналітичного компоненту CRM системи до інформаційної системи аптечної мережі.

Систему електронної комерції аптечної мережі створено на основі Spring MVC – фреймворку, придатного для створення гнучких та слабо пов'язаних веб-додатків. Шаблон проектування MVC допомагає розділити бізнес-логіку (Model), логіку уявлень (View) та логіку контролера (Controller). Моделі відповідають за інкапсуляцію даних програми. Views відображають відповідь користувачеві за допомогою об'єкта моделі. Контролери відповідають за отримання запиту від користувача та виклик функціоналу бізнес-логіки.

Принцип роботи фреймворку Spring MVC наведено на рисунку 5.1.

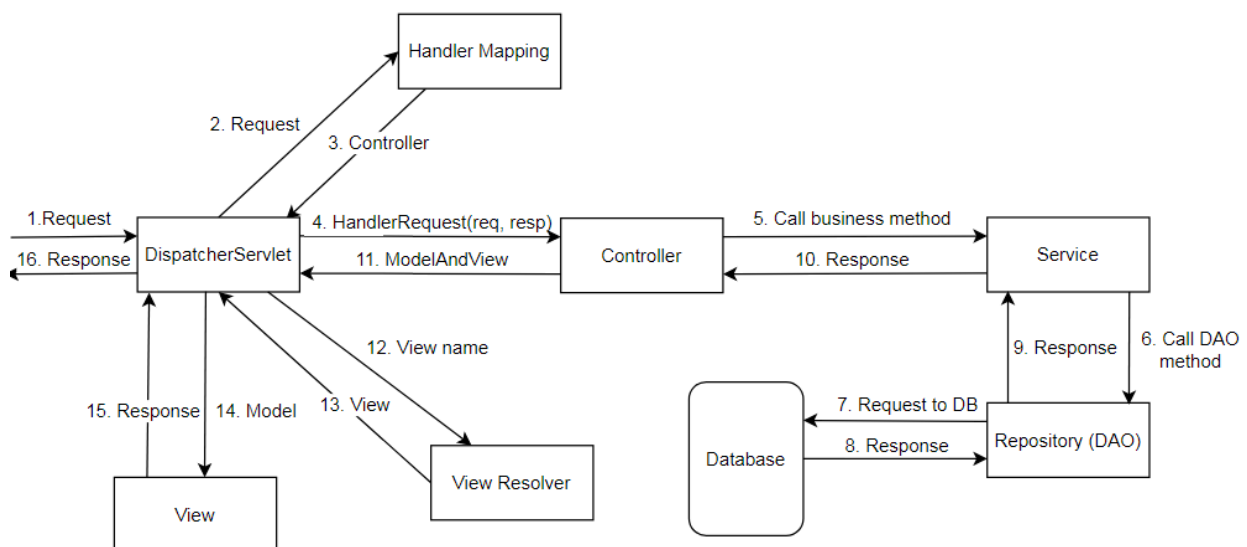


Рисунок 5.1 – Принцип роботи фреймворку Spring MVC

Запит від клієнта (браузера) потрапляє до DispatcherServlet-a – головного контролеру Spring, що визначає за допомогою Handler Mapping кінцевий контролер для обробки отриманого URL. На визначеному контролері викликається метод `handleRequest(req,resp)`, що містить у собі методи сервісів відповідно до бізнес-логіки застосунку. Як зазначено на рисунку, кожен метод звертається до бази даних через репозиторії для отримання кінцевого набору даних. Далі набір даних повторно обробляється у сервісі та повертається у готовому для відправки вигляді до контролера. Контролер повертає до DispatcherServlet об'єкт моделі та назву необхідного для повернення клієнтові уявлення. DispatcherServlet отримує уявлення за допомогою View Resolver-a та відправляє запит на формування результату до View, що обробить модель та поверне HTML сторінку. На останньому кроці DispatcherServlet відправляє результат до браузера [18, 19].

У якості уявлень використовуються JSP та JSLT [20].

JSP технологія дозволяє поєднувати статичний HTML з кодом Java та робити веб-сторінку динамічною. Основні переваги використання JSP:

- 47. створення інтерактивних сторінок;
- 48. забезпечення простого зчитування внесеної користувачем інформації;
- 49. введення особистих тегів;
- 50. легкість використання.

JSLT представляє собою колекцію корисних тегів, що спрощують вставку Java коду до JSP. Наприклад, по елементу ітерація списків або умовні оператори.

Взаємодію з базою даних реалізовано за допомогою реалізаційної СУБД MySQL [21] та фреймворку Hibernate [22].

MySQL є однією з найпопулярніших СУБД з відкритим кодом – дозволяє налаштовувати вихідний код відповідно до власних потреб. Іншими її перевагами у порівнянні з іншими СУБД є:

51. незалежність від платформи – працює як з операційною системою Windows, так і з Linux.

52. наявність драйверів для взаємодії з застосунками написаними на популярних мовах програмування – C++, Java, Python;

53. вбудовані засоби для покращення продуктивності застосунків: використання уявлень, тригерів, збережених процедур, можливість відкатувати транзакції у випадку помилок;

54. наявність візуального інструменту – MySQL Workbench –, що забезпечує інтерфейс для моделювання даних, запуску SQL запитів, представляє розробникам комплексні інструменти адміністрування для конфігурації сервера, резервного копіювання тощо.

Hibernate – фреймворк для створення об'єктно-реляційного відображення, що дозволяє не залучати розробника у процес перетворення класів Java у таблиці БД, а типів даних Java у типи даних SQL. Окрім цього фреймворк надає наступний ряд переваг:

55. має власну мову запитів HQL – на відміну від SQL працює з об'єктами, а не з таблицями;

56. забезпечує незалежність від обраної СУБД – досягається шляхом зміни параметра діалекту бази даних у конфігураційному файлі, таким чином при переході на іншу СУБД відсутня необхідність змінювання коду;

57. працює з найпоширенішими СУБД: MySQL, PostgreSQL та Oracle;

58. наявність механізму кешування, що складається з двох рівнів. Перший рівень прив'язаний до об'єкту сесії, а другий – до об'єкту-фабрики сесії. За замовченням ввімкнено лише перший рівень. Кешування зберігає об'єкти в пам'яті, що виключає повторні запити до бази даних для одного і того ж об'єкта. Завдяки цьому кількість операцій з базою даних зводиться до мінімуму.

Коректна інтеграції методу для покращення взаємодії з клієнтами повинна бути реалізована з урахуванням існуючих вимог та розробок.

Необхідно розглянути та розширити:

- 59.функціональні вимоги до інформаційної системи – відтворення функціональної моделі та діаграми потоків даних DFD [23];
- 60.діаграму варіантів використання;
- 61.провести логічне та фізичне моделювання.

5.2 Уточнення функціональних вимог

За нотацією IDEF0 функціональна модель представляє взаємопов'язані функціональні блоки. Кожна кінцева функція системи представлена функціональним блоком на відповідному рівні декомпозиції. Стрілки пов'язують функціональні блоки та набувають різного смислового значення.

Стрілки визначають:

- 62.вхідну інформацію – входять до функціонального блоку зліва, містять об'єкти або інформацію необхідну для роботи;
- 63.вихідну інформацію – виходять з правої частини функціонального блоку та означають результат його роботи;
- 64.механізми – об'єкти, що керують або задіяні у роботі функціонального блоку;
- 65.управління – всі документи, закони та правила, що регламентують обмеження для виконання функціонального блоку.

На рисунках 5.2 – 5.3 представлені контекстна діаграма та діаграма її декомпозиції відповідно, на яких відображено основну функцію інформаційної системи – «Продаж препаратів», та всю інформацію для її коректної роботи.

Вводимо додаткові дані для забезпечення роботи аналітичного компоненту CRM.

На вхід до функціонального блоку «Продаж препаратів» додатково подається наступна інформація:

- 66.перелік лікарських препаратів до рекламування;
- 67.визначені коефіцієнти для роботи алгоритму.

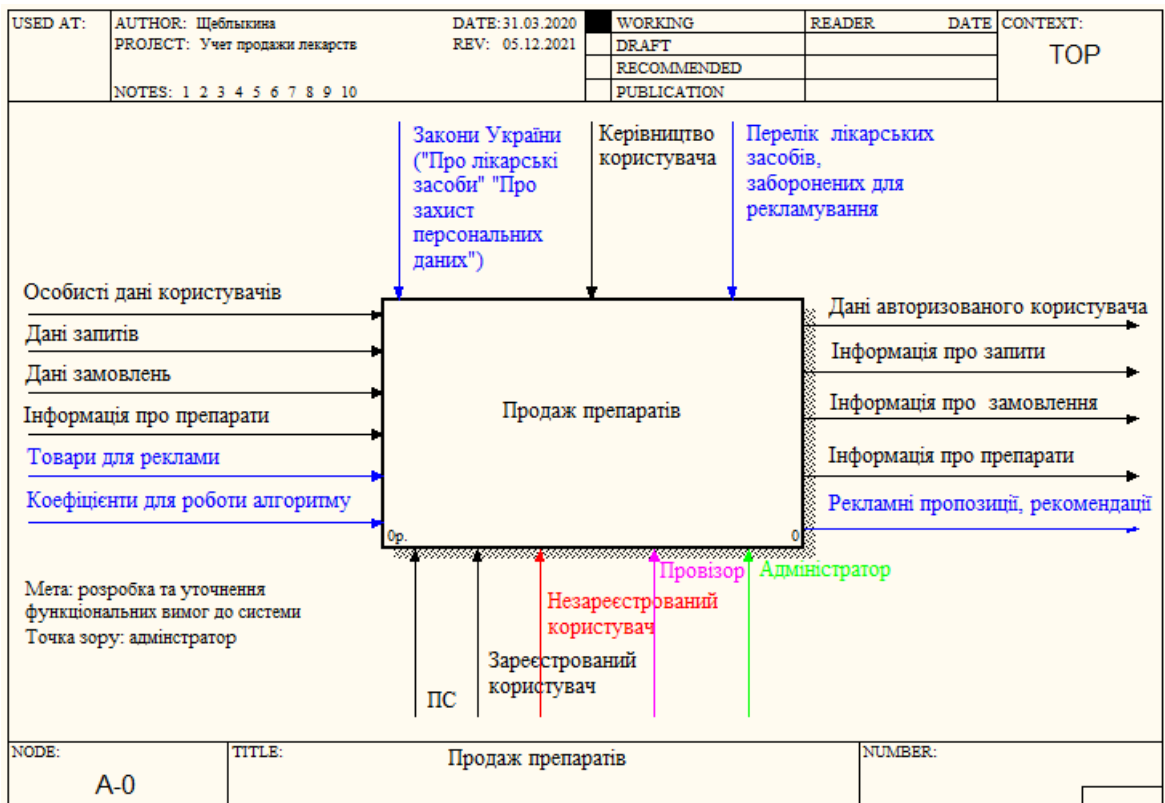


Рисунок 5.2 – Контекстна діаграма системи «Продажу препаратів»

У якості елементів управління додаються:

68. закони України, що визначають ЛЗ дозволені для продажу через Інтернет ресурси;

69. накази МОЗ України, щодо заборонених для рекламування ЛЗ.

Для розуміння взаємодії блоків системи проводимо декомпозицію функціонально блоку «Адміністрування» (рис. 5.4). Визначено наступні три підфункції:

70. відпуск замовлення;

71. оновлення ЛЗ;

72. створення рекомендацій та рекламних пропозицій.

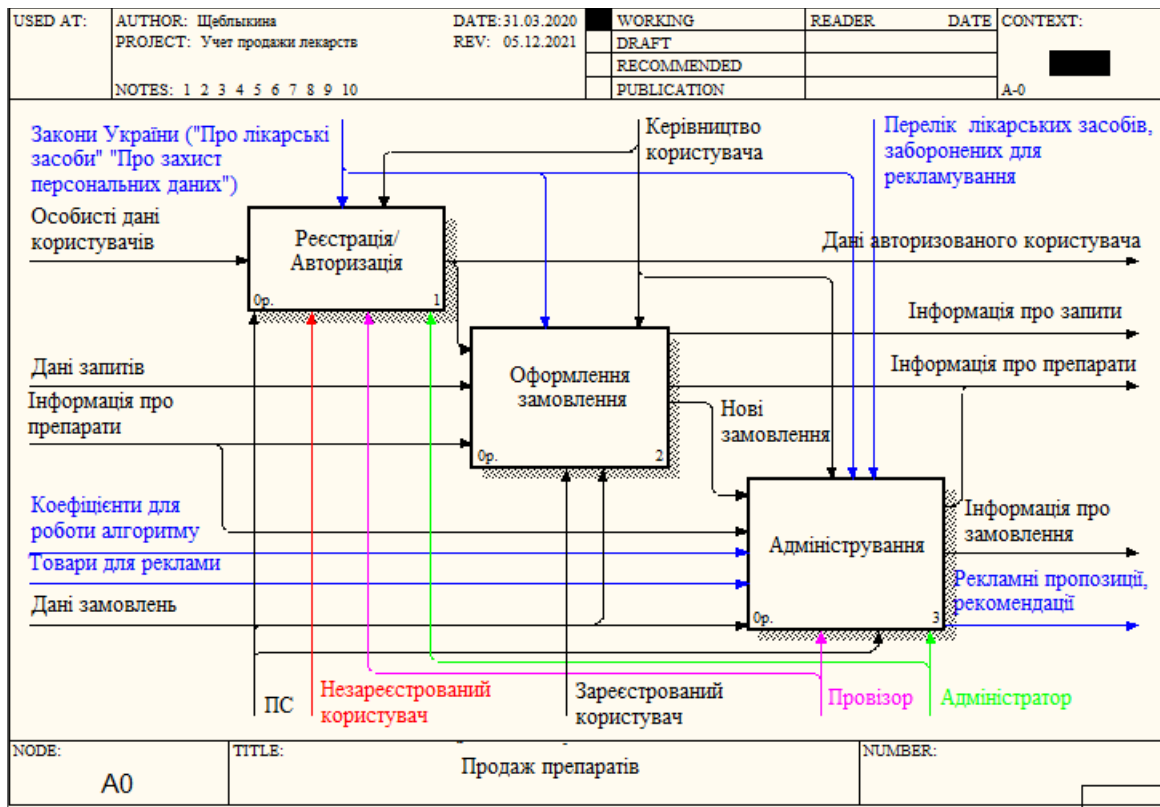


Рисунок 5.3 – Декомпозиція контекстної діаграми системи «Продажу препаратів»

Функціональний блок «Відпуск замовлення» виконується механізмом «Фармацевт» та реалізується при безпосередній передачі препаратів з замовлення відвідувачу в аптечному закладі. Результатом є збережене та виконане замовлення конкретного «Зареєстрованого користувача» системи. У подальшому використовується для генерації вподобань.

Функціональний блок «Оновлення лікарських засобів» виконується механізмом «Адміністратор» та реалізує зміну інформаційного опису ЛЗ або вносить нові лікарські препарати або медичну продукцію нелікувального призначення до каталогів системи.

Функціональний блок «Створення рекомендацій та рекомендаційних пропозицій» виконується механізмом «Адміністратор», що запускає алгоритм для створення або оновлення існуючих рекомендацій та формування даних для персоналізованих пропозицій у рекламній розсилці.

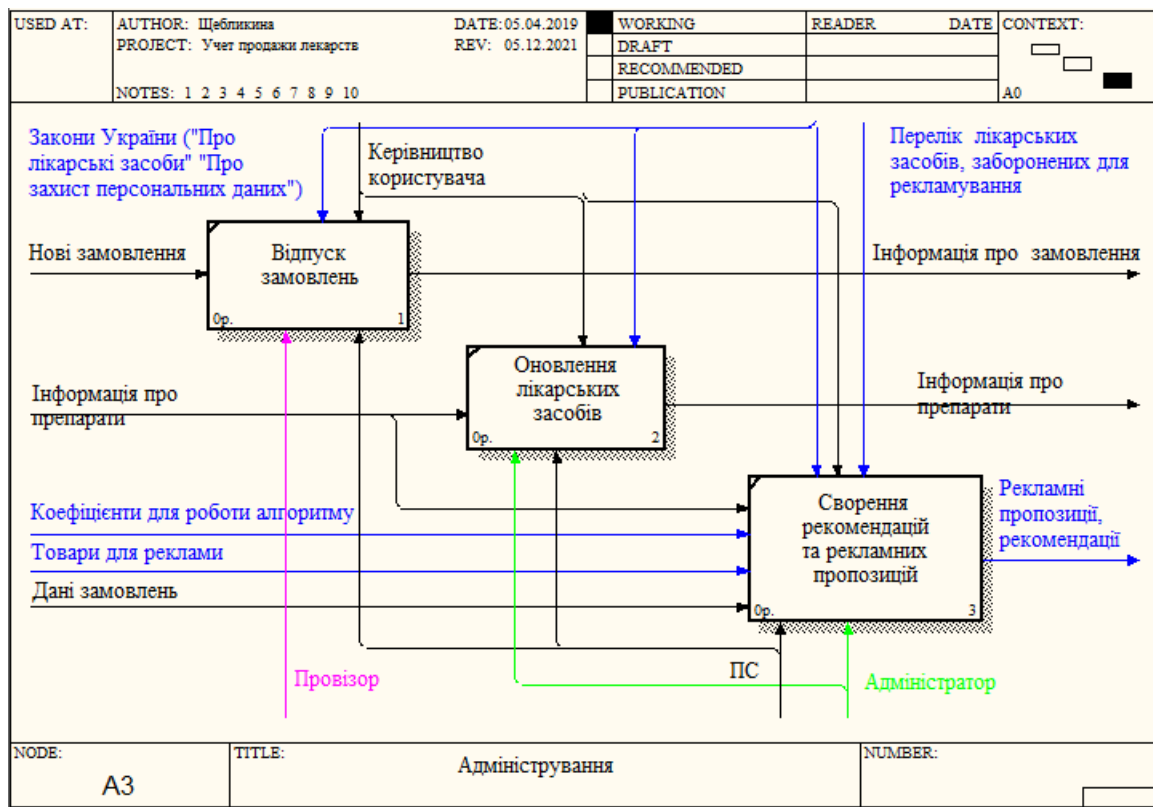


Рисунок 5.4 – Декомпозиція цільової функції «Адміністрування»

Для уточнення взаємодії системи зі сховищами даних та зовнішніми сутностями розглянуто та розширено DFD діаграму за нотацією Gane-Sarson.

На рисунку 5.5 представлено контекстну діаграму «Продажу препаратів», що містить зовнішні сутності, котрими виступають усі можливі клієнти системи, та головну роботу системи.

Для роботи аналітичного компоненту зовнішня сутність «Адміністратор» має передавати до системи:

- 73.перелік препаратів для реклами;
- 74.коефіцієнти для роботи алгоритму.

Оновлена система у результаті роботи буде повертати:

75.зовнішнім сутностям «Зареєстрованим» та «Незареєстрованим» користувачам рекомендації під час формування кошика;

76.«Зареєстрований» користувач отримуватиме персональні рекламні пропозиції.

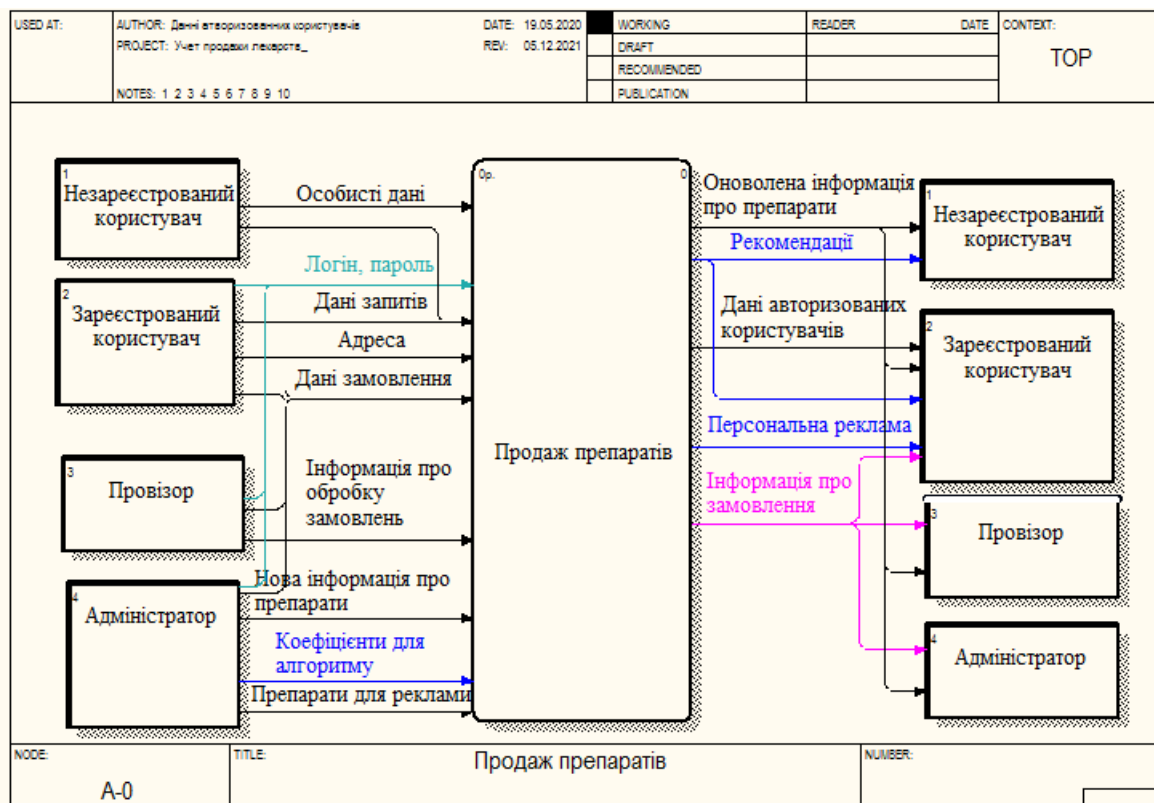


Рисунок 5.5 – Контекстна DFD діаграма «Продажу препаратів»

Декомпозиція контекстної DFD діаграми «Продажу препаратів» (рис. 5.6) дозволяє відобразити взаємодію робіт системи зі сховищами даних.

Розглянемо потоки, що взаємодіють з роботою «Адміністрування» зі сховищ даних та інших робіт.

В існуючій системі для виконання робіт з «Відпуску замовлення» та «Оновлення препаратів» залучаються потоки об'єктів зі сховища «Фрагмент БД «Замовлення»» та «Фрагмент БД «Препарати»». Таким чином, забезпечується вилучення об'єктів, їх опрацювання – зміна статусу поточних замовлень або оновлення існуючих описів лікарських препаратів, та їх збереження у сховищах даних.

Робота «Створення рекомендацій та рекламних пропозицій» (рис. 5.7) так само взаємодіє з вищенаведеними сховищами даних, додатково для вилучення та збереження знайдених рекомендацій вводиться сховище даних «Фрагмент БД «Рекомендації»».

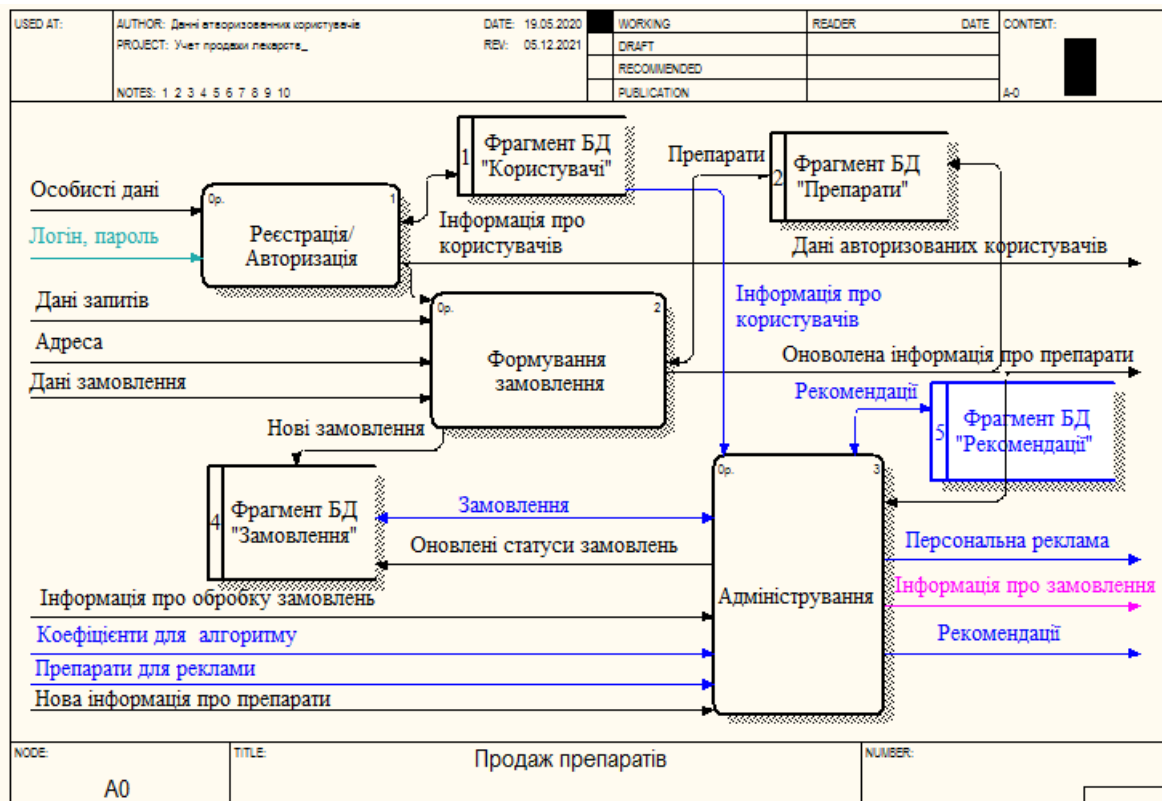


Рисунок 5.6 – Декомпозиція контекстної DFD діаграми «Продажу препаратів»

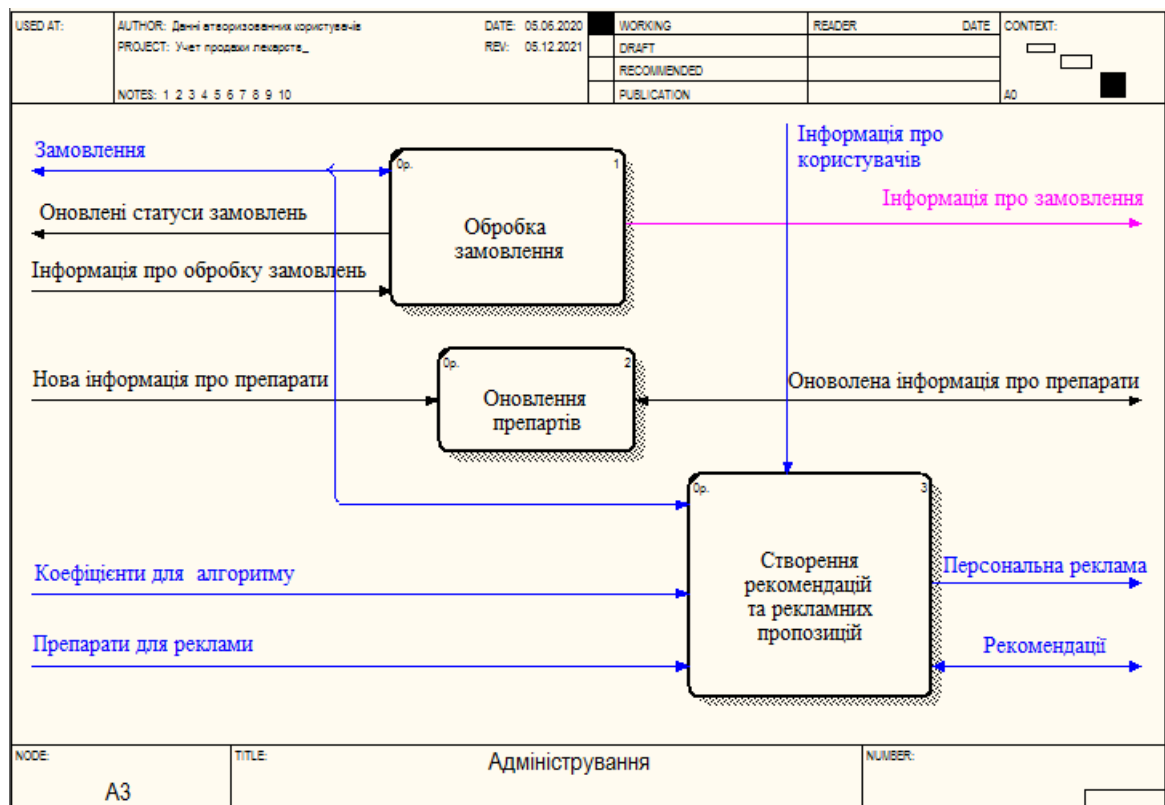


Рисунок 5.7 – Декомпозиція роботи DFD діаграми «Адміністрування»

У випадку необхідності швидкої генерації рекламних пропозицій, вже сформований перелік рекомендацій подається до роботи «Створення рекомендацій та рекламних пропозицій» зі сховища «Фрагмент БД «Рекомендації»». У протилежному випадку, рекомендації будуть розраховуватися знову й оновлювати зміст сховища даних «Фрагмент БД «Рекомендації»».

5.3 Розширення діаграми варіантів використання

Діаграма варіантів використання відображає, як системне оточення – актори, може взаємодіяти з частинами системи – прецедентами.

Діаграму використання для системи електронної комерції «Аптечної мережі» зображено на рисунку 5.8. Ставиться за мету внести до діаграми нові прецеденти.

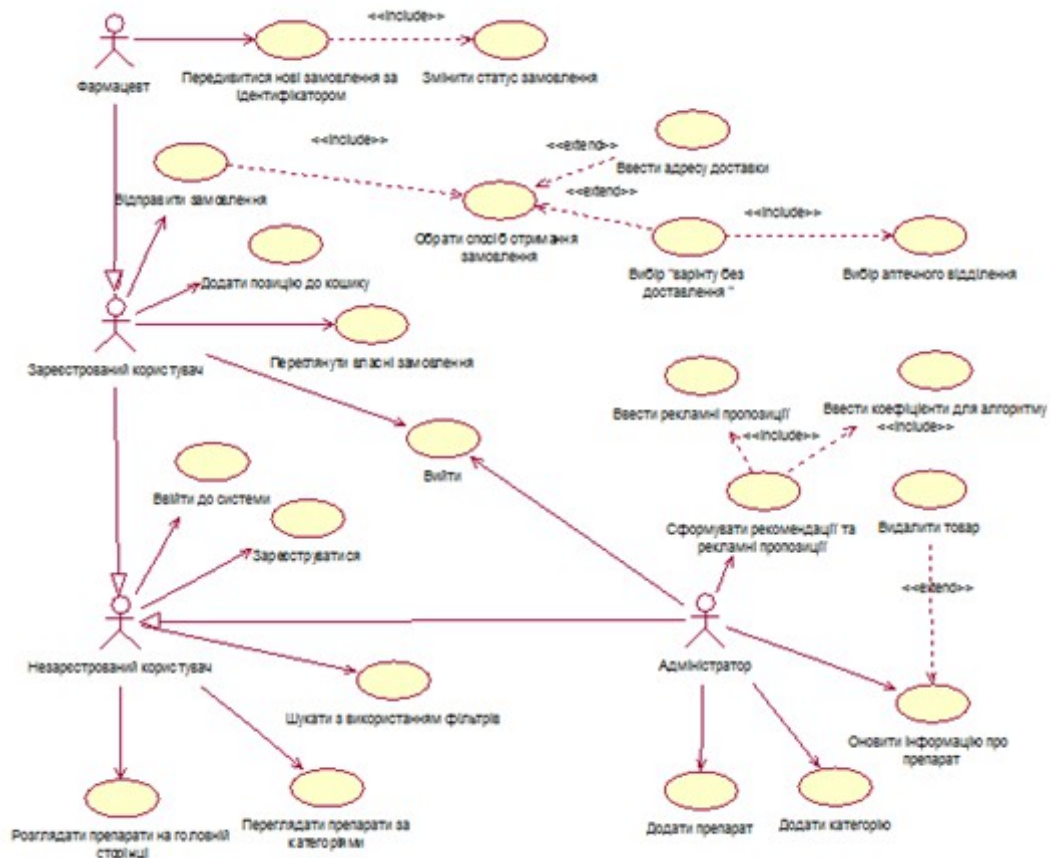


Рисунок 5.8 – Діаграма використання системи e-commerce «Аптечної мережі»

Система електронної комерції «Аптечної мережі» представлена наступними акторами, а саме «Адміністратором», «Фармацевтом», «Зареєстрованим користувачем» та «Незареєстрованим користувачем».

Актор «Незареєстрований користувач» має змогу оперувати прецедентами:

77. переглянути фармацевтичної продукції лікувального та нелікувального призначення на головній сторінці;

78. переглянути фармацевтичну продукцію лікувального та нелікувального призначення за обраною категорією;

79. здійснювати пошукову діяльність за представленими у системі фільтрами (виробник, країна походження, лікарська форма, міжнародне непатентоване найменування);

80. здійснити реєстрацію в системі;

81. здійснити вихід із системи.

Актор «Зареєстрований користувач» має доступ до усіх вищезазначених прецедентів та має змогу оперувати прецедентами:

82. додати позицію фармацевтичної продукції до кошика;

83. переглянути власні замовлення, зроблені під час попередньої взаємодії за системою;

84. відправити сформоване замовлення, яке включає в себе вибір бажаного способу одержання замовлення – скористатися директною доставкою або зазначити пункт видачі замовлення серед запропонованих у переліку аптечних відділень.

Актор «Фармацевт» має доступ до усіх вищезазначених прецедентів та має змогу оперувати прецедентом: здійснити перегляд доданих до системи нових замовлень за ідентифікатором, що включає в себе прецедент зміни їх статусу.

Актор «Адміністратор» має доступ до усіх прецедентів, що є у доступі для «Незареєстрованого користувача» та має змогу оперувати зазначеними нижче прецедентами:

85.здійснити оновлення (актуалізувати) інформацію, щодо препаратів, наявних у системі;

86.здійснити видалення препарату з завершеним терміном дії реєстраційного посвідчення або знятого з виробництва;

87.здійснити додавання нової позиції – препарату або медичної продукції нелікувального призначення;

88.здійснити додавання нової категорії;

89.формувати рекомендації та таргетовані рекламні пропозиції;

90.здійснити введення коефіцієнтів для роботи алгоритму;

91.внести перелік акційних пропозицій до системи – акції від виробника, сезонні пропозиції від аптечного закладу.

5.4 Розширення логічної та фізичної моделювання даних системи e-commerce «Аптечної мережі»

Проектування систем складається з різноманітних рівнів абстракції, що зумовлено потребами в деталізації системи кінцевих розробників.

Логічний рівень представляє розширення концептуального рівня – відображає сутності та зв'язки. На логічних моделях відображують атрибути сутностей – визначають у тривіальних типах даних – строкових, двійкових або цілочислених.

Моделювання на фізичному рівні додає до логічної моделі специфіки у відповідності до обраної СУБД. У загальному випадку тривіальні типи даних заміщуються більш вузько вживаними типами СУБД. Також, на функціональному рівні можна оперувати не тільки таблицями – розширений еквівалент сутностей логічного моделювання, а й більш специфічними об'єктами СУБД – збереженими функціями, уявленнями, тригерами, додатковими обмеженнями та індексами.

Результат логічного моделювання з урахуванням інтеграції аналітичного компоненту до інформаційної системи «Продажу препаратів» наведено на рисунку 5.9.

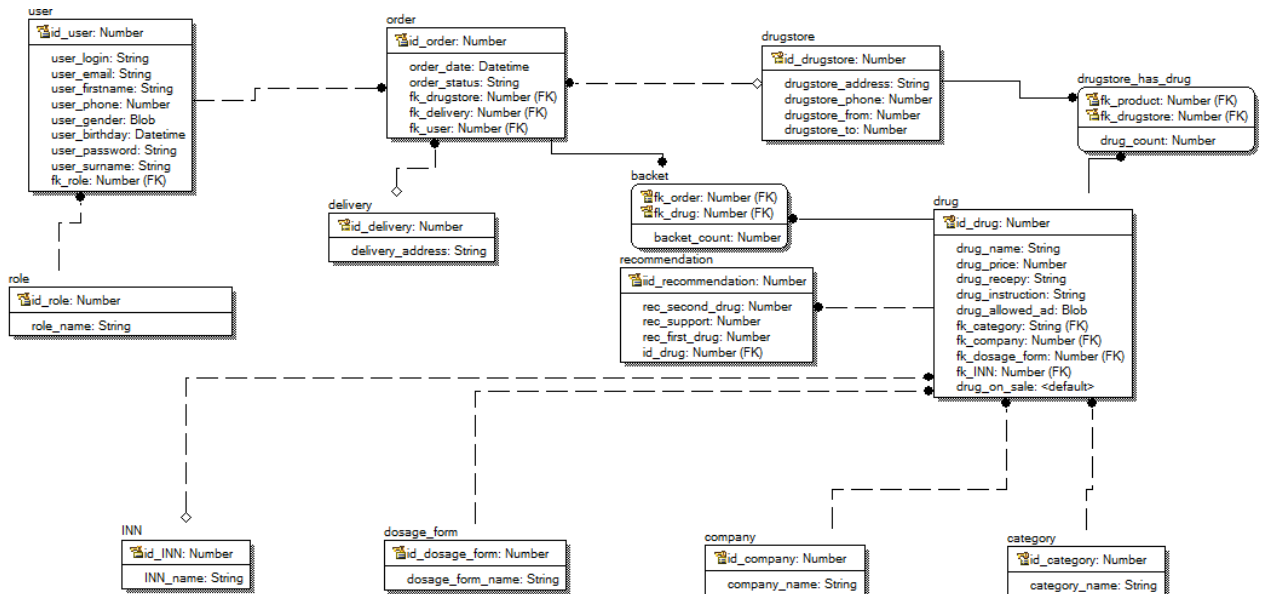


Рисунок 5.9 – Результат логічного моделювання

Функціонал CA Erwin Data Modeler [23] дозволяє деталізувати модель за рівнями та швидко переходити між логічною та фізичними рівнями моделі, що значно спрощує проектування.

До логічної моделі спроектовано наступний перелік сутностей: «user», «order», «drugstore», «drugstore_has_drug», «role», «delivery», «basket», «drug», «INN», «dosage_form», «company», «category» та «recommendation».

Сутність «user» використовується для збереження даних про користувачів системи, включає в себе відношення з відношенням «багато до одного» з сутністю «role» – визначає права користувача в системі. Заповнюється під час реєстрації користувача у системі.

Сутності «order», «delivery» та «drugshop» використовуються для збереження системних свідчень про замовлення – таких як дата створення та поточний статус, адреса доставки – якщо обрана ця функція або адресу аптечного відділу.

Сутність «basket» використовується для збереження обраних препаратів та їх кількості в рамках одного замовлення.

Сутність «drug» містить у собі лікарські препарати для продажу, має у відношення «багато-до-одного» з сутностями «INN», «dosage_form», «company» та «category», котрі в свою чергу визначають параметри для фільтрації під час пошуку препаратів. Дана сутність зазнала розширення, додано атрибути:

- «drug_allow_ad» – визначає дозвіл на рекламування препарату з боку законодавства;

- «drug_on_sale» – визначає чи доступний даний препарат до продажу. Даний атрибут введено з судження про виключення ситуації видалення препаратів, що більше не продаються з будь-яких причин – перереєстрації чи вилучення з асортиментної лінійки. Використання даного атрибуту дає змогу уникнути зайвий розрахунок та надання рекомендацій на вже неактуальний препарат.

Сутність «drugstore_has_drug» використовується для збереження препаратів та їх кількості в рамках одного аптечного відділу.

Сутність «recommendation» є новою для логічної моделі. Її введення зумовлено необхідністю зберігати рекомендації для кожного препарату системи. Кожен об'єкт сутності ідентифікується наступними атрибутами:

- «id_recommendation» – є унікальним ідентифікатором для об'єкту сутності, не має додаткового значення;

- «fk_drug» – є зовнішнім зв'язком з сутністю «drug», з встановленою потужністю «багато-до-одного» – кожен препарат може часто купуватися у різних комбінаціях;

- «rec_first_drug» – є унікальним ідентифікатором для вилучення рекомендованого препарату у зв'язці з вказаним у зовнішньому ідентифікаторові fk_drug;

- «rec_second_drug» – містить унікальний ідентифікатор для вилучення рекомендованого препарату у зв'язці з вказаним у зовнішньому ідентифікаторові fk_drug та попереднім атрибутом rec_first_drug;

– «rec_support» – визначена підтримка правила. Є основою для знаходження найбільш вигідних пропозицій.

У зв'язку з тим, що логічної моделі у випадку інтегрування методу, не є достатнім – спроектовано фізичну модель системи «Продажу ліків» (рис. 5.10).

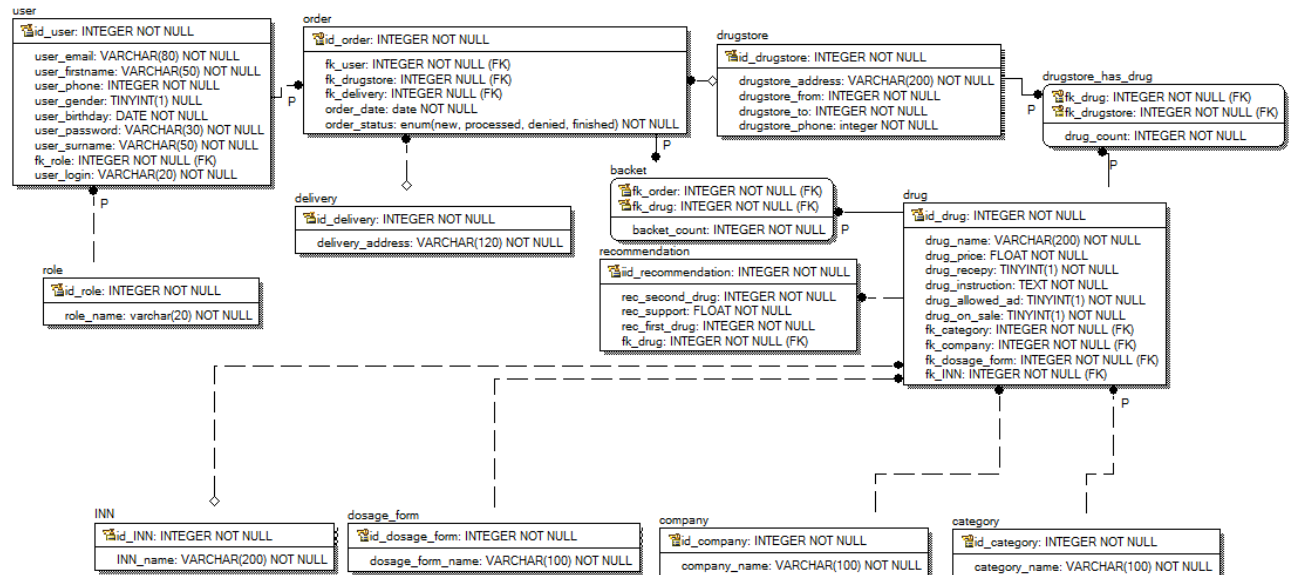


Рисунок 5.10 – Результат фізичного моделювання

Детальний опис кожної таблиці моделі, її атрибутів та типу, допустимих значень наведено у таблиці 5.1.

Таблиця 5.1 – Опис атрибутів фізичної моделі

Назва таблиці	Атрибут	Тип атрибуту	Null значення	Опис
User	id_user	INTEGER	NOT NULL	PK
	user_firstname	VARCHAR	NOT NULL	ім'я
	user_surname	VARCHAR	NOT NULL	прізвище

Продовження таблиці 5.1

	user_phone	INTEGER	NOT NULL	номер телефону користувача
	user_gender	TINYINT(1)	NOT NULL	стать (0 – жінка, 1- чоловік)
	user_email	VARCHAR	NOT NULL	email користувача

	user_birthday	DATE	NOT NULL	дата народження
	user_login	VARCHAR	NOT NULL	логін
	user_password	VARCHAR	NOT NULL	пароль
	fk_role	INTEGER	NOT NULL	FK таблиці role
Role	id_role	INTEGER	NOT NULL	PK
	role_name	VARCHAR	NOT NULL	назва ролі
Order	id_order	INTEGER	NOT NULL	PK
	order_date	DATE	NOT NULL	дата замовлення
	order_status	ENUM	NOT NULL	поточний статус замовлення
	fk_delivery	INTEGER	NULL	FK таблиці delivery
	fk_drugstore	INTEGER	NULL	FK таблиці drugstore
	fk_user	INTEGER	NOT NULL	FK таблиці user
drug-store	id_drugstore	INTEGER	NOT NULL	PK
	drugstore_address	VARCHAR	NOT NULL	адреса аптеки
	drugstore_phone	INTEGER	NOT NULL	номер телефону аптеки

Продовження таблиці 5.1

Drug	drugstore_from	INTEGER	NOT NULL	час відкриття відділу
	drugstore_to	INTEGER	NOT NULL	час закриття відділу
	id drug	INTEGER	NOT NULL	PK
	drug_name	VARCHAR	NOT NULL	назва препарату
	drug_price	FLOAT	NOT NULL	ціна препарату
	drug_recepy	TINYINT(1)	NOT NULL	дозвіл на продаж без рецепту (0 – можна, 1- заборонено)
	drug_instruction	TEXT	NOT NULL	інструкція
	drug_allowed_ad	TINYINT(1)	NOT NULL	дозвіл на рекламу (0 – можна, 1- заборонено)
	drug_on_sale	TINYINT(1)	NOT NULL	визначає чи можна придбати препарат (0 – можна, 1- заборонено)
	fk_category	INTEGER	NOT NULL	FK таблиці category
	fk_company	INTEGER	NOT NULL	FK таблиці company
	fk_dosage_form	INTEGER	NOT NULL	FK таблиці dosage_form

Продовження таблиці 5.1

	fk INN	INTEGER	NOT NULL	FK таблиці INN
	id_recommendation	INTEGER	NOT NULL	PK
	rec_first_drug	INTEGER	NOT NULL	рекомендований препарат у зв'язці з зовнішнім FK таблиці drug
	rec_second_drug	INTEGER	NULL	рекомендований препарат у зв'язці з зовнішнім FK таблиці drug та

				rec first drug
	rec_support	FLOAT	NOT NULL	підтримка правила
	fk_drug	INTEGER	NOT NULL	FK таблиці drug
drugstore_ has_drug	drug_count	INTEGER	NOT NULL	PK
	fk_drugstore	INTEGER	NOT NULL	FK таблиці drugstore, частина складеного PK
	fk_drug	INTEGER	NOT NULL	FK таблиці drug, частина складеного PK
Delivery	id_delivery	INTEGER	NOT NULL	PK

Продовження таблиці 5.1

	delivery_address	VARCHAR	NOT NULL	адреса, на котру необхідно доставити замовлення
Bucket	bucket_count	INTEGER	NOT NULL	кількість препарату, визначеного fk_drug
	fk_order	INTEGER	NOT NULL	FK таблиці order, частина складеного PK
	fk_drug	INTEGER	NOT NULL	FK таблиці drug, частина складеного PK
INN	id INN	INTEGER	NOT NULL	PK
	INN_name	VARCHAR	NOT NULL	міжнародне непатентоване найменування препарату
dosage_form	id dosage_form	INTEGER	NOT NULL	PK
	dosage_form_ name	VARCHAR	NOT NULL	назва лікарської форми
Company	id company	INTEGER	NOT NULL	PK
	company_name	VARCHAR	NOT NULL	назва компанії виробника

Category	id category	INTEGER	NOT NULL	PK
	category_name	VARCHAR	NOT NULL	назва категорії препарату

У результаті проведеного моделювання оновлено схему бази даних та сформовано засобами попередньої інженерії (forward engineering) – вбудованою функцією CA Erwin Data Modeler, DDL script.

5.5 Інтеграція методу до системи

У сформований на попередньому етапі script додано:

92.індекси у відповідності до логіки застосунку. У випадку пошуку за певним значенням за неіндексованим полем буде застосовуватися перегляд усіх кортежів таблиці, що погано впливає на швидкість виконання запитів. Задля запобігання виникнення даної проблеми на всі поля, за котрими найчастіше здійснюється пошук, додані індекси. Наприклад, поля «user_login» та «user_email» за логікою містять тільки унікальні значення, так гарантується ще й збереження цілісність логіки застосунку;

93.значення за замовченням для булевих атрибутів;

94.ключове слово auto increment для автоматичної генерації унікальних ключових ідентифікаторів вбудованими засобами СУБД;

95.уявлення – об’єктів БД, що оновлюються в момент звернення до них й дозволяють оперувати собою, в залежності від зазначеного алгоритму, як з таблицею. Основними перевагами використання уявлень є зниження вартості створення запиту за рахунок об’єднання даних з декількох таблиць в одну віртуальну – один запит до уявлення замість багато табличного.

Оновлений скрипт бази даних наведено у додатку А.

На рисунку 5.1 наведено послідовність обробки запитів та взаємодіє внутрішніх компонентів системи – зміни у схемі бази даних суттєво впливають на рівень репозиторію та сервісів.

Так як за мапінг таблиць у класи даних й навпаки відповідає ORM Hibernate додано:

96. всі оновлення у структурі бази даних до класів даних, що означає включення до класів нових полів та відповідних уточнюючих анотацій, створення класів даних для нових таблиці та уявлень;

97.HQL запити для отримання необхідних результатів з БД – а саме: вилучення усіх рекомендацій відповідного препарату, вилучення переліку усіх препаратів з заборорою для реклами, запити для видалення та занесення розрахованих рекомендацій;

98.в конфігураційний файл атрибут для створення та регуляції генерації бази даних (hibernate.hbm2ddl.auto) на validate, таким чином кожного разу при зверненні або розгортанні застосунку всі внесені дані будуть залишатися незмінними.

Рівень сервісів відповідає за поєднання даних з декількох репозиторіїв та додаткової бізнес логіки застосунку. Створено сервіс, для обробки рекомендацій, що реалізує:

99.завантаження необхідних за логікою об'єктів класів даних;

100. передачу аргументів до методу генерації рекомендацій та особистих таргетованих пропозицій з клієнта;

101. виклик методів для фіксації отриманих результатів у базі даних.

Взаємодія клієнта з застосунком відбувається через введення даних у браузері, їх відправки та опрацюванні на контролерах системи та поверненні результату до браузеру. Таким чином, додано в серверній частині за стосунку URL та отримуванні з клієнта параметри, за якими надається доступ до бажаного функціоналу. Для цього необхідно:

102. вказати над контролером URL доступу;

103. встановити над сутностями або параметрами, що надходять до контролеру з клієнта анотації для нетривіальних обмежень;

104. прописати виклик усіх задіяних сервісів.

Клієнтська частина має відображати простий та інтуїтивно зрозумілий користувацький інтерфейс. Представлені на елементи на сторінці в браузері мають в повній мірі забезпечувати приємне враження від взаємодії та мінімізувати шанс помилки. Для даної системи слід розробити сторінку з наступними елементами:

105. пошуковим полем для введення рекламних пропозицій, що за необхідно таргетовано запропонувати;

106. поля для встановлення мінімальної підтримки та достовірності;

107. наявність перемикача для включення (значення за замовченням) або вимкнення генерації асоціативних правил. Таким чином, можливо терміново робити рекламні розсилки.

108. кнопка для відправлення даних для обробки на сервері.

Код розробленого методу та код, необхідний для його інтегрування наведено у додатку Б та В відповідно.

ВИСНОВКИ

Під час виконання атестаційної роботи було розроблено метод аналітичного компоненту CRM системи на прикладі аптечної мережі.

Проаналізовано стан CRM систем відносно інформаційних систем, розглянуто компоненти CRM системи та їх основне призначення. Виокремлено аналітичний компонент та його функції.

Під час аналізу предметної області визначено слабкі та сильні сторони існуючих аптечних мереж на прикладі «Аптеки низьких цін» та «Аптеки 911», наведено доцільність введення аналітичного компоненту.

Окремо проведено дослідження для визначення найбільших пригодних методів для реалізації поставленої задачі:

- визначено групу методів, які не потребують залучення експертів;
- детально розглянуто методи кластеризації – ієрархічні та неієрархічні, рекомендаційні системи на основі колаборативної фільтрації, рекомендаційні системи на основі контенту, методи генерації асоціативних правил;
- для кожного розглянутого методу наведено перелік можливих метрик, мір, окремі випадки їх застосування;
- охарактеризовано переваги та недоліки методів.

За результатом дослідження виявлено методи та їх компоненти, що доцільно використовувати у комбінації для обраної предметної області, а саме: метод генерації асоціативних правил Apriori, один з його ключових компонентів – підтримка, та матрицю вподобань. Розроблений метод детально описано у четвертому розділі. Додатково приведено можливі варіанти оптимізації алгоритму методу.

Розглянуто необхідні для додання компоненти під час проектування системи. Наведено послідовність інтегрування розробленого методу до існуючої системи електронної комерції на прикладі аптечної мережі.

Роботу було апробовано на міжнародній конференції [25].

ПЕРЕЛІК ДЖЕРЕЛ

1. Кахоцкий П. С., Ремпель А. П., Хуткина Г. А. Перспективы внедрения и интеграции CRM-систем в фармацию. *Вестник фармации*. 2015. № 4 (70). С. 22–27.
2. Oliveira V. L. M. Analytical customer relationship management in retailing supported by data mining techniques. Universidade do Porto. 2012. URL: <https://repositorio-aberto.up.pt/bitstream/10216/64884/2/26561.pdf>
3. A Case for Analytical Customer Relationship Management. 6th International Conference on Knowledge Discovery and Data Mining (PAKDD-02). Research Collection School Of Information Systems / Srivastava Jaideep, Wang Jau-Hwang, Lim Ee Peng, Hwang San-Yih. 2002. URL: https://ink.library.smu.edu.sg/sis_research/970.
4. Gullo [Francesco](#) From Patterns in Data to Knowledge Discovery: What Data Mining Can Do [Physics. Procedia](#). 2015. Vol. 62. P. 18–22. DOI: [10.1016/j.phpro.2015.02.005](https://doi.org/10.1016/j.phpro.2015.02.005).
5. Чубукова И. А. Data Mining. URL: https://portal.tpu.ru/departments/kafedra/vt/Disciplines_VT/Data_storehouses/FilesTab/Tab/lections%20data%20mining.pdf#page50 (дата звернення: 05.10.2021).
6. Ершов К. С., Романова Т. Н. Анализ и классификация алгоритмов кластеризации. *Новые информационные технологии в автоматизированных системах*. 2016. № 19. С. 274–279. URL: <https://cyberleninka.ru/article/n/analiz-i-klassifikatsiya-algoritmov-klasterizatsii/pdf>.
7. Yildirim P., Birant D. K-Linkage: A New Agglomerative Approach for Hierarchical Clustering. *Advances in Electrical and Computer Engineering*. 2017. Vol. 17, N 4. P. 77–88. DOI: [10.4316/AECE.2017.04010](https://doi.org/10.4316/AECE.2017.04010)
8. Sandeep K. Raghuwanshi, Pateriya R. K. Recommendation Systems: Techniques, Challenges, Application, and Evaluation. In *Soft Computing for Problem Solving*. 817 / J. C. Bansal, K. N. Das, A. Nagar et al. Springer Singapore, Singapore, 2019. P. 151–164. URL: <https://www.researchgate.net/profile/Sandeep->

[Raghuwanshi/publication/328640457_Recommendation_Systems_Techniques_Challenges_Application_and_Evaluation_SocProS_2017_Volume_2/links/5cc7d508a6fdc1d49b9985b/Recommendation-Systems-Techniques-Challenges-Application-and-Evaluation-SocProS-2017-Volume-2.pdf?origin=publication_detail](https://raghuwanshi/publication/328640457_Recommendation_Systems_Techniques_Challenges_Application_and_Evaluation_SocProS_2017_Volume_2/links/5cc7d508a6fdc1d49b9985b/Recommendation-Systems-Techniques-Challenges-Application-and-Evaluation-SocProS-2017-Volume-2.pdf?origin=publication_detail)

DOI:10.1007/978-981-13-1595-4_12.

9. Чепикова Е. Д., Савкова Е. О., Привалов М. В. Исследование алгоритмов рекомендательных систем. *Информатика и кибернетика*. 2016. № 2 (4). С. 57–61. URL: http://infcyb.donntu.org/A_4_8.pdf.

10. Content-based video recommendation system based on stylistic visual features / Deldjoo Y., Elahi M., Cremonesi P. et al. *Lect. Notes Comput. Sci.* 2016. Vol. 5. P. 99–113. URL: <https://re.public.polimi.it/retrieve/handle/11311/1006121/161919/Journal%20of%20Data%20Semantics.pdf>.

11. Ткаченко А. Сравнение текста с помощью статистической меры TF-IDF. – URL: <https://newtechaudit.ru/kak-sverit-tekst-tf-idf/> (дата звернения: 27.09.2021).

12. . Акобир Шахиди Введение в анализ ассоциативных правил. URL: <https://loginom.ru/blog/associative-rules> (дата звернения: 23.09.2021).

13. Хайдаров К. А. Методы и средства анализа данных. Поиск ассоциативных правил. URL: <http://bourabai.kz/tpoi/analysis4.htm> (дата звернения: 01.10.2021)..

14. [Garg A.](#) Complete guide to Association Rules (1/2). URL: <https://towardsdatascience.com/association-rules-2-aa9a77241654> (дата звернения: 04.10.2021).

15. Левус Є. В., Нечипір Н. І., Поляняк Ю. В. Аналіз алгоритму Apriori для структурованих та неструктурованих даних. *Вісник Національного університету «Львівська політехніка»*. Серія: Інформаційні системи та мережі. Львів : Видавництво Львівської політехніки, 2017. № 872. С. 62–68. URL: <http://science.lpnu.ua/sites/default/files/journal-paper/2018/jun/13003/ilovepdfcom-62-68.pdf>.

16. Закон України "Про лікарські засоби" [Електронний ресурс] – Режим доступу до ресурсу: <https://zakon.rada.gov.ua/laws/show/123/96-%D0%B2%D1%80#Text>.

17. Наказ МОЗ "Про затвердження Переліку лікарських засобів, заборонених до рекламування, які відпускаються без рецепта" [Електронний ресурс] – Режим доступу до ресурсу: <https://zakon.rada.gov.ua/laws/show/z1948-12#Text>

18. Kumar M. Spring MVC Framework Tutorial. – URL: <https://www.javawebtutor.com/articles/spring/spring-mvc-tutorial.php>.

19. Web MVC framework. – URL: <https://docs.spring.io/spring-framework/docs/3.2.x/spring-framework-reference/html/mvc.html>.

20. The Java EE 5 Tutorial. – URL: <https://docs.oracle.com/javaee/5/tutorial/doc/bnagz.html#bnagz>.

21. Справочное руководство по MySQL [Електронний ресурс] - Режим доступу URL: <http://www.mysql.ru/docs/man/Introduction.html>.

22. Hibernate Getting Started Guide – URL: https://docs.jboss.org/hibernate/orm/5.5/quickstart/html_single/#tutorial_annotations.

23. Маклаков, С. В. BPwin и ERwin. CASE-средства разработки информационных систем [Текст] / С. В. Маклаков. – М.: Диалог, 1999. – 256.

24. Каюмова А. В. Визуальное моделирование систем в StarUML: Учебное пособие/ Казань. – Казанский федеральный университет, 2013. – 104 с.

25. Щебликіна Д. Ю. розробка аналітичного методу для компоненту CRM системи на прикладі аптечної мережі // Topical issues of modern science, society and education. Proceedings of the 5th International scientific and practical conference. SPC —Sci-conf.com.ua, Kharkiv, Ukraine..2021. Pp. 657-662. URL: <https://sci-conf.com.ua/v-mezhdunarodnaya-nauchno-prakticheskayakonferentsiya-topical-issues-of-modern-science-society-and-education-28-30-noyabrya-2021-goda-harkov-ukraina-arhiv/>.