ДИПЛОМНА РОБОТА
БАКАЛАВРА

ДРБ. ІП - 46.00.00.000 ПЗ

Група ІП-18-1

Склеповий Роман

2022

Міністерство освіти і науки України
Івано-Франківський національний технічний університет нафти і газу

Інститут інформаційних технологій

Кафедра інженерії програмного забезпечення

УДК 004.4

ДИПЛОМНА РОБОТА БАКАЛАВРА

Тема: " Розробка рішення та сервісу контролю витрат для клієнтів українських банків за допомогою технологій Nest.js, React.js "

(назва згідно з наказом ректора)

спеціальність – 121 "Інженерія програмного забезпечення"

ПОЯСНЮВАЛЬНА ЗАПИСКА

ДРБ.ІП - 46.00.00.000 ПЗ

(позначення)

Консультанти:

з питань нормоконтролю

доц., к.т.н_____ /В.Я. Піх/
(посада) (підпис) (дата)(розшифровка підпису)

Студент

ІП-18-1_____ /Р.Склеповий/
(шифр групи) (підпис) (дата) (розшифровка підпису)

Керівник роботи

к.ф.-м.н.,доц. /Р.І.Храбатин/
(посада) (підпис) (дата) (розшифровка підпису)

Допускається до захисту
Завідувач кафедри

д.т.н.,проф. /В.І.Шекета/
(посада) (підпис) (дата) (розшифровка підпису)

2022
(рік)

Івано-Франківський національний технічний університет нафти і газу

Інститут, факультет інформаційних технологій

Кафедра інженерія програмного забезпечення

Освітньо-кваліфікаційний рівень бакалавр

Спеціальність 121 – Інженерія програмного забезпечення

ЗАТВЕРДЖУЮ:
Зав. кафедрою **ІПЗ**
проф. В.І. Шекета
“ ____ ” _____2022 р.

ЗАВДАННЯ

НА ДИПЛОМНУ РОБОТУ БАКАЛАВРА СТУДЕНТОВІ

Склеповому Роману

(прізвище, ім’я, по-батькові)

1. Тема проекту (роботи) **“ Розробка рішення та сервісу контролю витрат для клієнтів українських банків за допомогою технологій Nest.js, React.js”**

керівник проекту (роботи) Храбатин Роман Ігорович, доцент,

затверджені наказом закладу вищої освіти від **“** ___ ” __травня___ 2022 р. № _____

2. Строк подання студентом проекту (роботи) 20 червня 2022 р.

3. Вихідні дані до проекту (роботи) Результати і матеріали отримані під час проходження переддипломної практики

4. Зміст розрахунково - пояснювальної записки (перелік питань, які потрібно розробити)

1. Аналіз предметної області рішення задачі конфіденційності в програмних сервісах
2. Розробка моделі рішення програмного сервісу
3. Представлення інтерактивного інструменту візуалізації
4. Представлення інтерфейсу
5. Виконання процесу тестування розробленого програмного забезпечення

5. Перелік графічного матеріалу (з точним зазначенням обов’язкових креслень)

1. Знімок графіка конфіденційності (рис. 1.1)
2. Приклад речення, розібраного у формі RST (рис. 2.1)
3. Огляд високого рівня підходу до виділення сигналу та розмежування (рис. 2.2)
4. Знімок екрана ProjectAcc (рис. 2.3)
5. Схема бази даних (рис. 3.1)

6. Консультанти розділів проекту (роботи)

Розділ	Консультант	Підпис, дата	
		Завдання видав	Завдання прийняв

7. Дата видачі завдання **25 травня 2022 р.**

Керівник _____
(підпис)

Завдання прийняв до виконання_____
(підпис)

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів дипломного проекту (роботи)	Строк виконання етапів проекту	Примітка
1	Аналіз предметної області рішення задачі конфіденційності в програмних сервісах	26.05.2022	виконано
2	Розробка моделі рішення програмного сервісу	29.05.2022	виконано
3	Представлення інтерактивного інструменту візуалізації	30.05.2022	виконано
4	Програмна реалізація інтерфейсу користувача	04.06.2022	виконано
5	Тестування ефективності розробленого програмного забезпечення	10.06.2022	виконано
6	Оформлення пояснювальної записки дипломної роботи завідувачем кафедри	17.06.2022	виконано

Студент – дипломник _____
(підпис)

Керівник роботи _____
(підпис)

АНОТАЦІЯ

В дипломній роботі проведено аналіз та вивчення обґрунтувань, що може бути корисним для розуміння онлайн-аргументації та міркувань клієнта в цілому. Враховуючи значення підмножини відношень у ідентифікації обґрунтування, ми запропонували новий підхід для автоматичного виявлення таких відношень. Спочатку ми провели аналіз на основі фінансових даних, щоб отримати набір n-грам як потенційних лексичних сигналів, які можуть сигналізувати про кортежі відношень. Ці дані потім використовувалися в імовірнісних моделях на основі графів для визначення синтаксичного контексту, в якому це сигналізує про наявність певного відношення. Результати оцінки були представлені для різних ознак відношення, що підтверджують значення синтаксичних ознак для завдання розшифровки. Крім того, було показано, що використання графів для кодування синтаксичної інформації є більш узагальненим та ефективним підходом у порівнянні з прямим використанням синтаксичних функцій.

КЛЮЧОВІ СЛОВА: КОНТРОЛЬ ВИТРАТ, ПРОГРАМНИЙ СЕРВІС, ПРОЦЕС МІРКУВАННЯ, ЛЕКСИЧНИЙ СИГНАЛ, СИНТАКСИЧНА ІНФОРМАЦІЯ, ІДЕНТИФІКАЦІЯ.

ANNOTATION

In the thesis, an analysis and study of justifications was carried out, which can be useful for understanding online argumentation and reasoning of the client as a whole. Given the importance of a subset of relations in rationale identification, we proposed a new approach to automatically detect such relations. First, we performed an analysis based on financial data to obtain a set of n-grams as potential lexical signals that can signal relation tuples. This data was then used in probabilistic graph-based models to determine the syntactic context in which this signaled the presence of a particular relation. Evaluation results were presented for different relational features, confirming the importance of syntactic features for the decoding task. In addition, it has been shown that the use of graphs to encode syntactic information is a more general and efficient approach compared to the direct use of syntactic functions.

KEY WORDS: COST CONTROL, SOFTWARE SERVICE, REASONING PROCESS, LEXICAL SIGNAL, SYNTAX INFORMATION, IDENTIFICATION.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ 7
ВСТУП 8
1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ РІШЕННЯ ЗАДАЧІ КОНФІДЕНЦІЙНОСТІ В ПРОГРАМНИХ СЕРВІСАХ 12
1.1. Техніка SGNS як неявна матрична факторізація 12
1.2. Модель латентного фактора на основі сусідства 15
1.3. Атрибути профілю для передбачення конфіденційності 16
1.5. Обчислювальний аналіз колективного інтелекту 21
1.6. Методологія системного огляду 25
1.7. Аналіз структури інтенцій 38
2. РОЗРОБКА МОДЕЛІ РІШЕННЯ ПРОГРАМНОГО СЕРВІСУ 45
2.1 Семантичний опис структури 45
2.2 Аналіз літературних джерел по темі дипломної роботи 47
2.3. Структура пропонованого підходу 50
2.4. Оцінка отриманих результатів 55
2.5. Обгрунтування отриманої концепції 60
2.6. Представлення інтерактивного інструменту візуалізації 64
3. РОЗРОБКА АЛГОРИТМІЧНОГО ТА ПРОГРАМНОГО РІШЕННЯ ЗАДАЧІ ПОБУДОВИ СЕРВІСУ 68
3.1 Складові програмного рішення 68
3.2 Розробка серверної частини проекту 70
ВИСНОВКИ 77
ПЕРЕЛІК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ 78
БІБЛІОГРАФІЧНА ДОВІДКА 82

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

ТЗ – Технічне завдання

БД – База даних

СУБД - Система управління базами даних

ПЗ - Програмне забезпечення

ВСТУП

Виявлення прогалин у дослідженні в контексті соціальної конфіденційності у порівнянні з іншими рішеннями для вирішення проблеми дихотомії конфіденційності в соціальних мережах, автоматичне визначення уподобань конфіденційності показало скромний успіх. Тому ми переглянули існуючу літературу про алгоритми, які використовують переваги соціальних слідів користувачів, щоб охарактеризувати їхні переваги щодо конфіденційності. Ми класифікували та переглянули існуючі дослідження щодо визначення переваг конфіденційності відповідно до типу даних, а саме: демографічні характеристики та особливості профілю, соціальний контекст та особливості мережі, а також спільний вміст. Потенціал та обмеження підходів далі обговорювалися, де було виявлено низку прогалин. Наприклад, хоча більшість досліджень зосереджено на конфігураціях конфіденційності, характерних для однієї платформи соціальних медіа, менше уваги було приділено загальному моделюванню та характеристиці користувачів. Крім того, необхідні подальші дослідження, щоб вивчити більш широкий спектр типів даних, доступних у соціальних мережах. Крім того, більшість досліджень використовували контрольовані методи. Однак контрольовані методи можуть бути неможливими в контексті соціальних мереж, де маркована інформація часто становить дуже малу частину доступних даних. У такому разі неконтрольовані та напівкеровані методи можуть бути привабливими альтернативами в цьому контексті. Наші дослідницькі зусилля щодо вирішення проблеми дихотомії конфіденційності були засновані на результатах аналізу літератури. Наприклад, ми включили в наш підхід різноманітні типи даних і застосували підхід до спільної фільтрації без нагляду.

Виявлення дослідницьких прогалин у ширшому контексті дискурсоцентричного колективного інтелекту: багатообіцяюча література, що аналізує колективний інтелект, заснований на діалозі. Наш аналіз цієї роботи виявив відсутність зосередженості на середовищі, орієнтованому на завдання, відсутність складних методів для аналізу впливу процесів групової взаємодії один на одного, а також відсутність зосередженості на вивченні впливу групової взаємодії. процеси взаємодії учасників з часом. Подібно до нашого напряму дослідження соціальної конфіденційності, наші рішення вивчати сліди міркувань були пов'язані з оглядовою літературою. Наприклад, ми зосередилися на середовищі, орієнтованому на завдання, яке в першу чергу зосереджено на переконанні та зміні переконань, і проаналізували фактори, які можуть вплинути на переконання учасників.

Розуміння поведінки та характеристик користувачів із різною поведінкою та ставленням до конфіденційності: як початковий крок до розуміння поведінки користувачів із різною поведінкою конфіденційності, ми перевірили, чи по-різному налаштовані атрибути профілю користувачів Twitter із різними налаштуваннями конфіденційності. У результаті аналізу було визначено набір функцій, які використовувалися для прогнозування налаштувань конфіденційності користувачів.

Протягом останніх кількох десятиліть Інтернет перетворився з платформи для простого споживання вмісту до місця, де люди можуть створювати, ділитися, змінювати й обговорювати вміст різних форм. Цей прогрес призвів до явища соціальної мережі, яка сьогодні значно впливає на наше повсякденне життя. Безпрецедентний рівень інтелекту, отриманий внаслідок таких масштабних онлайн-взаємодій, є одним із помітних ефектів соціальної мережі. З метою зробити внесок у галузі веб-колективної та бізнес-аналітики, ми вперше визначили два соціальних феномена конфіденційності та переконання, які мають неймовірний вплив на парадигму дослідження. Для подальшого вивчення концепції соціальної конфіденційності ми спочатку

провели огляд літератури. Потім ми вивчили зв’язок різних типів даних соціальних медіа з перевагами конфіденційності користувачів і запропонували модель для визначення профілів соціальних мереж, для яких діє дихотомія конфіденційності. Потім ми зосередилися на феномені онлайн-міркування, проглянувши відповідні дослідження. Потім були проаналізовані й обговорені механізми та стратегії онлайн-переконання. Враховуючи значення риторичних відносин у вивченні переконання в тексті, було запропоновано новий підхід для виявлення та уникнення неоднозначності явних риторичних зв’язків із текстом. Нарешті, було запропоновано інтерактивний інструмент візуалізації, щоб відобразити значення моделей інтелекту та структур для подальшого розвитку інтелекту. Потім ми провели серію експериментів, щоб перевірити, чи локалізовані налаштування конфіденційності в соціальних мережах. У результаті ми знайшли набір підказок того, що налаштування конфіденційності користувачів подібні до конфіденційності їхніх соціальних контактів, що свідчить про те, що параметри конфіденційності локалізовані в соціальних мережах. Нарешті, контекст сусідства був використаний для характеристики атрибутів людей, які не знають конфіденційності. Зокрема, ми виявили, що користувачі, які не знають конфіденційності, публікують більше особистої та приватної інформації порівняно з тими, хто стежить за публічною обстановкою.

Розуміння механізмів онлайн-переконання: проаналізувавши набір коментарів із людськими анотаціями, ми дослідили різні виміри мови, тимчасові аспекти коментарів, а також атрибути користувачів-учасників та їхнє відношення до переконання. процес.

Протягом аналізу ми знайшли великий набір атрибутів, які є специфічними для переконливих коментарів, і дослідили їх передбачувану силу, щоб виявити коментарі, які переконують. Більшість ознак пов’язано з переконливим впливом різних компонентів мови, включаючи психологічні атрибути мови, якість і витонченість письма. Ці особливості привели до

класифікації переконливих і непереконливих коментарів з розумною продуктивністю. Попереднє вивчення пояснень користувачів щодо того, чому і як змінилося їхнє переконання, дозволило проаналізувати передбачувані причини переконання та порівняти їх із реальними, знайденими під час попереднього аналізу коментарів. Порівняння наших висновків, отриманих з онлайн-платформи з літературою переконання в традиційних умовах, виявило низку подібностей і відмінностей між двома процесами.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ РІШЕННЯ ЗАДАЧІ КОНФІДЕНЦІЙНОСТІ В ПРОГРАМНИХ СЕРВІСАХ

1.1. Техніка SGNS як неявна матрична факторізація

Визначення параметрів конфіденційності за допомогою спільної фільтрації. Надання персоналізованого вмісту може принести взаємну користь як компаніям, так і клієнтам. Однак ефективне профілювання користувачів залежить від збору великої кількості даних від користувачів. Отже, конфіденційність є наріжним каменем будь-якої діяльності з персоналізації, яка, якщо її не враховувати, вплине на задоволення, отримане від персоналізації, і може призвести до непоправних проблем з довірою. Це питання може бути особливо проблематичним у контексті соціальних мереж, де сьогодні відбувається величезна кількість щоденних соціальних взаємодій.

Незважаючи на те, що користувачі мають можливість захищати свої дані за допомогою засобів контролю конфіденційності, такі рішення щодо конфіденційності часто є складними і вимагають ретельного вивчення компромісів між потенційною соціальною вигодою та можливими ризиками конфіденційності. Тому багато користувачів уникають клопоту з налаштуванням конфіденційності та дотримуються налаштувань за замовчуванням [12]. Однак вони зазвичай не знають, що налаштування за замовчуванням часто відкриті та дозволені. Навіть серед тих, хто докладає зусиль, щоб керувати своєю конфіденційністю, багато хто все ще не усвідомлюють наслідків своїх рішень [7].

Незважаючи на обмежену кількість, попередні дослідження поведінки щодо конфіденційності мали скромний успіх у прогнозуванні уподобань користувачів щодо конфіденційності, покладаючись на їхні соціальні сліди [4, 2]. Тут ми пропонуємо підхід спільної фільтрації (CF), який поєднує методи на

основі сусідства з моделлю латентного фактора, виведеною з соціального графіка користувачів.

Підхід прогнозування конфіденційності є новим і спрямований на виявлення користувачів, які піклуються про конфіденційність, з загальнодоступними профілями. Внесок цієї роботи полягає в наступному:

1) адаптація гібридного методу CF для визначення переваг соціальної конфіденційності,

2) вивчення корисності атрибутів профілю для визначення переваг конфіденційності,

3) встановлення та аналіз властивостей соціального графіка з покращеною конфіденційністю,

4) виявлення набору прихованих факторів, пов'язаних з атрибутами конфіденційності.

Наше дослідження проведено в Twitter, де контроль конфіденційності відповідає бінарній специфікації. У Twitter користувачі можуть або слідувати загальнодоступним налаштуванням за замовчуванням, які вказують на те, що їхні твіти та контакти є загальнодоступними, або вони можуть змінити налаштування на захищені, що робить їхні твіти та контакти доступними лише для їх затверджених підписників.

Методи CF показали великі перспективи при розробці рекомендаційних систем, де невідомі переваги користувачів визначаються за допомогою відомих переваг інших користувачів. Такі підходи можуть бути особливо цінними в контексті соціальних медіа через наявність додаткових соціальних відносин. Проте було зроблено обмежені спроби адаптувати методи CF для прогнозування переваг конфіденційності.

Наприклад, Squicciarini et al. [11] спочатку формують соціальні кола на основі характеристик користувачів (наприклад, стать та хобі). Коли основний користувач завантажує новий об’єкт, система шукає соціальні кола, які, швидше за все, мають справу з об’єктом так само, як і користувач. Тоді політика

конфіденційності, яку використовує вибране коло, є основою для прогнозування політики для доданого об’єкта.

У [10] активне навчання та властивості соціального графа вперше використовуються для виявлення набору інформаційних контактів, які мають бути позначені як навчальні зразки.

У процесі маркування користувач вказує, чи бажає він/вона поділитися певним елементом із вибраним контактом. Потім мітки поширюються від позначених екземплярів до немічених на графіку. Це поширення керується показником подібності користувачів, який обчислюється на основі інформації профілю контактів, а також показників їх мережі та спільноти відповідних мереж.

CF також дотримується в [3], де набір функцій профілю, інтересів користувачів і конфігурації конфіденційності використовується для пошуку набору користувачів, подібних до основного користувача. У своїй роботі користувачів спочатку характеризують відповідно до їхніх уподобань щодо конфіденційності як фундаменталістів конфіденційності, прагматиків конфіденційності або байдужих до конфіденційності. Користувачів призначають до цих категорій на основі кількості їхніх загальнодоступних, налаштованих та приватних фотоальбомів. Потім використовується алгоритм K-найближчого сусіда, щоб визначити, до якої категоризації конфіденційності належить основний користувач.

Методи прогнозування конфіденційності, засновані на CF, в основному використовують методи на основі сусідства, де основою прогнозу є спостереження за конфіденційністю від сусідніх або подібних користувачів. Однак моделі латентного фактора, які є сучасним станом для MB [1], ще не досліджені. У попередніх роботах також бракує вивчення гібридних методів, які показали значно кращі результати порівняно з моделями чистого сусідства та чисто латентного фактора в інших областях [6].

1.2. Модель латентного фактора на основі сусідства

Двома популярними методами CF є методи на основі сусідства та моделі латентних факторів. У методах, заснованих на сусідстві, спостереження сусідніх та/або подібних користувачів використовуються для виявлення атрибутів і переваг основного користувача. Виявлення налаштувань конфіденційності за допомогою таких методів, орієнтованих на користувача, створює унікальні проблеми через відсутність доступної інформації, пов'язаної з приватними обліковими записами.

У Twitter обмежена кількість атрибутів профілю доступна як для загальнодоступних, так і для захищених облікових записів. Отже, такі атрибути профілю можна використовувати для вимірювання схожості користувачів. Налаштування конфіденційності основного загальнодоступного користувача і можна визначити, використовуючи наступне

$$Σi0 ϵ Ωiωii0 xi0 yi = Σi0 ϵ Ωiωii0$$

де ωii0 вимірює подібність між користувачем і та його сусідами i0 ϵ Ωi. Крім того, xi0 ϵ {ε,1} вказує фактичне налаштування конфіденційності сусіда, де ε представляє дуже мале значення.

Виходячи з цієї формули, більше значення yi вказує на більш високий рівень занепокоєння конфіденційністю користувача i. Щоб отримати уявлення про можливості та обмеження атрибутів профілю Twitter для нашого завдання, ми провели ряд експериментів.

Ці експерименти, які коротко представлені в наступному розділі, припускають значення атрибутів профілю у висновку про переваги конфіденційності та вказують на їх потенціал для вимірювання подібності в підході на основі сусідства.

1.3. Атрибути профілю для передбачення конфіденційності

Тим часом моделі латентного фактора здобули величезний успіх у контексті рекомендаційних систем. Такі методи мають на меті виявити набір інформаційних прихованих факторів щодо користувачів і пізніше використовувати такі атрибути для визначення уподобань. Ми можемо обчислити ймовірність того, що публічний користувач і піклується про конфіденційність, використовуючи наступне:

$$p(yi) = p(yi|\theta 1,...,\theta k) = \Pi j=1,..,kp(yi|\theta j)$$

де ui = θ1,...,θk і p(yi|θj) вказує на ймовірність того, що користувач і буде стурбований конфіденційністю, якщо користувач і пов'язаний з прихованим фактором θj. Далі в цьому документі ми пропонуємо методику виявлення таких прихованих змінних із соціального графіка користувачів. На цьому графіку користувачі спочатку перетворюються на набір прихованих змінних. Ймовірність того, що латентний атрибут асоціюється з приватними особами, можна розрахувати на основі кількості та/або співвідношення його захищених сусідів. Нарешті, ці два підходи можна об'єднати в єдину модель, щоб ефективно фіксувати параметри конфіденційності [6].

Щоб проаналізувати зв'язок між атрибутами профілю та поведінкою конфіденційності, ми спочатку створили каталог користувачів Twitter, зібравши підписників кількох відомих акаунтів у Twitter (наприклад, «Facebook», «Katy Perry», «Obama»). Потім для кожного облікового запису ми обчислили відсоток захищених підписників до загальної кількості підписників. Результати показують, що цей відсоток значно вищий для «Крайні новини CNN» (11%) у порівнянні з іншими наборами підписників (від 5% до 7%) та середнім відсотком у Twitter (4,8%). Таким чином, ми вибрали цей набір для нашого дослідження,

оскільки дихотомія ставлення до конфіденційності та поведінки, здається, зведена до мінімуму.

Потім ми проаналізували цей набір користувачів, який включає збалансований набір із майже 1 мільйона користувачів, щоб отримати уявлення про потенційні відмінності в тому, як налаштовуються атрибути профілю загальнодоступних і захищених облікових записів. Ми зосередили свій аналіз на наборі функцій профілю, які легко доступні з облікових записів Twitter, а також додаткових функцій, розроблених на основі наявних атрибутів профілю. Наприклад, ми використали каталог англійських імен, щоб проаналізувати, чи оголошене ім’я містить справжнє ім’я людини. Крім того, ми дослідили лінгвістичні атрибути описів профілів за допомогою LIWC та аналізу частоти ключових слів.

В результаті розроблено набір ознак розміру, зведення якого наведено в таблиці. Перший стовпець таблиці показує функції Twitter API. У другому стовпці представлені лінгвістичні атрибути, отримані з описів профілів. На додаток до набору категорій LIWC2, цей список містить чотири функції на основі ключових слів, які вказують на наявність відповідного ключового слова в описі.

Атрибути Twitter	Лінгвістичні атрибути
Кількість твітів*	Слова з шістьма буквами
граф	Функціональні слова
Улюблений граф	Clout
Кількість списків	Емоційний тон
Фактична кількість імен	Автентичний
Гео-ввімкнено*	Аналітичне мислення
Має справжню назву	Вплив на слова
Ім'я користувача має ім'я	Соціальні процеси
Має URL-адресу	Когнітивні процеси
Має місцезнаходження*	Відносність
Має зображення за замовчуванням	Водії та потреби
Має профіль за замовчуванням*	«слідувати»

Відмінності між захищеними та публічними обліковими записами є статистично значущими для всіх цих функцій (визначено за результатами хі-квадрат або t-тест залежно від типу функції). Для чотирьох із цих функцій таблиця Функції профілю для виявлення захищених облікових записів.

Відмінності також є практично суттєвими (що визначають за V Крамера або d Коена залежно від типу ознаки). Ці чотири ознаки позначені зірочкою в таблиці. Таким чином, атрибути профілю є відмінними для захищених і загальнодоступних користувачів і виявилися корисними для нашого завдання визначення переваг конфіденційності в підході на основі сусідства. Використання цього набору функцій у контрольованому алгоритмі на основі регресії призвело до F-показу 0,72, що перевершує випадковий базовий рівень більш ніж на 20%.

1.4. Графік конфіденційності та приховані атрибути

Наш підхід спрямований на експлуатацію [8] і заснований на тому, що люди подібних інтересів мають тенденцію до спілкування один з одним. Тому замість того, щоб використовувати асиметричне відношення підписки чи друзів у Twitter, ми побудували ненаправлений взаємний граф користувачів, який включає лише ребра, які є взаємними. Очікується, що взаємні відносини вказують на міцніші стосунки між двома користувачами, і вони відрізняють розділ соціальної мережі сфери Twitter від її інформаційної мережі [9].

Починаючи з випадкового загальнодоступного користувача, ми ітераційно побудували взаємний графік користувачів методом пошуку в ширину (BFS). Для кожного загальнодоступного користувача ми спочатку підрахували кількість захищених спільних сусідів, а також відношення захищених до всіх спільних сусідів.

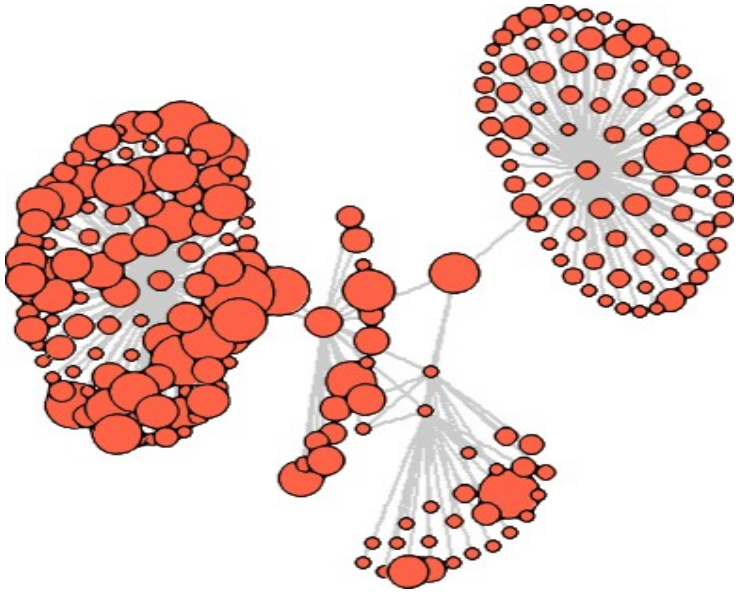


Рисунок 1.1 - Знімок графіка конфіденційності

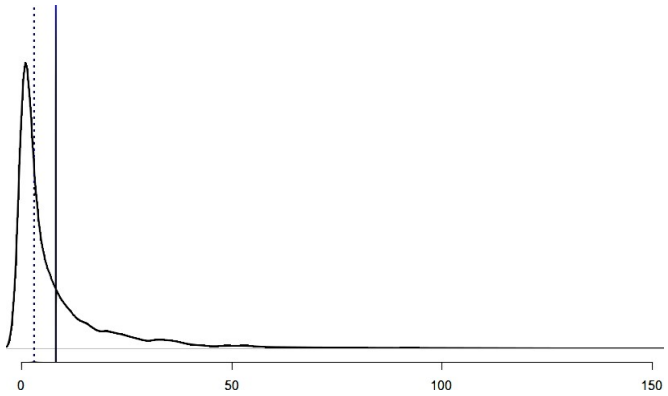


Рисунок 1.2 - Розподіл захищених контактів

Цей користувач потім анотується цими показниками та додається до графіка. Потім ми перевіряємо, чи має новий вузол взаємні зв'язки з будь-яким існуючим вузлом в графі, і додаємо відповідні ребра. Цей процес повторюється з новим відкритим користувачем, вилученим із черги BFS. Користувачі з менше ніж 10 твітів або менше ніж 30 підписників/друзів вважаються неактивними і, таким чином, не додаються до графіка. Крім того, не включаються перевірені користувачі та користувачі з понад 1000 підписників/друзів, оскільки вони часто представляють бренди та знаменитостей і не є представниками широкого загалу. Ми зібрали 3К загальнодоступних вузлів, які анотовані на основі їх показника коефіцієнта конфіденційності. На малюнку показано знімок невеликої частини

графіка, візуалізованого за допомогою примусово орієнтованого макета, де показник коефіцієнта конфіденційності зіставляється з розміром вузла. У цьому наборі даних кожен обліковий запис Twitter взаємно пов'язаний із в середньому 77 контактами. Серед цих сусідів у середньому 69 є державними та 8 захищеними. На рисунку 1.2 показано розподіл захищених сусідів, де середнє та медіана відзначені суцільною та пунктирною лінією відповідно. Як видно, розподіл перекошений вправо, що вказує на те, що, незважаючи на меншу кількість користувачів із великою кількістю захищених сусідів, ці числа є значно великими, тому середнє значення перетягується вправо.

Щоб переконатися, що застосовується в контексті конфіденційності, ми розрахували кореляцію між коефіцієнтом конфіденційності кожного вузла та середнім коефіцієнтом конфіденційності сусідів. Цей аналіз близько 700 користувачів із принаймні 10 взаємними контактами на графіку показав сильну позитивну кореляцію між двома змінними (коефіцієнт кореляції Пірсона r = 0,88). Цей результат вказує на те, що на поведінку користувачів щодо конфіденційності впливають або їхні близькі соціальні контакти, або люди з подібною поведінкою щодо конфіденційності, як правило, об'єднуються в соціальні мережі. У будь-якому випадку, цей висновок передбачає великий потенціал CF для прогнозування переваг конфіденційності.

Щоб виявити набір прихованих атрибутів, які цікавлять користувачів, які піклуються про конфіденційність, ми перетворили кожен вузол у набір атрибутів. Потім для кожного вузла атрибутів ми обчислили кількість і співвідношення захищених сусідів. Очікується, що отриманий графік дозволить обчислити $p(y_i|\theta_j)$, що вказує на ймовірність того, що користувач і буде приватним, якщо він/вона позначений атрибутом θ_j. Наприклад, таким значенням ймовірності може служити коефіцієнт конфіденційності кожного латентного фактора. Ми проводили експерименти, використовуючи уніграми та хештеги, витягнуті з твітів користувачів, як приховані змінні. Ці атрибути вибираються з 500 останніх твітів, опублікованих кожним користувачем. У

таблиці 1.2 наведено 10 найкращих хештегів та уніграм, які пов’язані щонайменше з 10% (n = 30) користувачів. Ці функції виділяються на основі показника коефіцієнта конфіденційності, який також наведено в таблиці.

Верхній хештег у таблиці — «#neverforget», який є пам’ятним політичним гаслом, який заохочує пам’ятати про національні та міжнародні трагедії. Високий коефіцієнт конфіденційності цього хештегу, а також найпопулярніші хештеги «#USA» та «#respect» можуть свідчити про те, що спільноти, зацікавлені в політичних темах, як правило, більш приватні. Цей висновок є цікавим, і його можна розглядати разом із високим відсотком захищених послідовників CNN. Однак, враховуючи те, що графік було зібрано за певний проміжок часу, ці результати можуть залежати від часу. Ще один орієнтований на час хештег у списку – «#aquarius». Збір даних протягом тривалого часу та групування хештегів у теми високого рівня може пролити світло на потенційні зв’язки між інтересами користувачів у політиці та знаках зодіаку та їхніми перевагами щодо конфіденційності.

Ми також дослідили зібрані твіти, щоб зрозуміти контекст, у якому використовується верхня уніграма «окремий». Більшість цих твітів пов’язані з любовним життям та стосунками користувачів. Цікаво, що «#relationshipgoals» і «#lastrelationshipnaughtme» мають високий рівень конфіденційності в нашому списку. Таким чином, можна зробити висновок, що користувачі, які розповідають про свої стосунки в соціальних мережах, швидше за все, піклуються про конфіденційність. Знову ж таки, групування хештегів у загальну концепцію та теми може дати більш точні результати.

1.5. Обчислювальний аналіз колективного інтелекту

Концепція інтелекту в людському мозку вже багато років активно досліджується в психологічних дослідженнях. За останні десятиліття в дослідженнях інтелекту відбувся зсув від розгляду пізнання лише як чогось у

мозку однієї людини до соціального та колективного інтелекту, що виникає внаслідок локальних взаємодій у групах індивідів [8]. Добре вивченою формою інтелекту в соціальних структурах є феномен «ройового інтелекту», який широко спостерігається в природі. Дослідження ройового інтелекту пояснюють, як групи тварин можуть виконувати надзвичайно складні й інтелектуальні завдання, які виходять далеко за межі когнітивних здібностей окремих тварин.

Наприклад, колонії мурашок давно захоплювали вчених, демонструючи комплексне вирішення проблем та інтелектуальні функції, такі як пошук найкоротшого шляху до джерела їжі або найкраще гніздо серед багатьох варіантів [7]. На додаток до колоній мурах, зграї птахів, випасання тварин, ріст бактерій та зграї риб є іншими прикладами самоорганізованих живих груп, які демонструють ознаки інтелекту [8].

Незважаючи на те, що колективний інтелект у людей та ройовий інтелект у тварин мають спільну ознаку «емерджент», вони відрізняються двома основними ознаками [6]. По-перше, рої тварин включають когнітивно простих агентів, тоді як колективний інтелект включає агентів зі складними когнітивними здібностями, тобто людей. По-друге, агенти кожного типу використовують різні методи взаємодії один з одним. Рої тварин зазвичай використовують взаємодії між особинами на основі стигмергії [6]. Стигмергія — це форма непрямої взаємодії, яка відбувається через подальші модифікації спільного середовища [4]. Наприклад, мурахи відкладають феромони, повертаючись до гнізда, коли знайшли джерело їжі. Потім феромон стимулює інших мурах і йдуть по слідах із сильнішим запахом. Оскільки мурахи проходять коротші шляхи за менший час, їх спільні дії, нарешті, призводять до пошуку найкоротшого шляху до їжі.

На відміну від роїв тварин, взаємодія в людських колективах може бути заснована на стигмергі (тобто непряма) або пряма. Крім того, взаємодія обох форм може відбуватися через два різні канали: вербальне спілкування або невербальні дії. Дотримуючись цих двох бінарних відмінностей, взаємодії між

людьми можна розділити на чотири різні групи: пряме вербальне спілкування, пряма невербальна дія, непряма вербальна комунікація та непряма невербальна дія. Обговорення та обговорення, пішоходи, які утворюють двоспрямовані смуги на переповненому тротуарі [4], спільне авторство та пошук оптимізованого маршруту через активне формування стежки [5] є традиційними прикладами кожної форми відповідно.

Тим часом прогрес у комп'ютерно-опосередкованій комунікації та поява Web 2.0 уможливили масовий обмін даними у глобальному масштабі, що призвело до нових і передових форм колективного інтелекту. Вікіпедія, Google, форуми запитань і відповідей та соціальні мережі – все це приклади веб-платформ колективної розвідки. Наприклад, Вікіпедія використовує словесні взаємодії на основі стигмергії, коли користувачі можуть змінювати локальні частини свого спільного віртуального середовища. У проектах з відкритим кодом дії учасників накладаються один на одного для створення складних продуктів, таких як операційна система Linux. З іншого боку, у таких середовищах, як форуми питань і відповідей, соціальні мережі та системи обговорення, колективний інтелект може бути викликаний прямим спілкуванням та онлайн-діалогом між користувачами. Малюнок ілюструє різні типи взаємодії в колективному інтелекті разом із відповідними прикладами на основі Інтернету.

Платформи колективного інтелекту з підтримкою Інтернет дозволяють натовпу збиратися і вносити свій внесок у вирішення проблем, які може бути важко або неможливо вирішити навіть найрозумнішим людям і найшвидшим комп'ютерам. Різні аналітичні методи розроблені, щоб зрозуміти, як інтелект виникає в соціальних взаємодіях, і визначити різні фактори мікро- та макрорівня, які можуть позитивно чи негативно впливати на феномен колективного інтелекту.

Отже, основним напрямком цієї роботи є виконання емпіричного аналізу різноманітних і великомасштабних наборів даних, створених користувачами, а також розробка обчислювальних інструментів і моделей для кращого розуміння

колективного інтелекту. Отримана аналітика буде потім використовуватися разом із візуалізацією інформації та принципами взаємодії людини з комп'ютером для проектування та розробляти веб-інструменти, які ефективно використовують інтелектуальну силу людських колективів. Це дослідження в основному зосереджено на даних, орієнтованих на онлайн-дискурс, які включають принаймні деякі прямі взаємозв'язки між учасниками (лівий верхній квадрант на рисунку 1.3). Незважаючи на те, що обчислювальні інструменти були розроблені для вилучення різних аспектів інтелекту з колекцій оглядів продуктів і фільмів, мікроблогів і заголовків новин, вони є незалежними творами, створеними окремими особами, і тому виключені з цього дослідження.

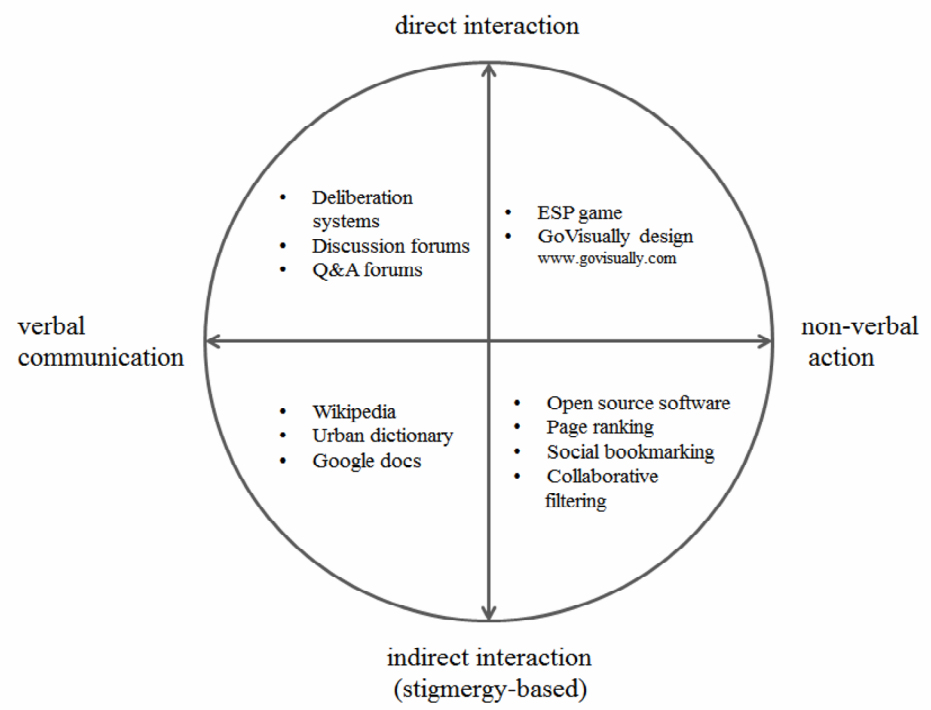


Рисунок 1.3 - Колективний інтелект можна розділити на чотири різні групи на основі типів взаємодії між людьми

Щоб визначити програму досліджень у напрямку розвитку та аналізу колективного інтелекту в Інтернеті, ми переглянули існуючі обчислювальні підходи, які використовували автоматизовані методи для вивчення

великомасштабних веб-платформ колективного інтелекту. Ми спеціально зосередилися на колективному інтелекті, який виникає внаслідок вербальних взаємодій. Тому наш огляд виключив платформи для онлайн-оглядів, мікроблогів та заголовків новин, оскільки ці текстові блоки є незалежними творами, створеними окремими особами. Крім того, ми зосереджуємось на потенційно повторюваних взаємодіях у обмежених середовищах, тому дослідження про Інтернет та соціальні мережі загалом, такі як [4], не включені. Інструменти чату та електронної пошти також виключені, оскільки вони в першу чергу призначені для підтримки індивідуальних і невеликих взаємодій і зазвичай не розглядаються як великомасштабні платформи.

У наступному ми описуємо структуру групового дослідження, яка використовується для класифікації існуючої літератури. Потім ми надаємо більш детальні пояснення досліджень у кожній категорії.

1.6. Методологія системного огляду

Хоча існують різні моделі обробки малих груп, ми не можемо знайти модель, яка б специфічна для масштабна групова діяльність. Макграт визначає групи як «соціальні агрегати, які передбачають взаємне усвідомлення та потенційну взаємну взаємодію». Спочатку призначена для вирішення відносно невеликих і структурованих груп, модель Макгратса – це не просто модель чи теорія діяльності малих груп, а основа для систематичного вивчення груп, враховуючи всі фактори, які потенційно впливають на процес і результат групової діяльності, і те, як вони взаємодія. Ми припускаємо, що ці фактори також існують у більших онлайн-колективах і мають подібні структури взаємодії, як описано в моделі Макгратса. Тому ми обрали цю модель для нашого систематичного огляду та застосували її до дослідження більших онлайн-колективів, які передбачають взаємну взаємодію між людьми.

Спрощена версія фреймворку McGrath показана на рисунку. У цьому контексті процес групової взаємодії розглядається як центральна ознака та сутність групи, і він відноситься до «шаблонних відносин між поведінкою індивідів» [4]. На додаток до цього основного компонента, структура охоплює кілька елементів, які якимось чином можуть бути пов’язані з процесами групової взаємодії. Наприклад, учасники впливають і можуть на них впливати процеси взаємодії; таким чином, властивості члена включені в структуру. Зразок стосунків між учасниками є важливим фактором у розумінні взаємодій всередині індивідів, і тому він виглядає як постійна група чи групова структура в рамках. Ще два важливі аспекти включають середовище, де відбувається групова взаємодія, і завдання (тобто неформально передбачувані цілі), які переслідують члени групи.

На додаток до загальної рамки групових досліджень, Макграт описує процеси взаємодії на мікрорівні, розглядаючи їх у термінах трьох різних етапів. Перший етап розглядається зі структурою і розподілом комунікацій між взаємодіючими людьми, який називається комунікаційним шаблоном. На другому етапі кожне спілкування розглядається з огляду на його зміст. Сам зміст потім розглядається з точки зору двох різних аспектів: міжособистісних стосунків і виконання завдань.

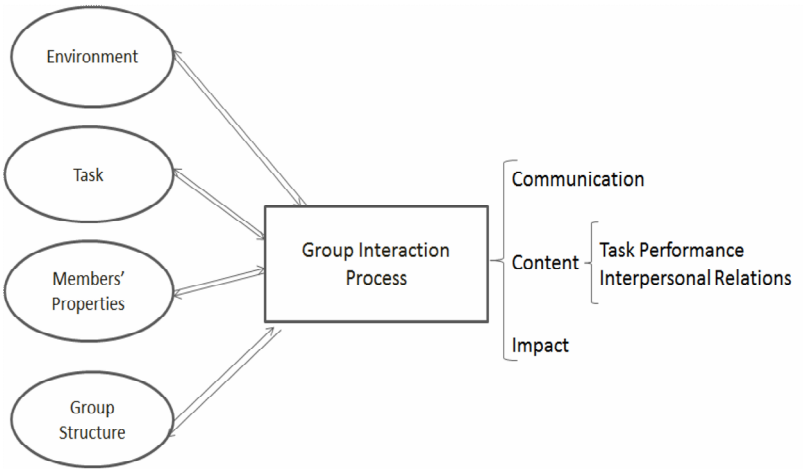


Рисунок 1.4 - Структура Макграта може використовуватися для класифікації літератури про онлайн-колективи.

Даний етап пов'язаний з впливом різних процесів групової взаємодії один на одного і на учасників. Сучасний напрям дослідження аналізу колективного інтелекту зазвичай розглядає властивості членів (наприклад, стать), структуру групи (наприклад, ієрархічну), середовище (наприклад, дискусійний форум) і завдання (наприклад, прийняття рішень) як контекст і незалежні змінні дослідження, а потім полегшує різні обчислювальні методи для вивчення компонента процесу групової взаємодії. Тому пов’язані роботи з аналізу колективного інтелекту поділяються на три основні групи: аналіз моделі комунікації, аналіз контенту та аналіз впливу. Контент-аналіз потім поділяється на дві категорії досліджень. Через різну природу віртуальних просторів компонент міжособистісних стосунків розглядається на вищому рівні, зосереджено на дослідженнях, які аналізують емоційні та сентиментальні аспекти змісту. Крім того, компонент виконання завдання аналізу вмісту називається виконанням завдання/цілі, оскільки такі параметри можуть бути менш орієнтованими на завдання.

Оскільки цей огляд зосереджений на обчислювальних методах та алгоритмах для аналізу колективного інтелекту, ми почали процес відбору статті, провівши кілька пошуків за ключовими словами в цифровій бібліотеці ACM [2]. Запити були сформовані шляхом запозичення ключових слів із системи групових досліджень McGrath, таких як «аналіз моделі спілкування» та «аналіз настроїв». Початковий аналіз 100 найкращих документів, отриманих за кожним запитом виконується шляхом читання заголовків, рефератів та вступів, щоб відфільтрувати набір, щоб включати лише ті статті, які входять до сфери цього дослідження. Таким чином, щоб визначити відповідні статті, ми використовували наступні критерії для кожного дослідження: 1) запропоновані автоматизовані підходи з мінімальним кодуванням і зусиллями людини та 2) провели аналіз на великомасштабних веб-платформах, які передбачають обмін інформацією та діалог. серед кількох учасників.

Після того, як початковий набір документів був обраний, для охоплення

відповідних статей в інших місцях використовувалися методи зворотного та прямого ланцюга. Метою було не охоплення кожного окремого дослідження, а надання репрезентативної вибірки статей, щоб отримати достатнє розуміння тем дослідження та прогалин у сучасному стані літератури.

Шаблон спілкування відноситься до форми або структури серії інтерактивних форм поведінки або спілкування, які мають місце серед людей [4]. Недавнє дослідження Woodley et al. показав, що моделі повороту учасників співвідносяться з колективним інтелектом групи. Цей висновок свідчить про важливість аналізу комунікаційних моделей для розуміння та розвитку інтелекту між людськими колективами.

Незважаючи на те, що попередні роботи в таких контекстах спиралися на основні статистичні показники, більш складні обчислювальні методи аналізу комунікаційних моделей дотримувалися двох основних напрямків: аналіз соціальних мереж та аналіз ниток розмов як деревоподібних структур. Крім того, замість того, щоб зосередитися на моделях групового рівня, ряд досліджень досліджував моделі спілкування як поведінку індивіда.

Велика кількість досліджень намагається створити соціальну мережу учасників, а потім використовувати різні методи аналізу графіків, запропоновані аналітиками соціальних мереж, щоб краще зрозуміти основні комунікаційні моделі. У соціальних мережах кожному учаснику відповідає вузол і V на графіку G = hV, Ei, де ребро (i, j) E на графіку вказує на якийсь соціальний зв’язок між двома учасниками і та j. При вивченні моделей спілкування краї зазвичай представляють повідомлення та онлайн-діалоги, якими обмінюються учасники. Різні типи соціальних мереж можуть бути сформовані на основі різних інтерпретацій того, як повинні бути пов’язані два учасники, наприклад, чи мають бути спрямовані краї, на основі відправника та одержувача повідомлення, чи краї мають бути зважені на основі кількості повідомлень. обмінялися.

Наприклад, Gomez et al. [2] досліджували три різні типи соціальних мереж, отримані з відповідей на коментарі на веб-сайті обміну новинами

Slashdot, які відповідають різним стратегіям визначення межі між парою вузлів. Далі вони проаналізували різні атрибути цих мереж, такі як максимальна кількість кластерів, середня довжина шляху, максимальна відстань між двома учасниками, взаємність, розподіл ступенів та асортативне змішування за ступенем і балом. Їхні висновки означають, що, незважаючи на те, що Slashdot поділяє деякі особливості моделі спілкування, такі як невелика світова власність і висока кластеризація, з традиційними соціальними мережами, її учасники продемонстрували нейтральне змішування за ступенем, ідентичне розподілення та просто помірну взаємність, що спостерігається рідко. в інших соціальних мережах.

В іншому дослідженні аналізуються моделі взаємодії на сторінках обговорення Вікіпедії [3], де виділяються три різні мережі відповідно до різних типів взаємодій: мережа відповідей на статті, яка включає прямі відповіді на сторінках обговорення статей; мережа розмов користувачів, яка включає прямі відповіді на сторінках обговорення користувачів; та настінна мережа, яка містить особисті повідомлення, розміщені на сторінках обговорень іншого користувача. Ці три мережі порівнюються за допомогою метрики перекриття країв. Результати показують відносно велике перекриття між двома мережами, витягнутими зі сторінок обговорення, в той час як мережа з особистих комунікацій було виявлено, що істотно відрізняється від двох інших. Крім того, аналіз спрямованих асортативно за ступенем показав, що користувачі, які відповідають багатьом іншим користувачам, як правило, відповідають недосвідченим користувачам, тоді як ті, хто отримує коментарі від багатьох користувачів, частіше взаємодіють один з одним.

Замість того, щоб формувати соціальні мережі з розмов, деякі дослідники досліджували комунікаційні моделі як інформаційні каскади або потоки, де дія відбувається внаслідок впливу діяльності інших [8]. У цих дослідженнях каскади моделюються як дерево T = hV, Ei, де кожен вузол i V представляє дію

учасника, а між і і j є ребро (i, j) E, якщо одне з цих дії є відповіддю на іншого або під впливом іншого.

У дослідженні [2] потоки обговорень відображаються на радіальні дерева, де повідомлення є центральним вузлом, а його коментарі додаються на різних рівнях вкладення залежно від їх рівня до центрального повідомлення. Аналізуючи коефіцієнт розгалуження b, тобто кількість коментарів, викликаних даним коментарем, вони виявили високий рівень неоднорідності в дискусійних ланцюгах. Крім того, результати аналізу вказують на різницю в процесі генерації коментарів початкового рівня та коментарів наступних рівнів. Крім того, метрика h-індексу, що використовується в науковій спільноті [2], прийнята для визначення рівня дискусії.

У [3] проаналізовано структуру ланцюжкових розмов у трьох різних наборах даних, а саме групах Usenet, Yahoo! Групи та Twitter. Досліджуються взаємозв'язки між розміром і глибиною, розподілом ступенів і рівнем потоку, а також розміром і кількістю авторів у потоках. На основі базового процесу розгалуження також розробляється математична модель. У процесі розгалуження кожен вузол генерує k дітей з ймовірністю p(k), де p — функція розподілу ймовірностей. Беручи до уваги початковий аналіз, автори включили в свою модель ступені вузлів, їх сучасність та ідентичність автора.

У роботі [2] проаналізували чотири різні набори соціальних даних і показали, що модель преференційної прихильності з ухилом до початкового повідомлення здатна охопити більшість структурних властивостей, а також моделі еволюції каскадів розмов. У моделі переваги приєднання ймовірність того, що існуючий вузол буде пов'язаний з новим вузлом, пропорційна його розподілу ступенів, що означає, що вузли з вищими ступенями, швидше за все, будуть пов'язані з новими. Для забезпечення більш повної моделі в [3] використовується інформація про останній час разом із упередженням щодо початкового повідомлення та факторів розподілу ступеня.

В [3] вивчали форму та розмір дерев обговорення на рівні статті у Вікіпедії та запропонували різні показники для характеристики сторінок обговорення, такі як ланцюжки прямих відповідей, максимальна глибина та h-індекс дерева. Після класифікації статей Вікіпедії та проведення їх аналізу на різні теми вони виявили значну різницю в структурах обговорення з різних семантичних областей.

Хоча вищезгадані дослідження були зосереджені на моделях спілкування на груповому рівні, були зроблені деякі спроби охарактеризувати комунікативну поведінку індивідів у широкомасштабних соціальних умовах. Наприклад, проведено велику кількість досліджень, присвячених моделюванню та аналізу часових інтервалів комунікативної діяльності індивіда [9]. Крім того, его-мережі використовувалися для розуміння комунікаційних моделей індивідів, таких як ідентифікація «осіб, які обговорюють» та «відповідаючих» бо ідентифікація лідерів [2]. Окрім мереж Адамік та ін. [3] запровадив показник ентропії, який фіксує концентрацію шаблонів відповідей учасника на різні теми на форумі питань і відповідей.

Була проведена велика робота з аналізу настроїв, дослідники досліджували різні методи для визначення суб'єктивних речень, оцінки рівня настроїв за двійковою (наприклад, позитивний/негативний) або порядковою шкалою (наприклад, від 1 до 5), щоб виявляти настрої та емоції (наприклад, гнівні чи радісні) і, нарешті, зрозуміти джерело, ціль та складні типи ставлення [5]. Позначаючи ту саму область дослідження, використовується багато інших термінів, серед яких найбільш поширеними є аналіз суб'єктивності та вивчення думки.

Для автоматичного аналізу настроїв було використано два основні методи. Методи, засновані на лексиці, спиралися на анотовані словники, які включають полярність настроїв та оцінки, щоб визначити настрої текстової одиниці на основі слів і фраз, які вона містить. Замість використання анотованих словників, підходи до навчання з учителем створюють класифікатори,

використовуючи мічені дані, щоб передбачити настрої речень, фраз або документів. Різні форми та комбінації цих методів використовувалися та вивчалися в загальних текстах та онлайн-оглядах, тоді як онлайн-дискусіям на соціальних платформах приділялося відносно менше уваги.

У попередніх роботах, присвячених методам на основі лексики, використовувалися як словники, створені людьми-анотаторами, так і автоматично створені, створені за допомогою методів навчання без нагляду [5]. Традиційний підхід до використання таких словників полягає у визначенні оцінки настрою текстової одиниці шляхом усереднення значень настрою окремих слів, витягнутих зі словника.

Через важливість контексту в аналізі настроїв, пізніші підходи вийшли за межі заперечень і врахували структуру дискурсу та інші контекстуальні особливості в процесі обчислення настроїв. Наприклад, в [3] використовували китайський словник, що включає списки позитивних і негативних слів, приватів і сентиментально зважених модифікаторів, щоб призначити рівень настроїв для постів для обговорення. Для кожного ключового слова в даній публікації рейтинг настроїв спочатку оцінюється на основі словника. Потім алгоритм перевіряє, чи присутній привілей у словах перед ключовим словом. Якщо це так, оцінка сентименту слова заперечується; інакше алгоритм переходить до останнього кроку. На цьому кроці, якщо в словах, що оточують ключове слово, присутній модифікатор, рейтинг настрою ключового слова множиться на вагу модифікатора, витягнутий зі словника. Результати аналізу настроїв потім включають у вектори ознак для кластеризації k-середніх, за допомогою яких можна виявити гарячі точки. Крім того, класифікатор Support Vector Machine (SVM) навчається на векторах ознак настроїв і результатах k-середніх для прогнозування наступного гарячого форуму.

Зростання доступності позначених екземплярів в Інтернеті разом із появою платформ краудсорінгу (наприклад, Amazon Mechanical Turk), де мічені дані можна ефективно та ефективно збирати, призвело до використання

різноманітних контрольованих методів класифікації в дослідженні настроїв. Набори функцій, які зазвичай використовуються в техніках вивчення настроїв, включають присутність і частоту термінів, позицію терміна, частину мови (POS), синтаксис і заперечення. Наприклад, Розенталь і МакКевон адаптували існуючий метод [4], який використовує лексичні оцінки слів, аналіз N-грам і полярність оточуючих синтаксичних складових, до даних соціальних мереж, беручи до уваги емоції, акроніми та орфографічні помилки.

Хоча методи навчання під наглядом були доведені як корисні та точні в завданнях класифікації настроїв, коли вони навчаються в іншій області, ніж область інтересу, їх ефективність значно падає [5]. У [4] незалежний від домену алгоритм [3], який називається аргументацією через пошук, використовується для класифікації настрою (тобто сердитий, сумний, радісний) коментарів новин на Yahoo!Buzz на основі позначених екземплярів, отриманих з LiveJournal. набір даних коментарів.

Щоб автоматично виявляти речення зі ставленням до інших учасників онлайн-дискусій [4] застосував дещо інший підхід, ніж основні дослідження аналізу настроїв. У їхньому підході мовна модель будується з використанням позначених даних. Однак замість того, щоб присвоювати мітки текстовим одиницям, мовна модель присвоює значення ймовірності [5]. Найрелевантнішою частиною речень зі ставленням є те, де включається займенник другої особи. Таким чином, цей фрагмент спочатку витягується. Потім за допомогою лексичних елементів, тегів POS, тегів полярності слів і відносин залежностей витягується набір шаблонів, пов’язаних з кожним фрагментом. Шаблони потім використовуються для побудови моделей Маркової для речень з ставленням і без нього. Тоді нове речення можна позначити як речення з ставленням або без нього.

Цінність груп полягає в їхній потенційній здатності ефективно виконувати широкий спектр завдань. Тому було проведено широке класичне дослідження соціальної психології за допомогою польових досліджень,

опитувань та лабораторних експериментів із застосуванням ручних методів для дослідження того, як і в яких умовах різні групи можуть бути ефективними [5]. Проте останніми роками поширення соціальних платформ на основі Інтернету вимагало автоматизованих підходів до обробки та оцінки великомасштабного текстового вмісту, створеного натовпом. Низка досліджень, присвячених онлайн-соціальним середовищам, намагається вирішити проблему перевантаження інформацією шляхом організації та опису особливостей основного інформаційного простору. Такі дослідження можна розглядати як попередні кроки, які можуть привести до легшого та більш точного аналізу виконання завдань машинами або користувачами. Інший набір робіт, з іншого боку, розробив нові обчислювальні методи для безпосередньої оцінки якості та релевантності розмовного тексту в Інтернеті.

Одне з ранніх досліджень з опису онлайн-діалогів було проведено Саком [9], в якому основні методи обробки природної мови використовуються для вилучення та представлення списку найбільш часто використовуваних термінів у дискусіях разом із їх семантичною мережею. Методи кластеризації [5] також були використані для опису змістових особливостей комунікаційного простору. Кластеризація відноситься до групування об’єктів на основі деяких показників подібності. Загалом, методи кластеризації тексту кодують вміст документа у відповідних структурах даних, таких як векторна модель простору, а потім використовують різні метрики відстані для обчислення міри подібності та формування кластерів документів. Основний принцип, що лежить в основі методів кластеризації на основі вмісту, полягає в тому, що чим більше ключових слів містить документи, тим більше вони схожі.

Кластеризація документів – добре вивчена тема; однак відсутність структури та редакційного контролю в розмовному тексті ставить нові виклики та складнощі. Таким чином, прийняття та розширення методів, що використовуються для вирішення загальних текстів, часто пропонуються для групування розмовного тексту. Наприклад, в [5] запропонували метод

кластеризації повідомлень на основі лідера (LPC), модифікацію алгоритму Leader [6], для кластерування повідомлень у ланцюжках обговорень. Їх підхід починається з головного стовпа нитки як першого ведучого. Використовуючи запропоновану ними метрику відстані, яка включає функції як вмісту, так і моделі комунікації (тобто тегування між дописами), розраховується схожість публікації з наявними кластерами. Якщо результуюча міра нижча за певний поріг, посада відноситься до відповідного кластеру; інакше він стає новим лідером. Цей процес продовжується до тих пір, поки всі дописи не будуть об'єднані в групу, що призведе до виявлення публікацій, які не відповідають темі, і дописів, що виходять за межі.

Близько пов'язаним підходом до кластеризації документів є тематичне моделювання, коли дослідники використовували різноманітні методи видобутку під наглядом і без нагляду для сегментації, позначення та відстеження різних тем [7]. Незалежно від проблем моделювання теми для розмовного тексту, деякі вчені застосовували нові методи в невеликих дискусіях; однак, подібно до кластеризації, тематичне моделювання менш досліджується для масштабних веб-розмов. Один приклад наведено Zhue et al. [7], який розширює основний алгоритм виявлення та відстеження тем (TDT), щоб відповідати цьому методу для даних, згенерованих в онлайн-спільнотах.

Замість кластеризації та тематичного моделювання, інший напрямок роботи з описового лінгвістичного аналізу онлайн-дискусій зосереджено на завданні пошуку та маркування висловлювань та одиниць тексту, які є цінними в аналізі ефективності, наприклад, ідентифікації претензій [8, 4], угоди чи розбіжності [1], обґрунтування чи аргументи [9] та ідеї [11]. Підходи, що використовуються в цьому напрямі дослідження, мають певну спільну мову з дослідженнями аналізу настроїв, оскільки обидва стосуються зіставлення фрагмента тексту з міткою з попередньо визначеного списку. Крім того, деякі підходи в цьому напрямку використовують суб'єктивний аналіз для визначення

певних ознак (наприклад, претензій та угод/розбіжностей). Таким чином, обчислення настроїв іноді включається в підходи.

У роботі [8], претензії автоматично виявляються за допомогою регресійного аналізу за різними ознаками, включаючи настрої, слова віри, POS та N-грами. Вони виявили, що настрої та POS-теги були важливішими в LiveJournal, тоді як віра та настрої були важливішими для форумів Вікіпедії. Ця відмінність свідчить про наявність відмінних мовних характеристик у різних соціальних контекстах.

Враховуючи пару речень, перше з яких є твердженням [9] розробили алгоритм машинного навчання, щоб автоматично визначити, чи є друге речення виправданням претензії. Їхній підхід витягує список показників відношень теорії риторичної структури (RST), які вони вважали відповідними для виконання завдання. Для кожного показника список пар слів, що зустрічаються разом, витягується з англійської Вікіпедії та становить основу функцій, які вони використовували у своєму підході до навчання під керівництвом.

У [1] розроблено і розроблено два інструменти визначення сенсу для систем управління ідеями. Перший інструмент, який називається ідеєю, використовує теорію мовленнєвих актів, щоб автоматично позначати ядро ідеї в неструктурованій текстовій інформації. Другий інструмент, званий інтерпретатором коментарів, шукає конкретні мовні характеристики в коментарях, створених для ідеї, а потім призначає коментар до однієї з трьох попередньо визначених категорій коментарів: реакції, пов’язані зі змістом, значенням або статусом ідеї. Використовуючи ці мітки, система проактивно пропонує реакції на ці коментарі на основі категорії, до якої він належить.

Інший напрям дослідження аналізу виконання завдання/цілі використовує обчислювальні методи для безпосередньої оцінки контенту, створеного користувачами. Наприклад, щоб підрахувати релевантність коментарів до фактичного повідомлення, [5] використовує TF-IDF, у якому коментарі розглядаються як запити, а дописи розглядаються як документи. Оскільки вони

застосовують свою техніку на сайті коментування новин, корпус новин Reuters використовується як базова колекція для обчислення показника IDF. Аналогічно, [4] представляє новий метод виявлення повідомлень не по темі на форумах. У цій техніці набір термінів витягується за допомогою методу інформативності слів BLTR для представлення кожного потоку. Кожна публікація також представлена набором термінів. Міра подібності цих двох векторів поєднується з трьома іншими мірами, які обчислюють подібність публікації з головною, з попередніми повідомленнями та з усіма попередніми повідомленнями, щоб обчислити релевантність кожного повідомлення для ланцюжка.

Запозичуючи алгоритм HITS з інформаційно-пошукових досліджень [9], Feng et al. [7] запропонував метод на основі графіків для виявлення найважливіших дописів у ланцюжковому обговоренні. Щоб сформувати графік із розмовного тексту, кожне повідомлення представлено вузлом на графіку. Запропоновано метод генерування посилань на основі ознак для розміщення зв'язків між парою вузлів на основі трьох ознак лексичної подібності, достовірності плакату та відносин мовленнєвого акту. Потім для кожного вузла ітераційно обчислюються два бали концентратора (тобто якість вузла як вказівника на інші) і авторитет (тобто якість вузла як ресурсу), доки алгоритм не збіжиться. В [6] застосував класифікатор SVM, заснований на різних внутрішніх характеристиках, таких як лексична, синтаксична та схожість вмісту, щоб оцінити якість публікацій (тобто хороших чи поганих) на дискусійних форумах.

Аналіз впливу має на меті зрозуміти потенційний вплив процесів групової взаємодії, тобто моделей спілкування, процесу виконання та сентиментальних аспектів, один на одного. Крім того, дуже цікавим є розуміння того, як ці процеси впливають на поведінку учасників, як-от їхнє ставлення, сприйняття, судження та навчання. Використовуючи обчислювальні методи, було дотримано дві різні процедури для вивчення впливу в онлайн-середовищі. Низка робіт була зосереджена на застосуванні існуючих методів для незалежного обчислення змінних, а потім використання основних статистичних

методів для оцінки їхньої кореляції. Кілька останніх підходів використовували більш складні методи для безпосередньої оцінки потенційного впливу однієї змінної на іншу.

Аналізуючи сайт коментування новин [5] дослідили зв’язок між актуальністю, тимчасовістю, настроєм і якістю коментарів новин. Під час аналізу актуальності були визначені теми, які викликали негатив і, таким чином, потребували більшої модерації. Була виявлена кореляція між негативністю та кількістю видалених коментарів, що свідчить про те, що настрої можуть бути показником якості коментарів. Крім того, у своєму часовому аналізі вони виявили позитивну кореляцію між частотою коментарів та негативністю коментарів користувачів, що також свідчить про потенційний зв’язок між настроями та якістю коментарів. У відповідному дослідженні на форумах BBC [10] досліджується зв’язок між активністю учасників та їхніми настроями [10], виявляючи, що негатив підвищує активність учасників, а більш активні користувачі схильні висловлювати негативні емоції. В [10] провели відносно складний аналіз онлайн-повідомлень на різних платформах, щоб зрозуміти вплив поточних емоцій та настроїв, висловлених в онлайн-спільноті, на емоції наступних публікацій. У їхньому підході ланцюжки дописів спочатку об’єднуються в групи послідовних дописів з однаковим значенням настроїв (тобто позитивні, нейтральні чи негативні). Порівняння сформованих кластерів з кластерами, сформованими на основі випадкових даних, показало значну різницю, причому кластери фактичних даних мають більший розмір. Їх статистичний аналіз показує, що умовна ймовірність для послідовних постів зростає як степеневий закон із розміром кластера, що схоже на преференційні процеси.

1.7. Аналіз структури інтенцій

Крім того, вони дослідили зв’язок між розміром потоку та емоціями, вираженими в ланцюжку, показавши, що коротші нитки, як правило,

починаються з меншого негативного настрою, а довші нитки мають більші варіації настроїв.

Щоб виявити прогалини в дослідженні з наявної літератури, ми розглянули основне та низькорівневе призначення робіт, використані обчислювальні методи, середовище їхнього фокусування та вторинні переваги, які вони можуть надати

технологіям Web 2.0.

Незважаючи на цінність і можливості, які полягають у середовищі, орієнтованому на завдання, більшість попередніх робіт були зосереджені на середовищах, в першу чергу, призначених для полегшення соціальних взаємодій (наприклад, форуми для дискусій та набори коментарів у новинах/медіа). Було докладено небагато зусиль, щоб зрозуміти орієнтовані на завдання онлайн-середовища, такі як інструменти обговорення та системи управління ідеями. Можливо, було багато досліджень щодо зручності використання та користувацького досвіду (UX), які досліджували вплив середовища за допомогою традиційних методів UX. Новим і недостатньо вивченим напрямком у літературі є розробка обчислювальних методів для вивчення цих середовищ шляхом виявлення та аналізу «слідів» колективного інтелекту. Очікується, що на «сліди» впливають різні фактори, такі як характеристики учасників, види розвідувальних завдань та дизайн середовища.

Відсутність складних автоматизованих методів аналізу впливу також видно з таблиці. У той час як велика група робіт була зосереджена на застосуванні існуючих методів для незалежного обчислення змінних і використанні основних статистичних методів для оцінки їх кореляції; лише кілька останніх підходів використовували більш складні методи для безпосередньої оцінки потенційного впливу одного процесу взаємодії на інший або на учасників. Методи навчання під наглядом і без нагляду можуть бути цінними та ефективними інструментами для моделювання та потенційного прогнозування таких впливів.

Крім того, як видно з таблиці, були зроблені деякі спроби проаналізувати вплив процесів взаємодії, тобто моделі спілкування, виконання завдання/мети та настроїв, один на одного. Однак менше уваги було приділено розумінню того, як ці процеси впливають на індивідуальну та колективну поведінку учасників з часом, наприклад на їхнє сприйняття, навчання та судження.

Приховані фактори, отримані з графіка конфіденційності, показало, що цей хештег в основному використовується в розмовах, де користувачів просять надати деяку інформацію, але вони відмовляються робити це, відповідаючи на твіт, який містить «#personal». Отже, цілком імовірно, що це люди, які знають конфіденційність і навмисно використовують публічне середовище, що відповідає тому, що вказує ймовірність моделі латентного фактора. Загалом, незважаючи на обмежений розмір графіка, результати є розумними і можуть виявити цікаву інформацію про приватні квартали в Twitter.

Щоб передбачити переваги конфіденційності, ми запропонували метод CF, який використовує як моделі сусідства, так і моделі латентного фактора. Ми проаналізували переваги використання атрибутів профілю для вимірювання схожості користувачів і використання хештегів та уніграм як прихованих функцій. Наразі процес збору даних триває, і ми проведемо подібні дослідження на кількох графіках, створених із використанням різних початкових користувачів. Ми також розглянемо ряд інших атрибутів користувача для моделі латентного фактора. Для перевірки корисності підходу будуть розроблені надійні методи оцінки. Хоча тут ми зосереджені на спрощеній формі конфіденційності (тобто бінарна специфікація), будуть зроблені спроби проаналізувати складні форми та стратегії захисту конфіденційності.

Щоб отримати уявлення про те, як інтелект виникає внаслідок онлайн-взаємодій, і зрозуміти різні фактори, які можуть вплинути на це явище, необхідні подальші дослідження, щоб заповнити ці прогалини. Ми вважаємо, що поєднання обчислювальних методів, які аналізують сліди та зміст колективного інтелекту, і технік візуалізації, які виявляють актуальні та тимчасові

закономірності з часом, є одним із напрямків дослідження, щоб усунути цю прогалину. Однак, незважаючи на те, що були зроблені деякі спроби надати визначення колективного інтелекту [6], відсутність узгодженого оперативного визначення для цього поняття призвело до відсутності емпіричних доказів для вивчення потенційного впливу кожного з них. ці фактори на колективний інтелект. Таким чином, пропонування добре обґрунтованих методів вимірювання колективного інтелекту може суттєво вплинути на цю сферу.

Людські суспільства завжди страждали від «злих» проблем, таких як зміна клімату, природні небезпеки та охорона здоров'я, і мали справу з ними. Дозволяючи широкомасштабне обговорення та обговорення, соціальні платформи Web 2.0 виходять за межі можливостей вирішення проблем невеликої групи органів влади і можуть ефективно використовувати колективну силу унікальних індивідів у безпрецедентних масштабах. Таким чином, дуже цінним і цікавим є виявлення різних моделей діяльності на основі дискурсу, які відповідають інтелекту.

Багатообіцяючим є поширена література, що аналізує колективний інтелект, заснований на діалозі. Ми провели систематичний огляд попередніх досліджень, які пропонували обчислювальні методи для аналізу колективного інтелекту у великомасштабному тексті дискурсу, створеному користувачами. Наш аналіз виявляє низку дослідницьких прогалин у цій галузі, включаючи відсутність зосередженості на середовищі, орієнтованому на завдання, відсутність складних методів для аналізу впливу процесу групової взаємодії один на одного, а також відсутність зосередженості на вивченні вплив процесів групової взаємодії на учасників у часі. Ми закликаємо до дослідницької діяльності, щоб усунути ці прогалини, і ми віримо, що така робота може сприяти покращенню нашого розуміння колективної поведінки людей, сприянню розвитку навичок колективної роботи та наданню цінного уявлення про розробку систем колективного інтелекту. Наші повсякденні соціальні взаємодії часто включають складний процес переконання та впливу. Такі взаємодії варіюються

від неформальних обстановок, таких як розмови в коридорі, до офіційних ситуацій, як-от ділові переговори та президентські виборчі кампанії. Тому вміння переконувати є ключовою навичкою, якою потрібно володіти, і вона може суттєво сприяти нашому соціальному житті. Починаючи з часів Аристотеля, у таких традиційних комунікаціях було проведено багато досліджень щодо переконання та переконливого дискурсу [3]. Однак із появою, постійним розвитком і зростанням веб-сайтів соціальних мереж, значна частина нашої соціальної взаємодії та взаємодії відбувається через нові та передові форми спілкування. Згенеровані комунікаційні дані, переважно у формі текстових повідомлень та коментарів, забезпечують чудовий ресурс для розуміння механізмів онлайн-переконання та його подібності та відмінності від традиційних процесів переконання. Проте такі повідомлення не були достатньо використані для вивчення переконливих дій в середовищі онлайнового спілкування, в основному через відсутність позначених даних.

Нішева спільнота субредітів пропонує значний ресурс для вивчення онлайн-переконання, сприяння зміні думок за допомогою онлайн-коментарів і вимоги до учасників коментувати успішні коментарі. Цей субреддіт під назвою Change My View (CMV) призначений для користувачів, які мають свою думку на тему, але бажають прислухатися до різних голосів і змінити свою думку. Користувачі, які публікують свої погляди в CMV, відомі як оригінальні плакати (OPs) у спільноті.

Версія цієї глави була подана для публікації в ACM Transactions on Social Computing. 109 користувачів надають коментарі з метою змінити думку ОП. Правила CMV вимагають, щоб ОП позначав(і) коментар(и), які успішно змінили його/її погляд, і коротко пояснював, чому подання змінено. У попередніх роботах, присвячених переконуванню та тексту, переконання широко визначають як інтерактивний процес, за допомогою якого дане повідомлення змінює точку зору людини, змінюючи знання, переконання чи інтереси, які лежать в основі цієї точки зору [3]. Це визначення спирається на відповідну

літературу в галузі грамотності, психології та філософії і застосовне до того, чого просять досягти користувачі CMV. Тому субреддіт CMV надає набір коментарів, які написані з метою переконання. Крім того, правило CMV щодо позначення коментарів дозволяє нам швидко розробити корпус переконання, оскільки коментарі вже опосередковано позначені ОП як переконливі або непереконливі.

Зосереджуючись на цьому корпусі, ми досліджуємо різні виміри переконливих коментарів у порівнянні з непереконливими. Спочатку ми вивчаємо текст коментарів щодо їх відповідності змісту оригінального допису. Через важливість часу та порядку в асинхронних комунікаціях між кількома учасниками, ми також враховуємо тимчасові атрибути коментарів та порядок їх введення. Згідно з нашими результатами, ці дві характеристики відіграють важливу роль в успіху коментаря щодо зміни переконань. Тому спочатку ми контролюємо ці дві змінні, а потім аналізуємо лінгвістичні показники, які не обов'язково доповнюють пропозиційний зміст коментаря, але мають на меті допомогти читачеві організувати, інтерпретувати та оцінити надану інформацію. Ці лінгвістичні особливості часто називають метадискурсом або не тематичним матеріалом і, як відомо, впливають на процес переконання [7].

У цій роботі багато з таких лінгвістичних показників фіксуються у словнику LIWC, який надає психологічно значущі категорії слів [5]. Крім того, відомо, що якість написання та організація тексту полегшують розуміння та пригадування аргументованого тексту, що згодом сприяє переконанню [4]. Тому ми проаналізували набір атрибутів, що стосуються узгодженості та витонченості написаного, таких як використання перехідних фраз, використання розділових знаків, а також лексичне різноманіття та перекриття в наших контрольованих наборах, і виявили відмінності між переконливими та не-переконливими. переконливі коментарі. Крім того, враховуючи, що текст, наданий із надійного джерела, швидше за все, спричинить переконання, ми беремо до уваги набір

атрибутів користувача. Прогностична сила, яку має кожна з цих ознак, далі аналізується.

Нарешті, як згадувалося раніше, CMV вимагає від користувачів включати пояснення, чому і як коментар може змінити їх погляд. З метою усвідомлення передбачуваних причин зміни переконань проведено попередній аналіз зібраних пояснень.

2. РОЗРОБКА МОДЕЛІ РІШЕННЯ ПРОГРАМНОГО СЕРВІСУ

2.1 Семантичний опис структури

Семантично обґрунтований текст складається з дискурсивних одиниць, пов’язаних між собою дискурсивними відносинами, які також називають риторичними відносинами. Незважаючи на зусилля зі створення надійних теоретичних основ і таксономій для таких відносин [8], сучасні методи їх автоматичного аналізу та виявлення в письмовому дискурсі ще мають удосконалити. Однак надання надійних моделей для аналізу та ідентифікації риторичних зв’язків може принести користь різним напрямкам дослідження в комп’ютерній лінгвістиці, таким як генерація тексту [9] та узагальнення [7], а також машинний переклад [8].

Однією з широко визнаних структур для аналізу та розуміння дискурсу є теорія риторичної структури (RST) [6]. У RST структура дискурсу має форму дерева, де листи відповідають елементарним одиницям дискурсу, а внутрішні вузли відповідають суміжним текстовим проміжкам. Кожен внутрішній вузол позначений риторичним відношенням, яке зберігається між його дочірніми вузлами. На малюнку наведено приклад дерева RST, взятого з корпусу RST [3]. Однією з помітних відмінностей RST від інших подібних теорій є те, що вона побудована на намірах авторів використовувати ці відносини [4]. Ця відмінна риса може ще більше ускладнити побудову моделей для автоматичної ідентифікації та класифікації риторичних відносин у контексті RST.

Риторичні відносини можуть бути як явними, так і неявними. Експліцитні відношення – це ті, які сигналізуються за допомогою реплік, таких як лексичні підказки, настрій, модальність та інтонація [4], тоді як в неявних зв’язках жодної репліки немає. У цьому дослідженні ми зосередимося на явних відносинах у письмовому тексті, про які сигналізує наявність лексичних ознак. Лексичні ознаки визначаються як мовні вирази, які функціонують як явні індикатори

дискурсивного відношення [7]. Наприклад, у реченні, представленому на малюнку, але і тому можна вважати лексичні ознаки, що сигналізують про існування відношення поступки та відношення пояснення-аргументація відповідно.

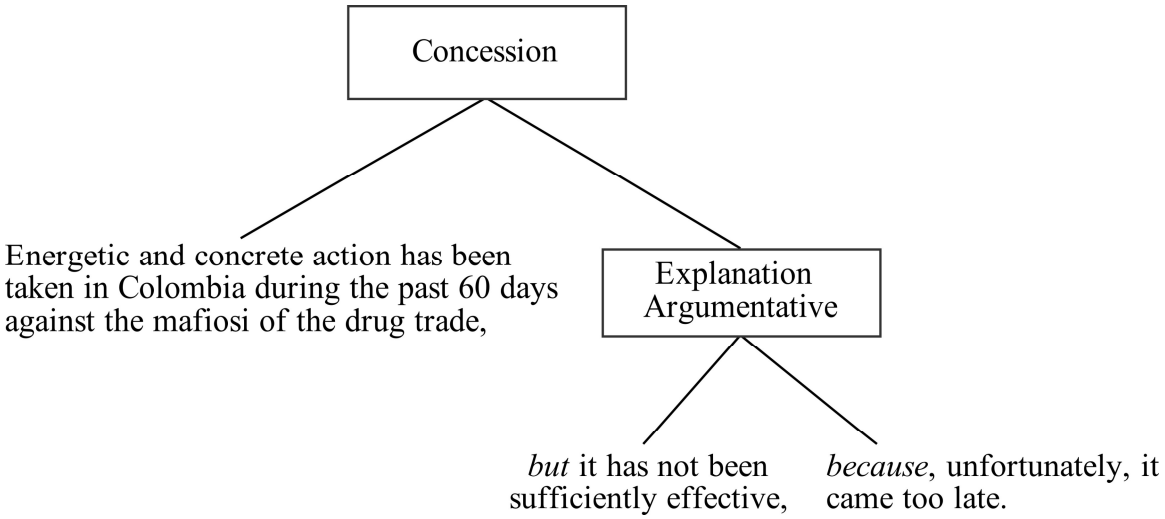


Рисунок 2.1 - Приклад речення, розібраного у формі RST

Оскільки це дослідження є частиною більшого проекту з визначення обґрунтування в письмовому дискурсі, ми зосереджуємось на трьох відносинах обставин, оцінки та розробки, які зазвичай присутні в обґрунтуваннях [8]. З метою запропонувати підхід, заснований на репліках, до виділення риторичних зв'язків, ми провели деякі експерименти на основі корпусів над анотованими корпусами RST. В результаті цих експериментів ми створили список ключових n-грам як потенційних лексичних ознак для кожного відношення. Такий метод на основі корпусу може призвести до виявлення недостатньо вивчених лексичних ознак.

Незважаючи на те, що лексичні ознаки можна використовувати для позначення риторичних відносин, вони не завжди є однозначними [2]. Деякі мовні вирази можуть функціонувати або не функціонувати як лексична підказка, або вони можуть сигналізувати про різні типи відносин у різних реченнях. Тому тут ми пропонуємо імовірнісну модель на основі графів, яка враховує

синтаксичні особливості речень. Ці моделі мають на меті визначити, в якому синтаксичному контексті лексичний сигнал дійсно сигналізує про наявність певного відношення.

Наведено та обговорено результати оцінки нашого підходу для відношення обставин. обставина вибирається як відношення фокусу, оскільки [1] виявило, що підходи, засновані на репліках, можуть бути добре підходять для виявлення обставин у різних жанрах, тоді як зв'язок розробки зазвичай не сигналізується [1]. Крім того, особливості основного текстового жанру можуть суттєво впливати на те, як оцінюється [1].

2.2 Аналіз літературних джерел по темі дипломної роботи

Більшість досліджень, зосереджених на аналізі дискурсу та класифікації дискурсу, повідомляють про результати як явних, так і неявних відносин [6]. Серед робіт, які особливо зосереджені на розшифровці лексичних ключів, велика частка проведена на Penn Discourse TreeBank (PDTB) [2], тоді як для вивчення інших теорій і рамок дискурсу було проведено менше досліджень.

Анотація PDTB є лексично обґрунтованою та теоретично нейтральною щодо структури дискурсу вищого рівня [2]. У ході анотації анотаторам було запропоновано шукати лексичні елементи, які можуть сигналізувати дискурсні відносини, а потім анотувати відповідні аргументи та відносини [2]. Навіть для неявних відношень анотаторів просили шукати сусідні речення, у яких відсутній один із цих сигналів. Коли можна було зробити висновок про дискурсне відношення, їх попросили спочатку позначити відношення лексичним елементом, який міг би служити сигналом, а потім анотувати значення відношення. Такий лексично орієнтований підхід до анотування дискурсивних відносин спонукав до роботи над розшифровкою лексичних ознак у PDTB.

Наприклад, в [9] використали набір синтаксичних особливостей разом із моделлю максимальної ентропії, щоб усунути три ознаки дискурсу про те, як,

відколи й коли. Їхній набір ознак включає форму допоміжного слова have, форму допоміжного слова be, форму голови та наявність модального. Вони отримали точність класифікації в 75,5% з урахуванням трьох можливих відчуттів, 71,8% для врахування чотирьох можливих відчуттів і 61,6% для трьох можливих відчуттів.

Також використовували набір синтаксичних ознак, щоб усунути неоднозначність сигналів дискурсу щодо їхнього дискурсного та недискурсового використання, а також розмежування сенсу. Їх особливості фокусування складаються з синтаксичної категорії самого маркера, його безпосереднього батька, а також його лівих і правих братів і сестер. Крім того, дві бінарні ознаки враховуються, щоб вказати, чи правий містить VP і чи правий брат/сестра містить слід. Їх найкращий набір функцій також включав функції парної взаємодії між сигналами та синтаксичними функціями, а також попарні взаємодії між синтаксичними функціями.

Їхній алгоритм навчання привів до F-бала 92,28% і точності 95,04% для дискурсу проти використання не-дискурсу і точності 94,15% для смислової класифікації.

Пізніше ці результати були вдосконалені в [10], де введено набір нових поверхневих і синтаксичних ознак, які об’єднані з набором ознак, представленим у [2]. Результати класифікатора максимальної ентропії, натренованого за допомогою цього набору функцій, дали точність 97,78% і F-показник 96,22%.

У більш широкому контексті побудови наскрізного аналізатора дискурсу для PDTB [5] створили класифікатор репліки, щоб визначити, чи функціонує лексичний елемент як репліка дискурсу чи ні. На додаток до функцій, використаних у [2], вони також включали особливості частини мови, а також функції, пов’язані з синтаксичним розбором шляху від репліки до кореня дерева. Використання їхнього набору лексико-синтаксичних і шляхових особливостей призвело до точності 97,25% і F-бала 95,36%.

Незважаючи на те, що RST є однією з найбільш широко прийнятих платформ для аналізу дискурсу, відносно мало уваги приділено анотованим корпусам RST щодо аналізу лексичних сигналів і тлумачення. На відміну від PDTB, анотації після RST не є лексично обґрунтованими, і кожне відношення визначається з точки зору намірів, які спонукають авторів використовувати ці специфічні відносини [4]. Таким чином, діаграма RST представляє деякі цілі або наміри авторів щодо включення кожної частини тексту [4]. Такі атрибути анотацій RST роблять складним завданням вивчення ролі лексичних одиниць у класифікації відношень і потенційне їхнє визначення неоднозначності.

В [7] спробували створити алгоритм риторичного аналізу, який бере необмежений вільний текст і генерує дійсне дерево риторичної структури. Корпусне дослідження було проведено, щоб зрозуміти, як можна використовувати сигнали для визначення елементарних дискурсивних одиниць і висунення гіпотези щодо їх відповідного співвідношення. Використовуючи попередні дослідження з аналізу дискурсу, він створив список із 450 ознак дискурсу для початку. Потім із корпусу було зібрано в середньому 17 текстових проміжків, пов’язаних з кожною реплікою. Потім усі речення були анотовані двома наборами метаданих: інформація, пов’язана з дискурсом (наприклад, маркер, використання та позиція) та алгоритмічні функції. Використовуючи ці анотації, які здебільшого фіксують орфографічне середовище сигналів, набір регулярних виразів був створений вручну для розпізнавання потенційних сигналів дискурсу. Якщо репліка дискурсу мала різні функції дискурсу в різних орфографічних середовищах, для кожного випадку створювався окремий регулярний вираз. Алгоритм привів до 84,9% F-показу від відкликання 80,8% і точності 89% для підзавдання ідентифікації сигналу. Для підзавдання класифікації відношень вони досягли відкликання 47,0% і точності 78,4%, що призвело до оцінки F 58,76%.

У [4] набір неоднозначних сигналів дискурсу вперше витягується з бази даних іспанських сигналів дискурсу. Контекст кожного сигналу потім

витягується з RST Spanish Treebank і передається синтаксичному аналізатору. Синтаксичні особливості контексту кожної репліки потім вручну аналізуються, щоб визначити потенційні мовні закономірності та закономірності. Використовуючи результати аналізу, розробляються лінгвістичні правила, щоб усунути неоднозначність лексичних ознак. Оцінюючи свої правила на тестовому корпусі RST Spanish TreeBank, вони досягли точності 60,65%.

Обидві роботи над анотаціями RST є напівавтоматичними та включають кроки вручну. Наша робота має на меті забезпечити повністю автоматизований підхід до виявлення потенційних лексичних сигналів, які можуть вказувати на риторичні відносини, і проаналізувати, чи може їхній синтаксичний контекст бути корисним для розшифровки сигналів і класифікації сенсу.

2.3. Структура пропонованого підходу

У цьому розділі ми спочатку опишемо два анотовані корпуси RST, які використовуються в цій роботі: корпус RST [3] та оглядовий набір даних SFU [5]. Потім представлено пояснення підходу, що використовується для виділення набору ключових n-грам як потенційних лексичних ознак, за яким слід опис нашого підходу, заснованого на графах, для роз’яснення лексичних ознак.

Ми використовували два анотовані людиною корпуси як наші базові набори даних для експериментів: корпус RST [3] та набір даних огляду SFU [5]. Обидва корпуси анотовані в структурі RST і створені за допомогою RSTTool. Корпус RST, який протягом багатьох років був доступним Консорціумом лінгвістичних даних, включає 385 статей Wall Street Journal і охоплює понад 178 000 слів. Серед екземплярів відносин у корпусі RST існує близько 700 випадків обставин, що становить майже 3% від загальної кількості екземплярів відносин.

Рецензійний корпус SFU — це колекція з 400 оглядових документів з фільмів, книг та товарів народного споживання. Цей набір даних містить понад

303 000 слів і був зібраний у 2004 році з веб-сайту Epinions. Існує близько 1300 випадків обставин, що становлять майже 7% анотованих випадків у корпусі.

Вибір лексичного сигналу

Текст новин має добре структурований офіційний стиль написання, тоді як онлайн-огляди відносно менш структуровані та неформальні, написані користувачами з широким спектром письмових здібностей. Тому для виділення лексичних ознак, пов'язаних із даним відношенням, ми використовували корпус RST.

Спочатку всі екземпляри відношень витягуються з корпусу RST і збираються в документ відношення, названий на честь відповідного відношення. Потім, дотримуючись підходу, запропонованого в [1], усі n-грами (аж до триграм) витягуються із складеного документа відношення. Для кожного n-грама потім обчислюється змінена версія метрики tf-idf. Показник idf все ще обчислюється на основі кількості документів, які містять n-грам, та загальної кількості документів у корпусі. Однак, оскільки кожен рядок відповідає одному екземпляру відношення, метрика tf розраховується на основі кількості рядків, які містять принаймні один екземпляр n-грама. Ця змінена метрика дозволяє нам компенсувати потенційне упередження, яке може бути викликано метрикою tf для слів, які з'являються більше одного разу в екземплярі відношення.

Потім список витягнутих n-грам (тобто лексичних ознак) фільтрується, щоб включати лише n-грами з їх tf-idf вище 0,5. Щоб відфільтрувати будь-які специфічні для корпусу n-грами, які можуть з'явитися в списку, n-грами, витягнуті з корпусу RST, застосовуються до набору даних огляду SFU, щоб ідентифікувати відповідне відношення. Потім F-показник кожного n-грама обчислюється незалежно. Нарешті, n-грами з F-показом вище 0,1 вибираються як потенційні лексичні ознаки. Нижче наведено покрокове пояснення процесу для відношення обставин, що призвело до виділення семи лексичних ознак: Коли, після, на, перед, з, поза, як. Ці результати детально проаналізовані в іншому документі [11]:

• Витягніть усі екземпляри обставин із корпусу RST та сформуйте документ про обставини

• Обчисліть змінену версію tf-idf для всіх n-грам (до триграмів), які з'являються в документі обставин.

• Опрацюйте список n-грам, щоб видалити ті, у яких tf-idf нижче певного порогу.

• Використовуйте оновлений список n-грам, щоб витягти обставини з корпусу огляду SFU та самостійно обчислити їх F-показ.

• Виберіть остаточні сигнали з F-оцінкою вище певного порогу.

Визначення лексичних ознак

Наш підхід до визначення неоднозначності в основному натхненний роботою [6] щодо виявлення речень із установками. Під час їхнього дослідження фрагмент тексту, який містить другий займенник, спочатку виділяється як найбільш релевантна частина речення. Потім ці фрагменти представляють за допомогою різних шаблонів, фіксуючи їх синтаксичні особливості та семантичну орієнтацію. Для кожного типу шаблонів моделі графіків будуються на основі пропозицій з ставленням і без нього. Нарешті, ймовірність створення нового речення на основі цих моделей використовується для прогнозування існування ставлення. Ми скористалися їхнім підходом для розмежування лексичних ознак. Наші графічні моделі побудовані на корпусі RST та оцінені на корпусі огляду SFU і навпаки. Тому процедура побудови графіка, описана нижче, проводиться для обох базових корпусів Data Collection.

Для кожного витягнутого сигналу ми спочатку створюємо два відповідних документи з анотованих корпусів. Один документ складається з усіх екземплярів відношень, які містять сигнал і анотовані відношенням фокусу (наприклад, усі екземпляри обставин, про які сигналізує коли). Відтепер такі випадки називатимуться позитивними. Інший документ складається з усіх екземплярів відношень, які містять сигнал і анотовані як будь-яке відношення, за винятком відношення фокусу (наприклад, усі екземпляри не-обставин, які

містять коли). У решті цього рукопису ми називатимемо ці випадки негативними.

RST постулює ієрархічну структуру в тексті, де екземпляр відношення може бути вбудований в інші екземпляри. Тому під час вилучення екземплярів ми переконалися, що не збираємо негативні екземпляри, які включають будь-який позитивний чи негативний підекземпляр. Ми також переконалися, що не збираємо жодних позитивних екземплярів, які включають негативні підекземпляри. Включення таких вбудованих екземплярів призвело б до надлишкових і неправильних точок даних. Наприклад, розглянемо такий позитивний екземпляр із корпусу RST.

Під час збору негативних екземплярів було виявлено, що цей екземпляр був вбудований у десять негативних екземплярів. Однак, оскільки фактично функціонує як ознака обставин у всіх з них, ці десять випадків не могли кваліфікуватися як негативні випадки, тому були виключені.

Після створення документів кожен екземпляр обробляється і перетворюється на два різних представлення, фіксуючи синтаксичні особливості екземпляра. Щоб створити перше синтаксичне представлення, слова в екземплярах замінюються відповідними тегами Part-OfSpeech (POS), а сам сигнал залишається таким, як є. Друге представлення включає найкоротший шлях від кореневого елемента до сигналу в дереві розбору залежностей. Нижче наведено приклад відношення обставин разом із двома відповідними синтаксичними репрезентаціями:

• Позитивний екземпляр з when як сигнал:

• Представництво на основі POS:

Коли NNP NNP VBD TO NN PRP VBD IN JJ NNS IN NN

• Представлення найкоротшого шляху: root advmod

Ми використовували інструментарій OpenNLP для токенізації та POS-тегування екземплярів і аналізатор залежностей для створення дерев розбору [2].

Графове моделювання

- Ми закодували синтаксичну інформацію екземплярів у графічних моделях. Будуємо орієнтований зважений графік G = (V, E),w, де:

- V — набір усіх можливих лексем, які можуть з'являтися в представленнях. Наприклад, для представлень POS, V є об'єднанням набору всіх POS-тегів і набору сигналів. • E = V × V — множина всіх можливих упорядкованих переходів між будь-якими двома лексемами.

- w → [0 − 1] — це зважувальна функція, яка призначає значення ймовірності ребру (i, j), що представляє ймовірність переходу від лексеми i до лексеми j.

Враховуючи набір синтаксичних уявлень, ймовірність переходу від лексеми i до лексеми j розраховується за оцінкою максимальної правдоподібності. Таким чином, ймовірність обчислюється шляхом ділення кількості разів, коли токен i одразу слідує за маркером j, на кількість разів, коли сам маркер i з'являється в наборі.

Цей метод побудови графіків подібний до мовного моделювання, але проводиться за набором синтаксичних уявлень [6]. Для кожного виду представлення ми будуємо один графік на основі набору позитивних екземплярів і один на основі набору негативних екземплярів. В результаті, враховуючи сигнал (наприклад, коли) і відповідне йому відношення (наприклад, обставина), ми будуємо чотири моделі графіка на основі наступних наборів: POS-представлення позитивних екземплярів, POS-подання негативних екземплярів, представлення позитивних екземплярів, аналізованих залежностями. екземпляри та залежність аналізували уявлення заперечних речень.

Модель розшифровки Cue

Нарешті, для нашої остаточної моделі розшифровки сигналів ми використовуємо значення ймовірності, отримані з наших моделей графіків, як набір функцій для стандартної моделі машинного навчання. За даними екземпляра та графіка ми обчислюємо ймовірність того, що його синтаксичні

уявлення будуть згенеровані з відповідних синтаксичних графіків. Ймовірність того, що синтаксичне представлення R, яке складається з послідовності лексем T1,T2,...,Tn, буде сформовано з графіка G, оцінюється за допомогою наступної формули. Зауважте, що W – це зважувальна функція або функція переходу ймовірності.

Враховуючи, що у нас є чотири моделі графіків, ми можемо створити чотири значення ймовірності як наш набір ознак. Ці функції далі використовуються в стандартному алгоритмі навчання з наглядом, щоб усунути неоднозначність сигналу та класифікувати відношення даного екземпляра. На малюнку наведено високорівневий опис всієї процедури виділення та розшифровки лексичних ключів.

2.4. Оцінка отриманих результатів

З огляду на те, що наша кінцева мета полягає в тому, щоб виявити обґрунтування з письмового дискурсу, наш підхід оцінюється для відношення обставин як єдиного відношення на основі репліки, яке, як відомо, часто присутнє в обґрунтуваннях [1, 8]. Ми проводили експерименти з використанням різних форм уявлень POS на основі кількості POS-тегів, що оточують сигнал, і деталізації тегів.

Таблиця 2.1 - POS-теги у наборі тегів Penn English Treebank, експерименти проводяться з використанням тегів, згрупованих відповідно до різних рівнів деталізації

Label	Medium Granularity	Coarse Granularity
JJ	JJ, JJR, JJS	JJ, JJR, JJS, DT, WDT
NN	NN, NNS, NNP, NNPS	NN, NNS, NNP, NNPS
PRP	PRP, PRP$	PRP, PRP$, WP, WP$
RB	RB, RBR, RBS	RB, RBR, RBS, WRB
WP	WP, WP$	-
VB	VB, VBD, VBN, VBP,	VB, VBD, VBN, VBP,

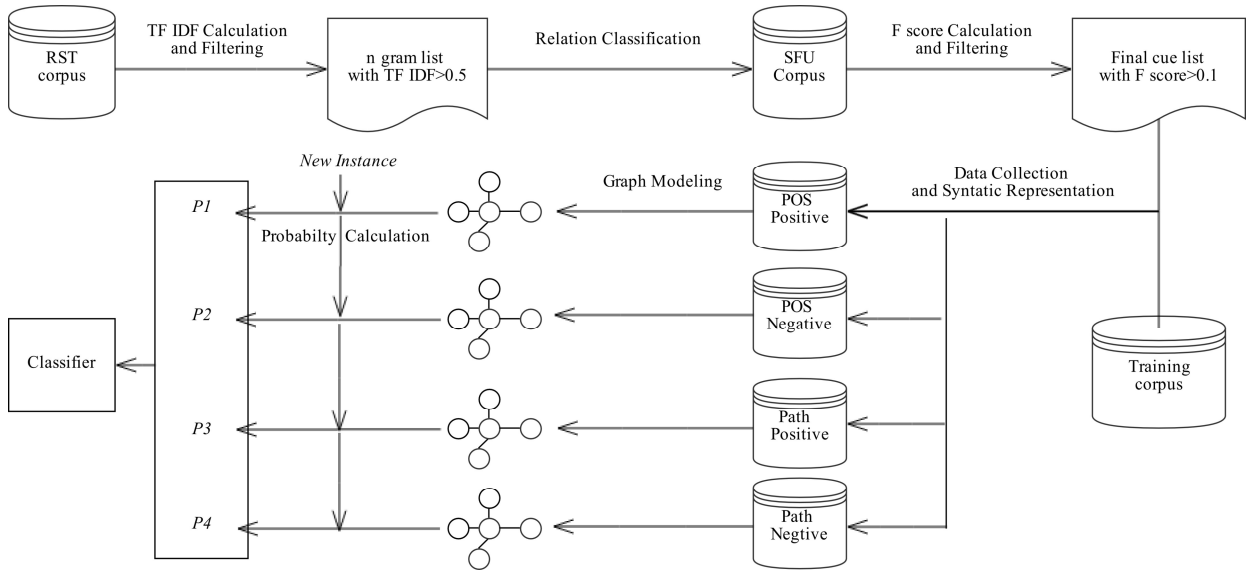


Рисунок 2.2 - Огляд високого рівня підходу до виділення сигналу та розмежування

Ми проводили експерименти, використовуючи весь екземпляр POS-тегів, використовуючи два POS-теги до і два теги після сигналу, а також використовуючи один до і один після сигналу. Ми також використовували три рівні деталізації POS-тегів, включаючи найтонший.

Набір тегів Penn English Treebank, який використовується OpenNLP. Ми також використовували середню та грубу деталізацію, які створюються шляхом категоризації подібних тегів в один високорівневий тег. Таблиця показує теги, які класифікуються на кожному з цих двох рівнів деталізації. Зауважте, що теги, не згадані в рівнях деталізації, використовуються як є.

Використання цих трьох варіантів двох атрибутів POS-тегів призвело до дев'яти різних налаштувань експерименту. Ми досягли найкращих результатів на обох корпусах, використовуючи один тег до і один тег після сигналу та середній рівень деталізації. У цьому розділі ми повідомляємо про результати експериментів із використанням цього конкретного налаштування POS.

Остаточний алгоритм, побудований на значеннях ймовірності, виконується та оцінюється за допомогою Weka Workbench та класифікує екземпляри за допомогою регресії. Для оцінки моделі використовується

стратифікований підхід десятикратної перехресної перевірки. Щоб отримати уявлення про ефективність нашої моделі у розрізненні різних сигналів, результати звітуються для кожного з семи ознак незалежно. Фільтр SMOTE використовувався, коли ми зіткнулися зі значним дисбалансом класів.

Було показано, що використання синтаксичного контексту для роз'яснення лексичних сигналів є корисним для роз'яснення символів у лексично орієнтованих відносинах (наприклад, відношення PDTB). У цьому дослідженні ми зосередили свої зусилля на анотованих корпусах RST і дослідили потенціал синтаксичного контексту для розшифровки сигналів у структурі RST. Ми продемонстрували, що синтаксичні ознаки можуть мати велике значення при класифікації явних риторичних відносин. Крім того, на відміну від більшості попередніх досліджень, присвячених розшифровці сигналів, ми закодували синтаксичний контекст сигналів у формі графіків. Очікувалося, що цей підхід на основі графіків забезпечить більш узагальнений та ефективний підхід. Щоб перевірити цю гіпотезу, ми провели експерименти, щоб порівняти нашу модель на основі графіків.

Таблиця 2.2 - Результати класифікації десятикратної перехресної перевірки на корпусі ДФУ

Cue	Precision	Recall	F-score
When	0.55	0.79	0.65
After	0.46	0.56	0.51
On	0.73	0.75	0.74
Before	0.62	0.60	0.61
With	0.76	0.80	0.78
Out	0.60	0.54	0.57
As	0.71	0.82	0.76
Average	0.63	0.69	0.66
Weighted Average		0.71	

Тут значення ймовірності, які використовуються як базовий набір ознак, виводяться з моделей графіків, побудованих на корпусі RST.

Таблиця 2.3 - Результати класифікації десятикратної перехресної перевірки набору даних RST

Cue	Precision	Recall	F-score
Коли	0.62	0.69	0.65
Після	0.61	0.57	0.59
Увімкнено	0.77	0.81	0.79
Раніше	0.70	0.40	0.51
З	0.77	0.81	0.79
Вихід	0.75	0.70	0.73
Як	0.73	0.67	0.70
Середній	0.71	0.67	0.68
Weighted Average		0.69	

Значення ймовірності, які використовуються як базовий набір ознак, виводяться з моделей графіків, побудованих на корпусі SFU. З конкретним використанням синтаксичних функцій, де безпосередньо використовувалися тег POS і відображення шляху залежностей. Логістичну модель спочатку навчали на корпусі RST та тестували на наборі даних SFU (див. таблицю 2.4), а потім навчали на корпусі SFU та тестували на корпусі RST (див. таблицю 2.5). Результати стабільно нижчі для всіх трьох показників, що підтверджує перевагу нашого підходу. У зв’язку з відсутністю попередньої роботи над розшифровкою сигналів у контексті RST, замість використання базової лінії, це порівняння зроблено, щоб підкреслити внесок нашої роботи.

Виходячи з результатів запропонованого нами підходу, можна було побачити, що три лексичні ознаки коли, після та до мають найнижчу продуктивність у корпусі RST (див. таблицю 2.3). Вони також входять до чотирьох ознак з найнижчим F-показом у наборі даних SFU (див. Таблицю 2.2). Цей висновок можна пояснити тим, що для цих трьох сигналів відповідні набори даних були одними з найменших наборів сигналів. Можливо, що більш важливо, ці три ознаки можуть функціонувати як тимчасові індикатори, що може ускладнити їх визначення. Наприклад, розглянемо наступні екземпляри, отримані з SFU

Таблиця 2.4 - Результати класифікації на корпусі SFU, коли синтаксичні ознаки використовуються безпосередньо для навчання моделі на корпусі RST

Cue	Precision	Recall	F-оцінка
Коли	0.50	0.41	0.45
Після	0.39	0.15	0.22
Увімкнено	0.65	0.28	0.39
Раніше	0.52	0.49	0.51
З	0.53	0.34	0.41
Вихід	0.51	0.19	0.28
Як	0.49	0.24	0.32
Середній	0.51	0.30	0.37

Корпус:

• Позитивний екземпляр;

• Негативний екземпляр.

Перше речення є прикладом відношення обставин, про який сигналізує when, тоді як у другому, коли має на увазі часовий аспект речення і не є сигнальною обставиною. Ми очікуємо, що певні лінгвістичні та контекстуальні особливості, пов’язані з текстом, можуть бути корисними для роз’яснення таких лексичних ознак, як-от час дієслова та/або атрибути агентів різних речень у реченні. Щоб вивчити ці особливості, потрібні подальші дослідження.

Два основних корпуси, використані в цьому дослідженні, належать до дуже різних текстових жанрів. Крім того, оскільки RST робить акцент на намірах письменника та ефекті відношення на читача [4], анотації RST за своєю суттю є суб’єктивними і засновані на розумінні тексту читачами [4]. Тому між двома корпусами можуть бути відмінності через потенційно різні знання, якими володіє кожен набір анотаторів щодо культури, ситуації та мови, які представляє текст [4].

Незважаючи на жанрову диспропорцію та різницю в анотаціях, за допомогою запропонованої моделі ми отримали обнадійливі результати. Однак очікується, що результати покращаться, коли моделі будуть побудовані на корпусах подібних жанрів і анотовані з використанням основних правил істини.

Таблиця 2.5 - Результати класифікації на корпусі RST, коли синтаксичні ознаки використовуються безпосередньо для навчання моделі на корпусі SFU

Cue	Точність	Пригадайте	F-оцінка
Коли	0.53	0.65	0.58
Після	0.31	0.14	0.20
Увімкнено	0.36	0.17	0.23
Раніше	0.33	0.22	0.26
З	0.62	0.20	0.30
Вихід	0.57	0.53	0.55
Як	0.52	0.25	0.34
Середній	0.52	0.25	0.35

2.5. Обгрунтування отриманої концепції

За останні кілька десятиліть було розроблено і розроблено багато програмних програм для вирішення різних аспектів управління проектами, таких як оцінка зусиль проекту [4], управління спільною діяльністю [3] та моніторинг змін і ризиків проектів [7]. У цьому дослідженні ми зацікавлені в наданні допомоги керівникам проектів у прийнятті ефективних рішень щодо змін, пов’язаних з проектом. Проекти не проводяться у вакуумі і зазвичай на них впливають різні динамічні фактори, такі як наявність бюджетних ресурсів, рівень пріоритету в організації та члени проекту. Отже, очікується, що менеджерам проектів зазвичай потрібно вносити зміни в проекти під час процесу управління.

Дослідники досліджували комп’ютерні засоби, які допомагають керівникам проектів приймати обґрунтовані рішення, коли проекти потрібно адаптувати до нової ситуації. Наприклад, Карвонен представив процес управління, підтримуваний комп’ютером, щоб проілюструвати, як комп’ютерні системи можуть підтримувати прийняття рішень менеджером проекту в ситуації змін під час виконання проекту, а також у безперервному вдосконаленні бізнес-процесів компанії. Sauve та ін. представили метод оцінки ризику, пов'язаного зі зміною, яка має бути внесена в інфраструктуру та послуги в управлінні IT-послугами. Завдяки метриці ризику цей метод автоматично призначає пріоритети змінам.

Наш підхід допомагає менеджерам проектів приймати рішення щодо змін, представляючи інтерактивне візуальне уявлення про історію змін попередніх проектів. Наше припущення полягає в тому, що, зіткнувшись із ситуацією змін, менеджери проектів можуть стати більш ефективними

Тому наш дизайн інтерактивного інструменту візуалізації був зосереджений на представленні та створенні асоціацій 1) між історією змін попередніх проектів, їх причинами та обґрунтуваннями їхніх рішень; 2) між рішеннями та статусом проекту як індикатором впливу рішень на проект. У цьому плакаті ми описуємо наш поточний прототип системи ProjectAcc. У решті розділів ми спочатку обговорюємо пов'язані роботи, а потім представляємо обґрунтування дизайну прототипу. Ми закінчуємо нашими планами оцінки користувачів і підсумком дослідження.

Після огляду інструментів управління проектами, таких як TeamSpace і TeamSCOPE, Smith et al. запропонував рекомендації щодо проектування інструменту управління проектами для розподілених команд дизайнерів. Вони стверджували, що для полегшення повторного використання знань з управління проектами з попередніх проектів важливо заархівувати як знання про продукти, так і знання про процес. Щоб проілюструвати свій підхід до конфігурації інформаційних систем управління проектами, Berzi¯ sa та Grabis представили, як знання попередніх запитів на зміни можна поєднати із запитом у поточному проекті. У їхньому підході запит має дві конфігурації в системі: атрибути запиту (наприклад, пріоритет, опис, залишок, термін виконання) і робочі процеси стану (наприклад, закрито, відкрито, погоджено). Коли зроблено запит у поточному проекті, система надасть запропоновані описи атрибутів і робочих процесів стану на основі попередніх проектів, тобто сховища знань з управління проектом.

У порівнянні з системою, представленою Berzi¯ sa та Grabis, яка полегшує конфігурацію поточних змін на основі історії попередніх проектів, наша система зосереджена на полегшенні процесів прийняття рішень

керівниками проектів, використовуючи знання про зміни з попередніх проектів. Крім того, ми використовуємо методи інтерактивної візуалізації, щоб дозволити керівникам проектів переглядати та витягувати інформацію про зміни, що стосується поточної ситуації проекту, а також міркувати про зв'язок між змінами та їх впливом на проекти.

Наш прототип, так званий ProjectAcc, розроблений і побудований на базі даних історії програмного забезпечення з відкритим вихідним кодом EGroupware (http://www.egroupware.org/). Програмне забезпечення EGroupware призначене для компаній, щоб керувати контактами, зустрічами, проектами та списками справ. Це активно використовувана програма з надійною репутацією користувача. Наприклад, його оцінка користувачів становить 4,2 з 5 на SourceForge.net (164 голоси), а на веб-сайті для спільної роботи підприємства є 9352 теми та 27071 допис для форуму користувачів EGroupware станом на 13 вересня 2013 року. Таким чином, ми розуміємо, що історична база даних EGroupware широко прийнята та використовується в реальних випадках управління проектами, тому ми використовували її як базову базу даних для ProjectAcc.

База даних історії EGroupware надає різноманітні атрибути для проектів (наприклад, назва, опис, пріоритет, використаний бюджет) і кілька атрибутів для змін (наприклад, проект, змінений атрибут проекту, мітка часу). Ми змінили цю базу даних історії, щоб включити причини змін. Завдяки причинам, доступним керівникам проектів, вони можуть легко отримати рішення щодо змін, які відповідають тим же причинам, що й поточна ситуація. Під час модифікації проекту менеджерів проекту просять призначити одну або кілька причин зміни. Ці причини можна вибрати із попередньо визначеного списку потенційних причин, витягнутих з літератури [2], включаючи зовнішні причини (наприклад, політичні, економічні, потреби клієнтів, ринкова сила), а також внутрішні причини (наприклад, дефіцит персоналу або бюджету, стратегічні рішення,

контроль якості). Щоб дозволити керівникам проектів адаптувати причини до своєї конкретної організаційної ситуації, їм дозволено додавати нові до списку.

Стверджуючи про важливість надання обґрунтування змін керівнику проекту, ми також включили цю інформацію до бази даних. У нашому дослідженні ми визначаємо обґрунтування зміни як інформацію, яка обґрунтовує рішення про зміну. Соціальні психологи досліджували види інформації в спілкуванні, які впливають на ставлення людей [6]. Ці дослідження показали, що поширена інформація, яка є правдоподібною та логічною та додає щось нове до проблеми, швидше за все, спричинить зміну ставлення [5]. Фабрігар, Прістер, Петті та Вегенер виявили, що коли люди мали більший доступ до різних поглядів, якість аргументації повідомлення мала більший вплив на переконання. Ми віримо, що представлення не тільки рішень щодо змін, але й їх обґрунтування може бути корисним для менеджерів проектів при вирішенні поточної ситуації. Дослідження Сяо та Керролла вказує на те, що на навички міркування людей впливає чітке обґрунтування завдання під час групової діяльності. Очікується, що завдяки архівації та поширенню обґрунтування змін у проектах організації керівники проектів покращать свої навички міркування щодо прийняття рішень у довгостроковій перспективі. Покращені навички міркування можуть позитивно вплинути на якість рішень щодо змін, оскільки рішення приймаються після кращих процесів міркування.

Дизайн інтерфейсу

Інтерфейс ProjectAcc розділений на два окремі компоненти, розташовані вертикально на екрані: компонент огляду історії та компонент детального перегляду проекту.

Кожен компонент системи надає певні скоординовані уявлення про різні виміри базового набору даних, підтримуючи менеджерів проектів у глибшому уявленні про дані . Компонент огляду історії складається з трьох різних уявлень, а саме: сітка історії, панель причин і панель обґрунтування. Сітка історії призначена для наочного огляду змін, внесених у різні атрибути різних проектів.

У цьому поданні кожен рядок представляє проект у базі даних, а кожен стовпець — атрибут.

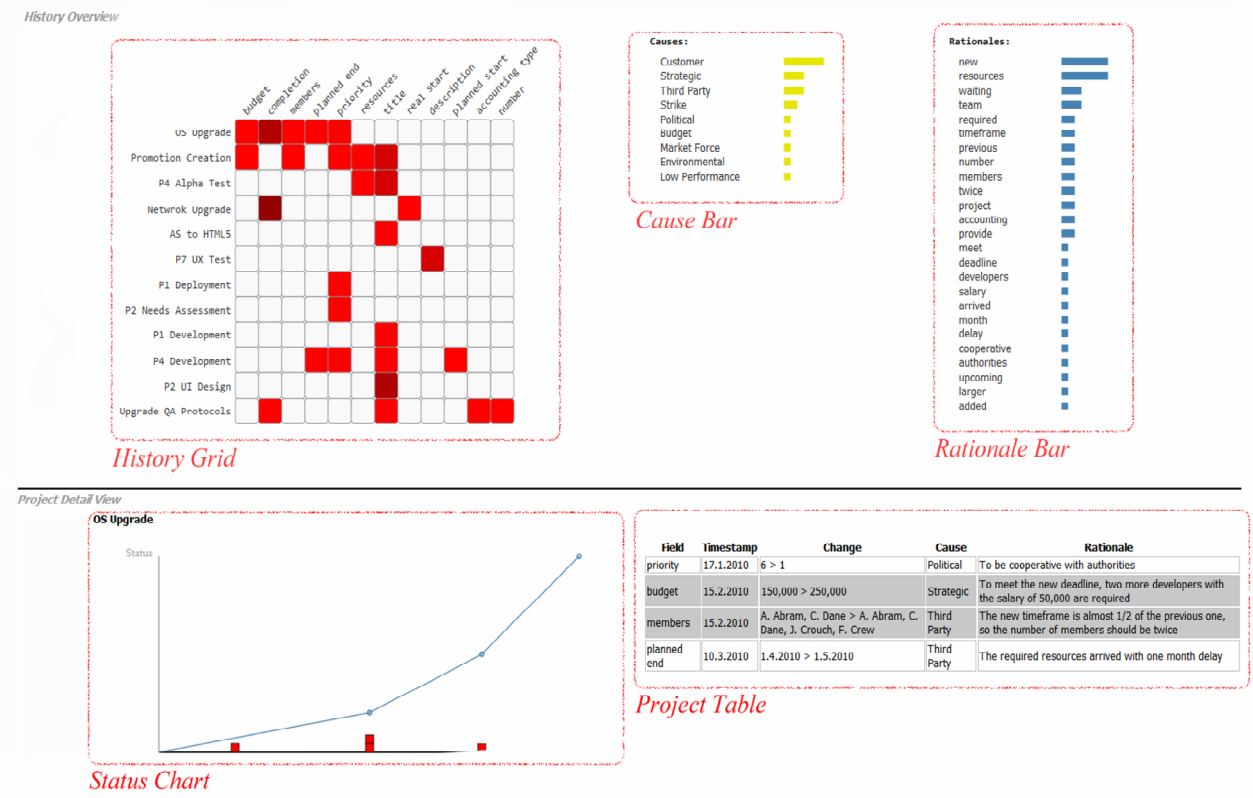


Рисунок 2.3 - Знімок екрана ProjectAcc

2.6. Представлення інтерактивного інструменту візуалізації

Оскільки колір є візуальною ознакою попередньої уваги і може бути окремо декодований від просторового положення зоровою системою людини [11], кольорове кодування використовується для представлення кількості змін в різних атрибутах проектів. Якщо зміни не відбулося, клітинка забарвлюється в світло-сірий колір. У разі зміни кількість змін кодується за допомогою каналу яскравості червоного кольору; таким чином, клітинки з високою частотою зміни виглядають темнішими, а клітини з низькою частотою змін — яскравішими.

Сітка історії додатково збагачується механізмом взаємодії, що дозволяє менеджерам проектів сортувати проекти за кількістю змін у певному атрибуті,

клацнувши мітку атрибута. Аналогічно, клацання на мітці проекту призводить до сортування списку атрибутів відповідно до кількості змін для цього проекту. Таке візуальне уявлення дозволяє менеджерам проектів швидко отримати уявлення про частоту змін у різних проектах і різних атрибутах, а також дослідити цю інформацію для проектів і атрибутів, що цікавлять, за допомогою функції сортування.

На додаток до сітки історії, компонент огляду історії пропонує дві вертикальні гістограми, щоб наочно відобразити огляд причин та обґрунтування змін. Панель причин представляє відсортований вигляд причин від найчастішої до найменш поширеної, тоді як панель обґрунтування показує найчастіші слова, які використовуються в обґрунтуваннях змін. Оскільки стоп-слова не мають значення при дослідженні визначних слів у обґрунтуванні, вони вилучаються з панелі обґрунтування. Історична сітка тісно пов’язана з причиною та планкою обґрунтування за допомогою техніки зв’язування та розчісування. Наведення курсору на кожну клітинку в сітці історії виділяє відповідні причини й обґрунтування в обох гістограмах. Крім того, якщо менеджери проекту наводять курсор на причину або обґрунтування, відповідні клітинки в сітці історії будуть виділені.

Компонент детального перегляду проекту

Вивчаючи огляд історії змін, менеджери проекту можуть бути зацікавлені детально зосередитися на конкретному проекті. Подвійне клацання на мітці проекту завантажує компонент детального перегляду проекту з двома пов’язаними видами діаграми рядка стану та таблиці проекту. На діаграмі рядка стану вісь x представляє тривалість проекту як часову шкалу. Ця вісь доповнена червоними гліфами, кожен з яких представляє зміни, що відбулися в проекті в певний час. Представлення змін на часовій шкалі може дозволити менеджерам проекту дослідити послідовність змін і побачити, як конкретна зміна викликала

наступні зміни. Вісь Y діаграми відображає статус проекту, що дозволяє керівникам проекту оцінити вплив певної зміни або серії змін на статус проекту.

Нарешті, таблиця проекту надає менеджерам проекту детальну інформацію про зміни, включаючи змінений атрибут, час зміни, старі та нові значення, причини зміни та повне обґрунтування, у текстовому форматі. Ці два уявлення також пов’язані з технікою зв’язування та використання пензля, оскільки наведення курсора миші на гліф у діаграмі чи рядок у таблиці викликає виділення відповідного представлення змін в іншому поданні. Ці два представлення дозволяють менеджеру проекту визначити потенційно цікаву зміну в одному поданні, а потім виявити ту саму зміну в іншому поданні для подальшого дослідження.

Управління змінами «це цілісний процес, пов'язаний з усіма внутрішніми та зовнішніми факторами проекту, що впливають на зміни проекту; до можливої зміни прогнозу; виявлення вже відбулися змін; планувати превентивні впливи; до координації змін у всьому проекті» [9]. У цій роботі ми представили прототип програмного забезпечення, головна мета якого — допомогти менеджерам проектів приймати більш ефективні та ефективні рішення під час прийняття рішення про можливі зміни. Завдяки нашому поточному дизайну ProjectAcc надає менеджерам проектів можливість переглядати, досліджувати та отримувати уявлення про історію причин, рішень та обґрунтування попередніх змін за допомогою методів інтерактивної візуалізації. Основне обґрунтування нашого дизайну полягає в тому, що архівування причин та обґрунтування змін настільки ж важливе, як і самі рішення щодо змін для подальшого повторного використання. Зараз ми проводимо оціночне дослідження, щоб вивчити зручність використання та наше обґрунтування дизайну.

Вивчення та аналіз відношень, як будівельних блоків узгодженості в дискурсі, може сприяти розробці складних лінгвістичних додатків і алгоритмів. З метою полегшення автоматичного виявлення явних риторичних зв’язків у тексті,

ми розробили алгоритм, щоб спочатку виявити потенційні лексичні ознаки, а потім усунути їх, передбачивши відношення.

Змінена версія tf-idf була використана для виділення потенційних сигналів, а модель на основі графіка, побудована на синтаксичних ознаках, була використана для вирішення завдання розшифровки сигналу. Загалом, результати оцінювання вказують на ефективність синтаксичних ознак у розкритті лексичних ознак та передбаченні явних риторичних зв'язків у різних жанрах. Наші експерименти показали перевагу нашого підходу до кодування таких синтаксичних ознак у імовірнісному графі порівняно з їх прямим використанням.

Як згадувалося раніше, це дослідження є нашою першою спробою визначити обґрунтування в тексті. Обгрунтування — це пояснення причин, що лежать в основі рішень, висновків та тлумачень. Попередні дослідження обґрунтування формулювання та обміну свідченнями свідчать про те, що це сприяє контролю якості, управлінню знаннями та повторному використанню знань. Проте було лише кілька спроб застосування обчислювальних методів для визначення обґрунтування погано структурованого тексту. Наші майбутні дослідницькі зусилля спрямовані на розробку алгоритмів для виявлення легкосигнальних і неявних відносин, а також на подальше дослідження потенціалу та обмежень використання таких риторичних зв'язків у визначенні обґрунтування в тексті дискурсу. Крім того, ми плануємо оцінити розширюваність нашого підходу до PDTB та безпосередньо порівняти його з належним базовим рівнем серед попередніх досліджень щодо PDTB.

3. РОЗРОБКА АЛГОРИТМІЧНОГО ТА ПРОГРАМНОГО РІШЕННЯ ЗАДАЧІ ПОБУДОВИ СЕРВІСУ

3.1 Складові програмного рішення

BudgetAssistant – назва кінцевого програмного продукту. Його можна розділити на 3 складові – Базу даних, серверну частину (Back-end) та користувацький інтерфейс (Front-end).

База Даних

Розробку програмного продукту було розпочато з його основи - бази даних. В якості СКБД було обрано *MongoDB* - документо-орієнтовану систему керування базами даних з відкритим вихідним кодом. Цю СКБД було обрано через ідеальну інтеграцію з вибраними технологіями для реалізації серверної частини. Також MongoDB ефективно забезпечує зберігання великих бінарних об'єктів (якими являються банківські виписки) та підтримує зберігання документів в JSON-подібному форматі зі створенням індексів для атрибутів. Останнє надає можливість оптимізувати та пришвидшити отримання даних для UI частини програмного продукту.

Як можна побачити на рисунку 3.1, схема складається з чотирьох таблиць: “Users”, “BankConnections”, “Statements”, “ExchangeRates”.

Таблиця “Users” містить в собі всю необхідну інформацію про користувача. Поле “email” слугує унікальним індентифікатором користувача та використовується для логіну. “Passord” – містить зашифрований перед збереженням пароль користувача. Поле “dateOfBirth” потрібне для валідації віку користувача, так як доступ до рахунків у використаних банківських системах надається лише повнолітнім користувачам. Інші поля зберігають загальну інформацію про користувача (ім’я та номер телефону).

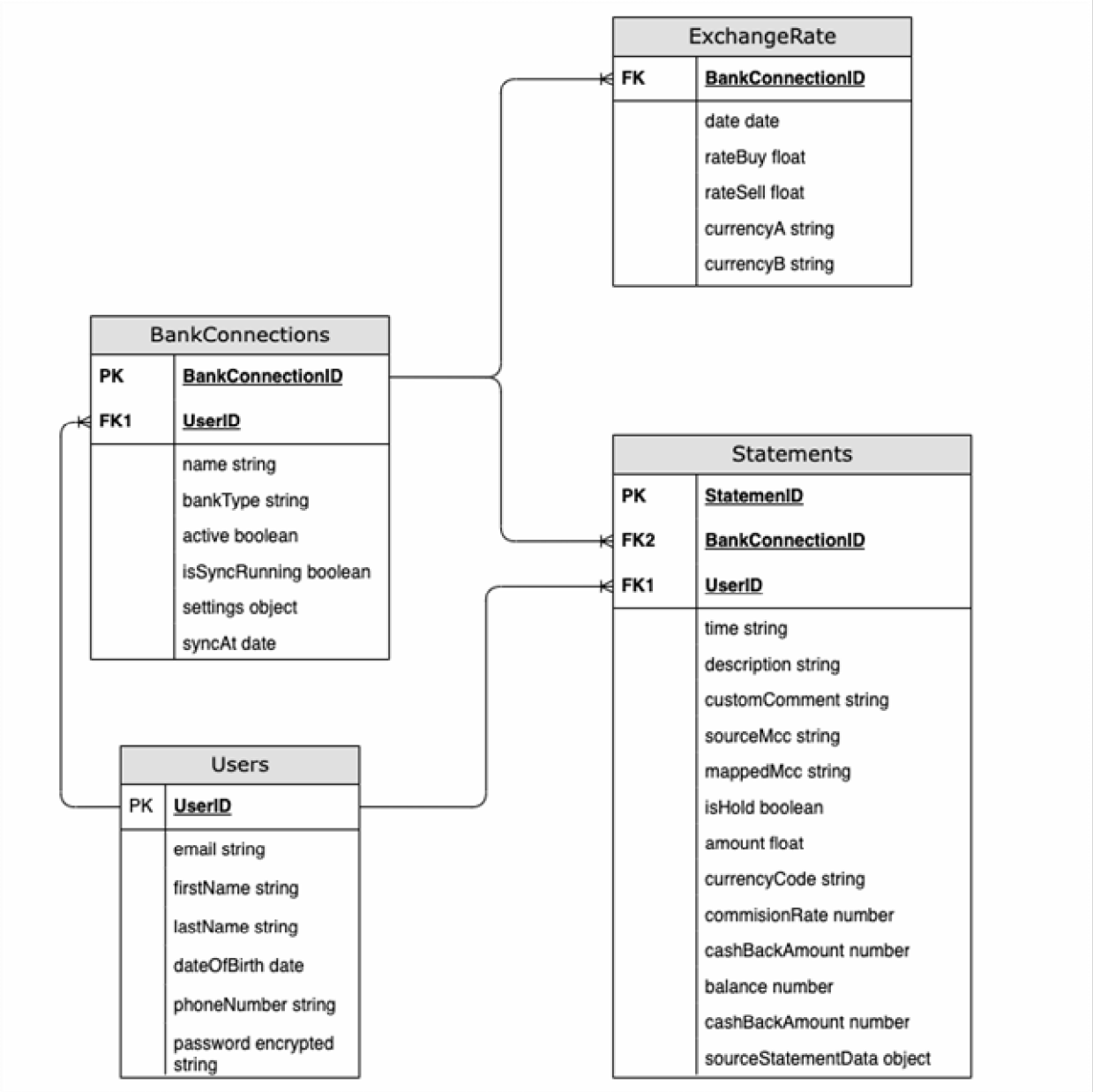


Рисунок 3.1 - Схема Бази Даних

Таблиця “BankConnections” існує для збереження інформації про підключені банківські рахунки. Містить у собі поле “settings” в якому зберігаються зашифрований токен для доступу по API та унікальний ідентифікатор обраного банківського рахунку. Також у цій таблиці зберігаються службові поля стану синхронізації виписок: “isSyncRunning” вказує чи запущено зараз синхронізацію та “active” для призупинення наступних синхронізацій за необхідності.

Таблиця “Statements” слугує для збереження банківських виписок. Містить собі усі головні дані транзакції. Наприклад: “sourceMcc” – код категорії платежу, “description” – опис операції від банку, “comment” – коментар до платежу, “balance” – баланс рахунку після операції, “amount” – сума (може бути як від’ємною, так і додатною).

Таблиця “ExchangeCourses” - потрібна для збереження історії курсів валют.

Базу даних було розміщено у докер контейнері що забезпечує можливість використовувати її локально та швидко розгорнути на віддаленій машині. При подальшому масштабуванні програмного рішення слід розглянути можливість використання хмарного сервісу Mongo Atlas.

3.2 Розробка серверної частини проекту

Основні переваги платформи node.js над іншими сучасними рішеннями:

- Висока продуктивність

Програмні рішення побудовані з використанням багатозадачності node.js отримують значну перевагу над іншими. На відміну від інших платформ, однопотокова архітектура, керована подіями, ефективно обробляє декілька одночасних запитів, не забиваючи оперативну пам’ять. Крім того, його цикл подій і неблокуючі операції вводу-виводу дозволяють виконувати код зі швидкістю що сильно впливає на загальну продуктивність програми.

- Підтримка Google

Node.js побудований на рушії Google Chrome V8 і написаний на C++. Він розбиває функції Javascript на машинні коди з високою ефективністю та швидкістю. Це важливо тому, що Google інвестує у свій механізм, щоб покращити продуктивність, безпеку та забезпечити чудову підтримку сучасних функцій JavaScript. Це безпосередньо впливає на продуктивність V8 і, зрештою, на Node.js.

- Легка масштабованість

Гіганти сучасного ринку, з постійно зростаючою базою споживачів, що налічує десятки мільйонів користувачів, розглядають Node.js як життєздатне рішення для масштабованості .

Платформа Node.js має безліч функцій, до прикладу кластерний модуль. Це полегшує балансування навантаження між кількома ядрами ЦП, що дозволяє легко досягати бажаних результатів за допомогою менших модулів, не витрачаючи процес RAM. Крім того, Node.js використовує неблокуючий механізм циклу подій, який забезпечує високу масштабованість і дозволяє серверу безперешкодно обробляти запити.

- Економічна ефективність

Node.js дає розробникам можливість писати код на стороні сервера на Javascript. Тепер є можливим використання однієї мови для реалізації користувацького інтерфейсу та серверної логіки. Це одна з найбільших переваг node.js, яка заощаджує багато часу, коштів та енергії для загальної розробки проекту.

Реалізація серверної частини

Для реалізації обрано мікросервісну архітектуру яка забезпечить легке масштабування у майбутньому. На даному етапі було реалізовано один сервіс “budget-api” та два допоміжні пакети для нього “monobank-api” та “budget-utils”.

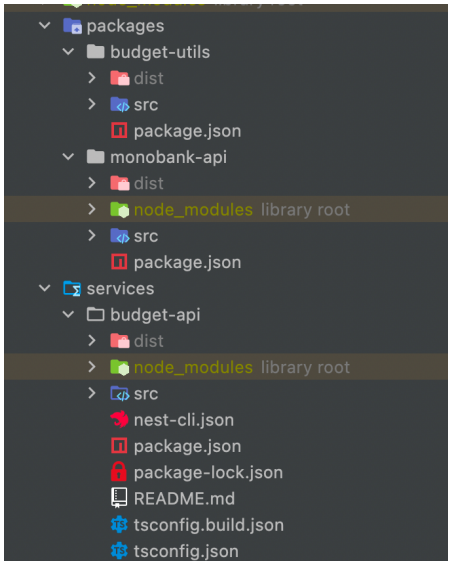


Рисунок 3.2 - Структура Проекту

“Budget API” сервіс

Головним завданням цього сервісу є реалізація API точок доступу для їх використання Front-end частиною.

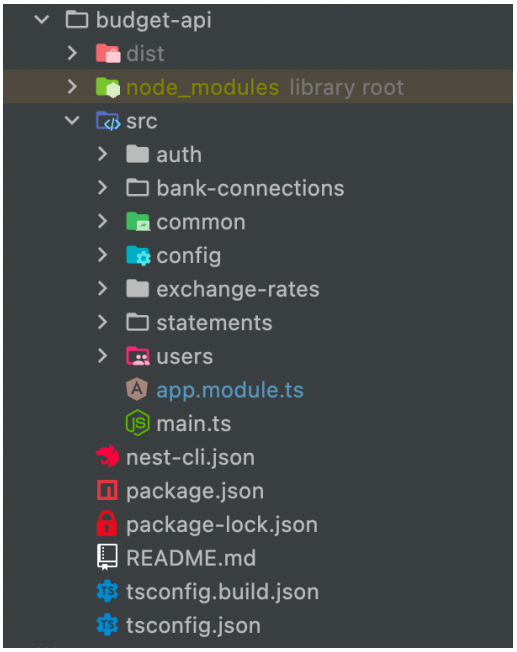


Рисунок 3.3 - Структура Budget-API сервісу

Сервіс було створено за допомогою Nest.js.

NestJS — фреймворк для створення серверних додатків Node.js. Це просте середовище з чіткою архітектурою і широкими можливостями, містить у собі все необхідне, щоб запустити проект, але при цьому підтримує інтеграцію будь-яких компонентів Express.JS, бібліотек і модулів TypeScript та JavaScript, а також багато іншого.

Сервіс складається з модулів – логічно виокремленого та поділеного функціоналу для окремих сутностей. Як можна побачити на рисунку, сервіс складається з таких модулей:

- **auth**: модуль що віповідає за безпеку та логіку авторизації;
- **bank-connections**: модуль містить логіку створення підключень до банківських рахунків;

- **exchange-rates**: модуль відповідає за збереження інформації про курси валют;
- **statements**: модуль що містить увесь функціонал банківських виписок, їх синхронізацію та обробку з подальшою фільтрацією для запитів з Front-end частини;
- **users**: модуль для керування користувачами
- **common**: модуль що містить загальні налаштування для всіх інших модулів, такі як: підключення до бази даних, форматування API відповідей, логування вхідних запитів;
-

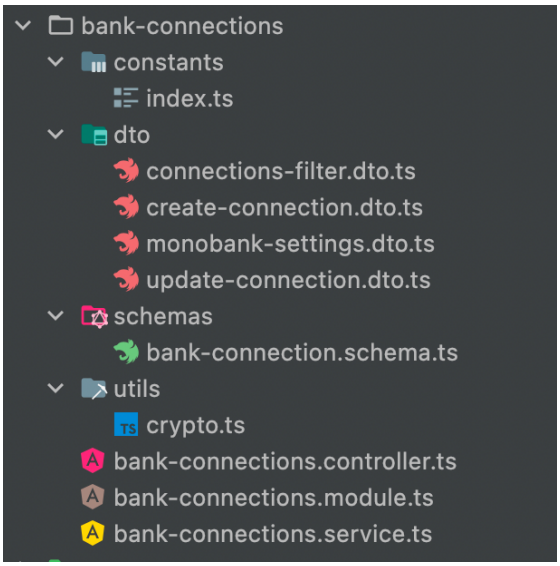


Рисунок 3.4 - Структура “bank-connections” модуля

Більш детально структуру модуля можна розглянути на прикладі “**bank-connections**” модуля:

В каталозі "DTO" знаходяться об’єкти передачі даних (Data Transfer Object). Data Transfer Object (DTO) — один із шаблонів проектування, який використовують для передачі даних між підсистемами програми, а в даному випадку між Front-end та Back-end частинами.

В каталозі "Schemas" зняходяться схеми колекцій бази даних. У даному випадку це тільки схема “bank-connection” колекції.

Файл контролерів містить у собі тільки опис точок доступу, та має наступний вигляд:

```
@ApiTags('bank-connections')
@Controller('bank-connections')
export class BankConnectionsController {
      constructor(private readonly service:
 BankConnectionsService) {}

  @ApiBearerAuth()
  @Get(':id')
        async findById(@Param('id', ValidateMongoId) id:
    string) {
    return this.service.findOneById(id);
  }
  @ApiBearerAuth()
  @Get()
        async      find(@Body()      connectionsFilterDto:
    ConnectionsFilterDto) {
    return this.service.find(connectionsFilterDto);
  }
  @ApiBearerAuth()
  @Post()
        async      create(@Body()      createConnectionDto:
    CreateConnectionDto) {
    return this.service.create(createConnectionDto);
  }
  @ApiBearerAuth()
  @Patch(':id')
  async update(
    @Param('id', ValidateMongoId) id: string,
    @Body() updateConnectionDto: UpdateConnectionDto,
  ) {
    return this.service.update(id, updateConnectionDto);
```

```
}}
```

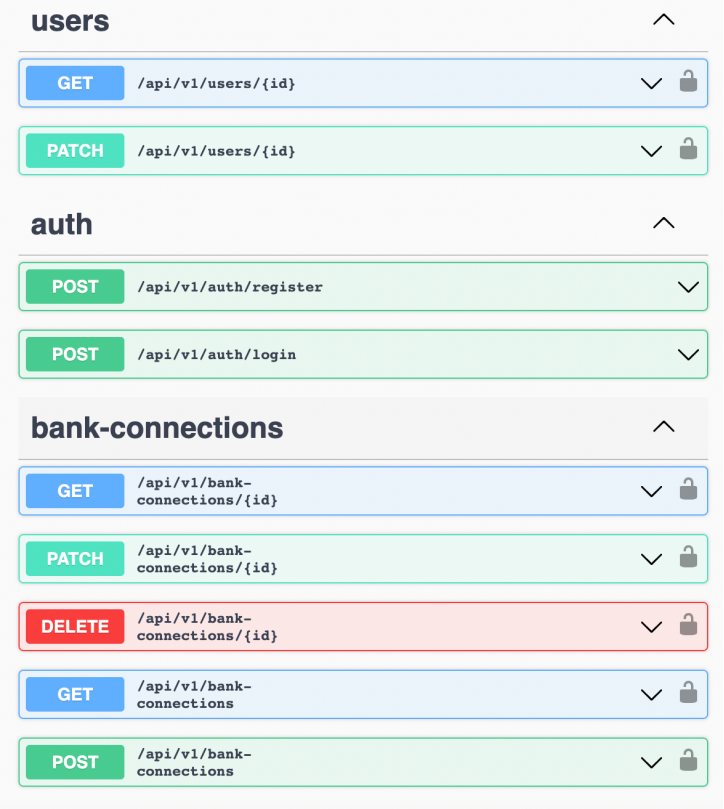


Рисунок 3.5 - Swagger документація

Файл сервісів містить логіку обробки API запиту та формування запиту до бази даних. Для прикладу реалізація методу оновлення одного “bank-connection” :

```
async update(id: string, updateConnectionDto:
UpdateConnectionDto) {
  const updatedConnection = await
this.connectionModel.findByIdAndUpdate(
    id, updateConnectionDto, { returnDocument: 'after' },
  );

  if (!updatedConnection) {
    throw new HttpException('Connection not found',
HttpStatus.BAD_REQUEST);
  }
```

```
        const mongoObject = getMongoObject(updatedConnection);
        return decryptConnection(mongoObject);
    }
```

ВИСНОВКИ

В дипломній роботі крім аналітичних переваг, які може дати вивчення поведінки клієнта, моделі та рамки можна ефективно використовувати для проектування та розробки інтерфейсів, які можуть додатково сприяти розвитку інтелекту. Щоб продемонструвати цю перевагу в конкретному контексті бізнес-аналітики, ми розробили і розробили інтерактивну систему візуалізації. Ця система надає керівникам проектів можливість переглядати, досліджувати й отримувати уявлення про обґрунтування попередніх змін, внесених у проекти. Основне обґрунтування нашого дизайну полягало в тому, що архівування причин і обґрунтування змін настільки ж важливо, як і самі рішення щодо змін для подальшого повторного використання. Проте розробка такої системи була б неможлива без базової моделі, яка може ефективно та ефективно виявляти обґрунтування внутрішніх корпоративних комунікацій.

Дослідження концепцій інтелекту, конфіденційності та переконання в соціальних взаємодіях вже давно проводяться в різних дисциплінах, таких як соціологія, психологія та антропологія. Ця робота робить внесок у ці галузі шляхом повторного вивчення деяких припущень і висновків щодо такої соціальної взаємодії в контексті соціальних мереж онлайн. Крім того, представлені нові висновки щодо структури та змісту масових соціальних взаємодій в Інтернеті, які вивчалися та спостерігалися в невеликих традиційних дослідженнях. Прогалини та закономірності, знайдені в літературі, запропоновані моделі та набори ознак, разом з їх оцінкою та аналізом, можуть дати нове розуміння цих дисциплін і запропонувати нові шляхи для майбутньої дослідницької діяльності. Нарешті, наші висновки також можуть бути корисними для розробки інтерфейсів і платформ наступного покоління, сприяючи сферам візуалізації інформації та взаємодії людини та комп’ютера в фінансових задачах зокрема.

ПЕРЕЛІК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

[1] A N R Hami & A Y Cheng 2002 A risk perception analysis on the use of electronic payment systems by young adult order 6(8.4) 6-7

[2] P Aigbe & J Akpojaro 2014 Analysis of security issues in electronic payment systems. (International journal of computer applications) 108(10)

[3] S k Antwi, K Hamza & S W Bavoh 2015 Examining the effectiveness of electronic payment system in Ghana: the case of e-ZWICH in the Tamale metropolis (Research Journal of Finance and Accounting) 6(2) 163-177

[4] C K Ayo & W I Ukpere 2010 Design of a secure unified e-payment system in Nigeria: A case study African Journal of Business Management) 4(9) 1753-1760

[5] Z Bezovski 2016 The Future of the Mobile Payment as Electronic Payment System (European Journal of Business and Management) 8(8) 127-132 IC-ELINVO IOP Conf. Series: Journal of Physics: Conf. Series 1140 (2018) 012033 IOP Publishing doi:10.1088/1742-6596/1140/1/012033

[6] I Bukhori 2018 Analisis Faktor-Faktor Yang Mempengaruhi dan Dampak Adopsi Mobile Payment Di Kalangan Merchant Case Study fasilitasi Mocash pada merchant di Kab. Pacitan (Master's thesis, Universitas Islam Indonesia)

[7] I Daştan & C Gürler C 2016 Factors affecting the adoption of mobile payment systems: An empirical analysis (EMAJ: Emerging Markets Journal) 6(1) 17-24

[8] S Preetha, V Divya 2016 Customers Trust on the Security of Online Payment System Using Credit/ Debit Cards with Reference to Banking Customers in Chennai (American International Journal of Research in Humanities Arts and Social Sciences)

[9] B Fung, M Molico & G Stuber 2014 Electronic money and payments: Recent developments and issues (No. 2014-2) Bank of Canada Discussion Paper

[10] T S Gandawati 2007 Analisis Proses Adopsi Electronic Payment System dengan menggunakan UTAUT model (Makasar: Universitas Gunadarma)

[11] A Hamza & A Shah 2014 Gender and mobile payment system adoption among students of tertiary institutions in Nigeria (International Journal of Computer and Information Technology) 3(1) 13-20

[12] K H Hanzaee & S Alinejad 2012 An investigation about customers perceptions of security and trust in e-payment systems among Iranian online consumers (Journal of Basic and Applied Scientific Research) 2(2) 1575-158

[13] T D Hascaryani 2013 Analisis Faktor Yang Mempengaruhi Preferensi dan Aksesibilitas Terhadap Penggunaan Kartu Pembayaran Elektronik (Jurnal Ilmiah Mahasiswa FEB) 2(2)

[14] A N Hidayanto, L S Hidayat, P I Sandhyaduhita & P W Handayani 2015 Examining the relationship of payment system characteristics and behavioural intention in e-payment adoption: a case of Indonesia (International Journal of Business Information Systems) 19(1) 58-86

[15] P O Igudia 2017 A Qualitative Evaluation of the Factors Influencing the Adoption of Electronic Payment Systems (SMEs) by SMEs in Nigeria (European Scientific Journal, ESJ) 13(31)

[16] S Junadi[a] 2015 A model of factors influencing consumer's intention to use e-payment system in Indonesia

[17] M A Kabir, S Z Saidin & A Ahmi 2015 Adoption of e-payment systems: a review of literature. In International Conference on E-Commerce pp. 112-20

[18] M Karismariyanti 2014 Perancangan Sistem Informasi Akuntansi Penjualan dengan Pembayaran Elektronik (e-Payment) (Jurnal Teknologi Informasi) 1(7) 238-244

[19] K Kaur & A Pathak 2015 E-payment system on e-commerce in India (International Journal of Engineering Research and Applications) 5(2) 79-87

[20] B U I Khan, R F Olanrewaju, A M Baba, A A Langoo & S Assad 2017 A compendious study of online payment systems: Past developments, present impact, and future considerations (International Journal of Advanced Computer Science and Applications) 8(5), 256-71

[21] M Maadi & M Javidnia 2016 Identification of factors influencing building initial trust in ecommerce (Iranian Journal of Management Studies) 9(3) 483-503

[22] M Maqableh, M T R E Masa, R O Shannak & K M Nahar 2015 Perceived Trust and Payment Methods: An Empirical Study of MarkaVIP Company (International Journal of Communications) Network and System Sciences 8(11) 409-427

[23] F Maroofi, R Hashemi & Z Nargesi 2012 Survey of Customers (Asian Journal of Business Management) 4(3) 275-285

[24] M Masihuddin, B U I Khan, M M U I Mattoo & R F Olanrewaju 2017 A survey on e-payment systems: elements, adoption, architecture, challenges and security concepts (Indian Journal of Science and Technology) 10(20)

[25] D Mathur 2017 A Survey of Awareness about Security in E-payment System (International Journal of Mechanical Engineering and Information Technology) 5(03) 1846-1850

[26] H Mulyasari, T T B Dan & A B M Wijaya Analisis Jenis Sistem Pembayaran Elektronik Dalam Transaksi E-Commerce di Indonesia IC-ELINVO IOP Conf. Series: Journal of Physics: Conf. Series 1140 (2018) 012033 IOP Publishing doi:10.1088/1742-6596/1140/1/012033 7

[27] D Niousha, B Masoud & B Hamid 2015 Factors influencing the adoption of electronic payment cards in urban micro-payments (Basic Research Journal of Business Management and Accounts) 4(2) 62-70

[28] E F Nwaolisa & E G Kasie 2012 Electronic retail payment systems: User acceptability and payment problems in Nigeria (Arabian Journal of Business and Management Review (OMAN Chapter)) 1(9) 111-123

[29] J Okifo & R Igbunu 2015 Electronic Payment System in Nigeria: Its Economic Benefits and Challenges (Journal of Education and Practice) 6(16) 56-62

[30] A Omotubora & S Basu 2018 Regulation for E-payment Systems: Analytical Approaches Beyond Private Ordering (Journal of African Law) 62(2), 281-313

[31] A Premchand & A Choudhry 2015 Future of Payments–ePayments (International Journal of Emerging Technology and Advanced Engineering) 110-115

[32] A M Qatawneh, F M Aldhmour& S M Alfugara 2015 The Adoption of Electronic Payment System (EPS) in Jordan: Case Study of Orange Telecommunication Company (Journal of Business and Management) 6(22) 139-148

[33] S Roy & I Sinha 2014 Determinants of customers' acceptance of electronic payment system in Indian banking sector–a study (International Journal of Scientific and Engineering Research) 5(1) 177-87

[34] S Roy and I Sinh 2017 Factors affecting Customers' adoption of Electronic Payment : an Empirical Analysis (IOSR Journal of Business and Management (IOSR-JBM))

[35] P H K Seno 2012 Analisis Persepsi Nasabah Atas Keamanan dan Kepercayaan dalam Sistem Epayments (Jurnal Ekonomi & Bisnis PNJ) 11(2)

[36] A S K S Shahazad, M H Khan & M Chandra 2012 A Review: Secure Payment System for Electronic Transaction (International Journal) 2(3)

БІБЛІОГРАФІЧНА ДОВІДКА

Тема дипломної роботи бакалавра: «Розробка рішення та сервісу контролю витрат для клієнтів українських банків за допомогою технологій Nest.js, React.js»

Обсяг пояснювальної записки: 81 аркуш

Дата закінчення дипломної роботи 14 червня 2022 р.

Підпис студента _____