

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КІЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»**

Факультет прикладної математики

Кафедра прикладної математики

«На правах рукопису»
УДК 330.46

«До захисту допущено»
Завідувач кафедри

_____ О.Р. Чертов
«___» 2015р.

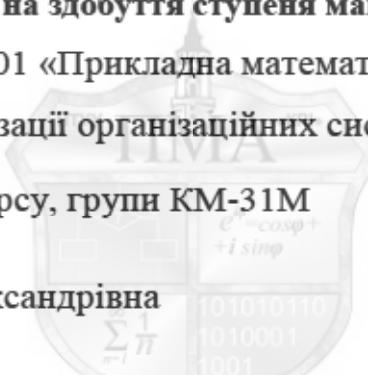
Магістерська дисертація

на здобуття ступеня магістра

зі спеціальності 8.04030101 «Прикладна математика»

на тему: Модель кластерізації організаційних систем

Виконала: студентка 2 курсу, групи КМ-31М



Оріщенко Анастасія Олександрівна

_____ (підпис)

Науковий керівник доцент, канд. техн. наук, ст. наук.
співроб. Маслянко П. П.

_____ (підпис)

Консультант із старший викладач Мальчиков В. В.

_____ (підпис)

Консультант зі старший викладач Темнікова О. Л.

_____ (підпис)

Рецензент професор, д-р техн. наук, проф.
Бірюк П.І.

_____ (підпис)

Засвідчую, що у цій магістерській дисертації
немає запозичень з праць інших авторів без
відповідних посилань.
Студент _____

**Національний технічний університет України
«Київський політехнічний інститут»**

Факультет прикладної математики

Кафедра прикладної математики

Рівень вищої освіти – другий (магістерський)

Спеціальність 8.04030101 «Прикладна математика»

ЗАТВЕРДЖУЮ
Завідувач кафедри
_____ О. Р. Чертов
(підпис)

«____» 2015 р.

**ЗАВДАННЯ
на магістерську дисертацію студентці
Оріщенко Анастасії Олександрівні**

1. Тема дисертації: «Модель кластеризації організаційних систем», науковий керівник дисертації Маслянко Павло Павлович, канд. техн. наук, ст. наук. співроб., затверджені наказом по університету від «20» березня 2015 року № 785-С.
2. Термін подання студентом дисертації: «18» червня 2015 р.
3. Об'єкт дослідження: організаційні системи (Орг.С.), види і класи Орг.С., класифікація та критерії класифікації Орг.С., класифікація Орг.С. за організаційно-правовою формою господарювання та за видами економічної діяльності, кластеризація Орг.С., концептуальні моделі Орг.С, засоби автоматизації кластерного аналізу.
4. Предмет дослідження: уніфікована модель кластеризації Орг.С. на основі методів кластерного аналізу.

5. Перелік завдань, які потрібно розробити:

- проаналізувати існуючи види, класи Орг.С., методи та засоби їх відління;
- розробити концептуальну та уніфіковану модель кластеризації Орг.С. на основі методів кластерного аналізу;
- дослідити математичні моделі кластеризації для реалізації уніфікованої моделі;
- розробити програмне забезпечення з використанням уніфікованої моделі кластеризації Орг.С., попередньо обравши математичну модель

6. Орієнтовний перелік ілюстративного матеріалу:

- схема існуючих класифікацій Орг.С. за різними критеріями;
- порівняльна таблиця існуючих засобів кластерного аналізу;
- діаграма компонентів моделі кластеризації Орг.С.;
- діаграми діяльності моделі кластеризації Орг.С.;
- діаграми діяльності моделі кластеризації Орг.С. у вигляді «плавальних доріжок».

7. Орієнтовний перелік публікацій:

- 17 міжнародна науково-технічна конференція «SAIT 2015»;
- VI наукова конференція магістрантів та аспірантів «Прикладна математика та комп’ютинг – ПМК’2015»;

8. Консультанти розділів дисертації

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Математичні моделі кластеризації Орг.С.	Темнікова О. Л., старший викладач		

9. Дата видачі завдання «25» жовтня 2013 р.

Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації	Примітка
1	Вибір напряму дослідження та узгодження тематики МД з керівником	15 вересня–30 жовтня 2013	
2	Грунтовне ознайомлення з предметною областю	30 жовтня 2013–15 лютого 2014	
3	Вивчення літератури, пошук додаткової інформації	15 лютого–1 вересня 2014	
4	Проведення дослідження, розроблення програмного забезпечення	1 вересня 2014–1 березня 2015	
5	Завершення роботи над основною частиною МД, переддипломна практика, робота над публікаціями	1 березня–1 травня 2015	
6	Оформлення текстової і графічної частини МД	1 травня–1 червня 2015	
7	Попередній захист МД	1 червня–15 червня 2015	

Студентка

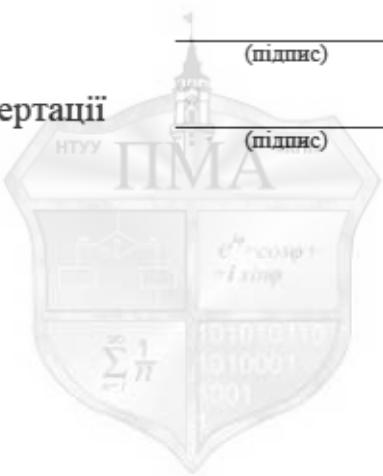
(підпис)

А. О. Оріщенко

Науковий керівник дисертації

(підпис)

П. П. Маслянко



РЕФЕРАТ

Актуальність теми. Організаційні системи (Орг.С.) є невід'ємною частиною сучасного життя. Вони існують в усіх його сферах: політичній, економічній, соціальній, культурній та інших.

Усі вони взаємодіють між собою для досягнення кращих результатів. Тому перед людиною часто виникають задачі пошуку потрібної підмножини Орг.С. за певними критеріями і формування кластерів Орг.С. для ефективного управління ними.

В ході аналізу існуючих методів виділення видів та класів Орг.С. було встановлено, що одним з основних підходів для вирішення цієї задачі є класифікація, яка дозволяє виділяти види та класи Орг.С. за однією ознакою.

Ще одним з методів, які вирішують цю задачу є кластерний аналіз. Він дозволяє за набором ознак, використовуючи різні методи, метрики, підходи для обчислення вагових коефіцієнтів виконувати кластеризацію Орг.С.

На сьогоднішній день не існує спеціальних засобів автоматизації процесу кластеризації Орг.С. Проте, існують програмні засоби, які дозволяють проводити кластеризацію об'єктів будь-якої природи, в тому числі і Орг.С. Такі засоби не враховують особливості Орг.С. для виділення кластерів.

Тому актуальною є задача розробки моделі кластеризації Орг.С., яка дозволить автоматизувати процес виділення класів та видів Орг.С. на основі обраних ознак.

Об'єктом досліджень є Орг.С., види і класи Орг.С., класифікація та критерії класифікації Орг.С., класифікація Орг.С. за організаційно-правовою формою господарювання та за видами економічної діяльності, кластеризація Орг.С., концептуальні моделі Орг.С, засоби автоматизації кластерного аналізу.

Предметом досліджень є уніфікована модель кластеризації Орг.С. на основі методів кластерного аналізу.

Мета дослідження: полягає у розробці концептуальної та уніфікованої моделі для автоматизації виділення класів Орг.С. за встановленими ознаками та формалізації процесів кластеризації.

Методи дослідження. В роботі використовуються методи математичного моделювання, методи системного аналізу, апарат кластерного аналізу.

Наукова новизна роботи полягає в наступному:

1. Вперше запропонована концептуальна та уніфікована модель кластеризації Орг.С., як опорний інструмент для розробника програмного забезпечення.
2. Розроблено компонентну та динамічну моделі кластеризації Орг.С., що чітко визначають складові будь-яких системи виділення видів та класів Орг.С. та загальний принцип їх роботи.
3. На основі уніфікованої моделі розроблено мобільний додаток для пошуку туристичного агентства за обраними ознаками.

Практична цінність отриманих в роботі результатів полягає в тому, що запропонована уніфікована модель дає змогу виконувати кластеризацію Орг.С. не за однією, а за цілим набором ознак, а також використовувати методи кластерного аналізу, які найкраще підходять для вирішення конкретної задачі. Така модель дозволяє автоматизувати процес виділення класів та видів Орг.С., а також реалізововувати програмні засоби для вирішення різних задач.

Апробація роботи. Основні положення і результати роботи були представлені та обговорювались на VII науковій конференції магістрантів та аспірантів «Прикладна математика та комп’ютинг» ПМК-2015 (Київ, 15-17 квітня 2015 р.) та опубліковані у збірнику тез доповідей, Просвіта, 2014, а також на 17-й Міжнародній науково-технічній конференції SAIT 2014 (Київ, 22-25 червня 2015 р.) та опубліковані у збірнику «Системний аналіз та інформаційні технології» ННК “ПСА” НТУУ “КП”, 2015.

Структура та обсяг роботи. Магістерська дисертація складається зі вступу, чотирьох розділів, висновків та додатків.

У вступі надано загальну характеристику роботи, виконано оцінку сучасного стану проблеми, обґрунтовано актуальність напрямку досліджень, сформульовано мету і задачі дослідження, показано наукову новизну отриманих результатів і практичну цінність роботи.

В першому розділі виконується огляд існуючих видів, класів Орг.С., методів та засобів їх виділення, а також подається їх порівняльна характеристика.

У другому розділі будується концептуальна модель кластеризації Орг.С., а також уніфікована модель у вигляді компонентного та динамічного представлень а також діаграми «плавальних доріжок», які є складовими уніфікованої моделі.

В третьому розділі виконується огляд методів кластерного аналізу, а саме метрик, правил зв'язку а також методу парних порівнень для обчисленн вагових коефіцієнтів ознак.

В четвертому розділі розглядається архітектура програмної реалізації мобільного додатку для пошуку туристичних агентств, обґрунтовується вибір методів кластерного аналізу.

У висновках проаналізовано отримані результати роботи.

У додатках наведено фрагменти коду програмного продукту, тестові дані для перевірки роботи програмного забезпечення, та слайди презентації для супроводу доповіді.

Робота виконана на 103 аркушах, містить 3 додатки та посилання на список використаних літературних джерел з 32 найменувань. У роботі наведено 13 рисунків та 7 таблиць.

Ключові слова: Орг.С., види і класи Орг.С., класифікація Орг.С., кластеризація Орг.С., уніфікована модель, компонентна модель, динамічна модель, кластерний аналіз.

РЕФЕРАТ

Актуальность темы. Организационные системы (Орг.С.) есть неотъемлимой частью современной жизни. Они существуют во всех ее сферах: политической, экономической, социальной, культурной и других. Они все взаимодействуют между собой для достижения лучших результатов. Поэтому перед человеком часто возникают задачи поиска нужного подмножества Орг.С. по определенным признакам и формирования кластеров Орг.С. для эффективного управления ими.

В ходе анализа существующих методов выделения видов и классов Орг.С. было установлено, что одним из методов для решения этой задачи есть классификация, которая позволяет выделять виды и классы Орг.С. по одному признаку.

Еще одним методом, которые решают эту задачу есть кластерный анализ. Он позволяет по набору признаков, используя разные методы, метрики, подходы к расчету весовых коэффициентов признаков выполнять кластеризацию Орг.С.

Сегодня не существует специальных средств автоматизации процесса кластеризации Орг.С. Однако, существуют программные средства, которые позволяют проводить кластеризацию объектов любой природы, в том числе и Орг.С. для выделения кластеров. Поэтому актуальна задача разработки методики кластеризации Орг.С., которая позволит автоматизировать процесс выделения классов и видов Орг.С. по определенным признакам

Объектом исследований являются Орг.С., виды и классы Орг.С., классификация и критерии классификации Орг.С., классификация Орг.С. по организационно-правовой форме и по виду экономической деятельности, кластеризация Орг.С., концептуальные модели Орг.С., средства автоматизации кластерного анализа.

Предметом исследований является унифицированная модель кластеризации Орг.С. на основе методов кластерного анализа.

Цель исследования: разработка концептуальной и унифицированной модели для автоматизации выделения классов Орг.С. по определенным признакам и формализация процессов кластеризации.

Методы исследования. В работе используются методы математического моделирования, методы системного анализа, аппарат кластерного анализа.

Научная новизна работы заключается в следующем:

1. Впервые предложена концептуальная и унифицированная модель кластеризации Орг.С., как опорный инструмент для разработчика .
2. Разработаны компонентная и динамическая модели кластеризации Орг.С., которые четко определяют составляющие любых системы выделения классов и видов Орг.С. и общий принцип их работы.
3. На основе унифицированной модели разработано мобильное приложение для поиска туристических агентств по определенным признакам.

Практическая ценность полученных в работе результатов заключается в том, что предложенная унифицированная модель позволяет выполнять кластеризацию Орг.С. не по одному, а по набору признаков, а так же использовать методы кластерного анализа, которые лучше всего подходят для решения конкретных задач. Такая модель позволяет автоматизировать процесс выделения классов и видов Орг.С. так же реализовывать программные продукты для решения разных задач

Апробация работы. Основные положения и результаты работы были представлены и обсуждались на VII научной конференции магистрантов и аспирантов «Прикладная математика и компьютеринг » ПМК-2015 (Киев , 15-17 апреля 2015 г.) и опубликованы в сборнике тезисов докладов , Просвіта , 2014 , а также на 17-й Международный научно-технический конференции SAIT 2015 (Киев , 22-25 июня 2014 г.) и опубликованы в сборнике «Системный анализ и информационные технологии» УНК " ИПСА " НТУУ " КПІ ", 2015 .

Структура и объем работы. Магистерская диссертация состоит из введения, четырех глав, заключения и приложений.

Во введении дана общая характеристика работы, выполнена оценка текущего состояния проблемы, обоснована актуальность направления исследований, сформулированы цель и задачи исследований, показана научная новизна полученных результатов и практическая ценность работы .

В первом разделе выполняется обзор существующих видов и классов Орг.С., методов и средств их выделения а также дается их сравнительная характеристика .

В втором разделе строится концептуальная модель кластеризации Орг.С., а так же унифицированная модель компонентная и динамическая модели, а так же унифицированная модель в виде структурного и динамического представления и диаграммы “плавающих дорожек”.

В третьем разделе выполняется обзор методов кластерного анализа, а именно метрик, правил связи, а так же метода попарных сравнений для расчета весовых коэффициентов признаков.

В четвертом разделе рассматривается архитектура программной реализации мобильного приложения для поиска туристических агентств, обосновывается выбор методов кластерного анализа.

В заключении проанализированы полученные результаты работы .

В приложениях приведены фрагменты кода программного продукта, тестовые данные для программного обеспечения и слайды презентации, для сопровождения доклада.

Работа выполнена на 103 листах, содержит 3 приложения и ссылки на список использованных литературных источников из 32 наименований. В работе приведены 13 рисунков и 7 таблиц.

Ключевые слова: Орг.С., виды и классы Орг.С., классификация Орг.С., кластеризация Орг.С., унифицированная модель, компонентная модель, динамическая модель, кластерный анализ.

ABSTRACT

Theme urgency. Organizational systems (Org.S.) is an integral part of modern life. They exist in all its spheres: political, economic, social, cultural and others.

They interact with each other to achieve the best results. So before people often exist the problem of finding relevant subset Org.S. according to certain criteria and clustering of Org.S. for effective management.

The analysis of existing methods and tools of partition Org.S. by classes and types of it was found that one of the main approaches to solving this problem is the classification that allows to select types and classes of Org.S. by one attribute.

Another method that solve this problem is cluster analysis. It allows by the range of attributes, using different methods, metrics and approaches for calculating the weighting coefficients to perform the clustering of Org.S.

Today there are no special tools of automation the process of clustering of Org.S. However, there are software tools that allow to cluster objects of any nature, including Org.S. These tools do not address specific Org.S. for clusters.

So the urgent task is a developing the model of Org.S. clustering, that will automate the process of selection of classes and types of Org.S. based on selected attributes.

The object of research is Org.S., types and classes of Org.S., classification and classification criteria of Org.S., classification of Org.S. by organizational form and management of economic activity, clustering of Org.S. conceptual model of Org.S, tools for cluster analysis.

The subject of research is unified model of Org.S. clustering, based on the methods of cluster analysis.

Research objective: is to develop the conceptual and unified model to automate the selection of classes Org.S. by attributes and formalization of clustering task.well. Providing a reduction of energy consumption for climate control system.

Research methods. Methods of mathematical modeling, methods of system analysis, cluster analysis.

Scientific novelty consists in the following:

1. Firstly proposed conceptual model and unified model for clustering Org.S., as reference tool for software developers.
2. A component model and dynamic clustering of Org.S. clearly defining the components of any system of allocation of classes and types of Org.S. and the general principle of their work.
3. Based on the unified model developed the mobile application for finding travel agency on selected attributes.

Practical value of the results is that the proposed unified model enables Org.S. clustering not by one, but by a set of attributes, and let to use different methods of cluster analysis, for the best solve of a particular problem. This model automates the process of selection of classes and types Org.S. and to develop software for solving various problems.

Approbation. The main points and the results were presented and discussed at the VII Scientific Conference of master students and graduate students "Applied Mathematics and Computing" AMC 2015 (Kyiv, 15-17 April 2014) and published in the book of handouts, Prosvita , 2015, as well as 17th International Scientific and Technical conference SAIT 2015 (Kiev, 22-25 June 2014) and published in the "System Analysis and information Technologies" ESC " IASA " "KPI", 2015 .

Structure and content of the thesis. The master's thesis consists of the introduction, five chapters, conclusions and appendixes.

The introduction presents the general description of the work and gives an estimation of current state of the problem , shows the urgency of research areas. The goal and objectives of research shows scientific novelty of the results and the practical value of the work .

In the first chapter performs a review of existing types, classes of Org.S., methods and tools, and served their comparative characteristics.

In the second chapter a conceptual model Org.S. clustering is created, as well as a unified component model, dynamic representations and diagrams "swimming lanes" that are parts of a unified model.

In the third chapter a review of cluster analysis methods, such as metrics, rules of communication and the method for calculating weight coefficient of attributes performs.

In the fourth chapter examines the architecture of the program implementation of mobile applications for finding travel agencies justified the selection of cluster analysis.

In the conclusion the obtained results are analyzed.

The appendixes present fragments of software code, test data for program and presentation slides to accompany the report.

The thesis is presented on 103 pages, it contains 3 appendixes and 32 references to the used information sources. 13 drawings and 7 tables are also presented.

Keywords: Org.S., types and classes of Org.S., Org.S. classification, clustering of Org.S., unified model, dynamic model, cluster analysis.

ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ, УМОВНИХ ПОЗНАЧЕНЬ, ТЕРМІНІВ.....	16
ВСТУП	18
1 ОГЛЯД ІСНУЮЧИХ ВІДІВ, КЛАСІВ ОРГ.С., МЕТОДІВ ТА ЗАСОБІВ ЇХ ВІДІЛЕННЯ	20
1.1 Види Орг.С.....	20
1.2 Класифікація Орг.С.....	21
1.2.1 Класифікація за організаційно-правовою формою	23
1.2.2 Класифікація за видом економічної діяльності	24
1.3 Кластеризація Орг.С.....	25
1.4 Засоби автоматизації кластеризації Орг.С.....	29
Висновки до розділу	31
2 МОДЕЛЬ КЛАСТЕРИЗАЦІЇ ОРГ.С	33
2.1 Структурне представлення та формалізація поняття Орг.С.....	33
2.2 Концептуальна модель кластеризації Орг.С.....	34
2.3 Уніфікована модель кластеризації Орг.С	35
2.3.1 Уніфікована мова моделювання UML	36
2.3.2 Структурне представлення моделі кластеризації Орг.С.....	40
2.3.3 Динамічне представлення моделі кластеризації Орг.С	41
2.3.4 Діаграма “Плавальні доріжки” кластеризації Орг.С	42
Висновки до розділу	45
3 МАТЕМАТИЧНІ МОДЕЛІ КЛАСТЕРИЗАЦІЇ ОРГ.С	46
3.1 Кластерний аналіз, як метод виділення класів Орг.С.....	46
3.2 Класифікація методів кластерного аналізу	47
3.2.1 Ієрархічні методи кластерного аналізу	48
3.2.2 Обчислення вагових коефіцієнтів ознак методом парних порівнянь..	53
3.3 Формалізація постановки задачі кластеризації Орг.С	49
3.3.1 Метрики	50
3.3.2 Обчислення вагових коефіцієнтів ознак методом парних порівнянь..	53
3.3.3 Методи виділення кластерів Орг.С	57
Висновки до розділу	59
4 ПРОГРАМНА РЕАЛІЗАЦІЯ УНІФІКОВАНОЇ МОДЕЛІ КЛАСТЕРИЗАЦІЇ ОРГ.С.....	60
4.1 Застосування уніфікованої моделі для реалізації мобільного додатку “Your Travel Agency”	60
4.2 Вибір методів кластерного аналізу для кластеризації туристичних агентств.....	61
4.3 Архітектура програмного забезпечення	62
4.3.1 Діаграма класів	62
4.3.2 Концепція “Модель - вид - контроллер”	63
4.4 Інтерфейс користувача.....	65
4.5 Результати роботи програми	67

Висновки до розділу	67
ВИСНОВКИ	68
СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ	70
ДОДАТОК А Фрагменти коду програмного забезпечення	74
ДОДАТОК Б Тестові дані мобільного додатку	86
ДОДАТОК В Ілюстративний матеріал	90



ПЕРЕЛІК СКОРОЧЕНЬ, УМОВНИХ ПОЗНАЧЕНЬ, ТЕРМІНІВ

ДИНАМІЧНА МОДЕЛЬ – це абстрактне графічне відображення основних фізичних властивостей технічного об'єкта і характеристик взаємодії із зовнішнім середовищем, що відображає поведінку об'єкта у часі.

КЛАСИФІКАЦІЯ ВИДІВ ЕКОНОМІЧНОЇ ДІЯЛЬНОСТІ (КВЕД) – нормативний документ, який установлює основи для підготовлення та поширення статистичної інформації за видами економічної діяльності [1].

КЛАСИФІКАЦІЯ ОРГАНІЗАЦІЙНО-ПРАВОВИХ ФОРМ ГОСПОДАРЮВАННЯ (КОПФГ) – класифікатор, за допомогою якого органи державної влади (в першу чергу, органи державної статистики) класифікують юридичні особи за організаційно-правовими формами.

КОМПОНЕНТНА МОДЕЛЬ – статична структурна модель, що показує розбиття програмної системи на структурні компоненти та зв'язки (залежності) між компонентами. В якості компонентів можуть виступати файли, бібліотеки, модулі, виконувані файли, пакети і т. п.

ОРГАНІЗАЦІЙНА СИСТЕМА (Орг.С.) – це об'єднання автономних підприємств, організацій або структурних підрозділів з економічної, соціальної чи державної сфери діяльності, які функціонують під централізованим керівництвом і вирішують спільні завдання [2].

УНІФІКОВАНА МОВА МОДЕЛЮВАННЯ (UML) – уніфікована мова моделювання, використовується у парадигмі об'єктно-орієнтованого програмування. Є невід'ємною частиною уніфікованого процесу розробки програмного забезпечення. UML є мовою широкого профілю, це відкритий стандарт, що використовує графічні позначення для створення абстрактної моделі системи, називаної UML-моделлю [3].

УНІФІКОВАНА МОДЕЛЬ – абстрактна модель системи, що відображає характер основних процесів, компонентів і зв'язків, а також описує поведінку

системи у часі. Є аналогом реальної системи, створеної за допомогою обраної мови моделювання.

INTEGRATED DEVELOPMENT ENVIRONMENT (IDE) – система програмних засобів, яку використовують програмісти для розробки програмного забезпечення.

INTERNATIONAL STANDARD INDUSTRIAL CLASSIFICATION OF ALL ECONOMIC ACTIVITIES (ISIC) – це міжнародна довідка класифікації виробничої діяльності, основною метою якої є запровадження категорій діяльності, які можуть бути використані для збору статистичної інформації та подання звітності відповідно до виду діяльності [4].

IPHONE OPERATING SYSTEM (IOS) – це власницька мобільна операційна система від Apple що використовується тільки на власних продуктах Apple.

NOMENCLATURE GÉNÉRALE DES ACTIVITÉS ÉCONOMIQUES DANS LES COMMUNAUTÉS EUROPÉENNES (NACE) – система для класифікації видів економічної діяльності, розроблена на основі ISIC [5].

ВСТУП

Організаційні системи (Орг.С.) є невід'ємною частиною сучасного життя. Вони існують в усіх його сферах: політичній, економічній, соціальній, культурній та інших. В них переплітаються інтереси особистості та груп, технології та інновації, стимули та обмеження, жорстка дисципліна та вільна творчість, нормативні вимоги та неформальні ініціативи.

Орг.С. розвиваються, коли мають чітку стратегію та ефективно використовують ресурси. Вони перестають існувати, коли не можуть виконувати свої задачі. Розуміння сутності Орг.С. та закономірностей їх розвитку дає можливість управлюти ними, ефективно використовувати їх потенціал, освоювати нові технології їх діяльності [6].

Сьогодні існує дуже велика кількість Орг.С. [7]. Усі вони взаємодіють між собою для досягнення кращих результатів. Тому перед людиною часто виникають задачі пошуку потрібної підмножини Орг.С. за певними критеріями і формування кластерів Орг.С. для ефективного управління ними. Великі об'єми інформації унеможливлюють вирішення їх вручну, та потребують автоматизації процесу розв'язання.

Одним з основних способів виділення класів та видів Орг.С. є класифікація. Існують класифікації Орг.С. за різними критеріями, деякі з них затверджені у нормативних документах [1,4,5,8]. Проте, основним недоліком такого підходу є врахування лише однієї ознаки Орг.С.

Більш гнучким способом, що дозволяє врахувати декілька ознак для виділення класів та видів Орг.С., є кластерний аналіз. Велика кількість різноманітних методів дозволяє обрати найкращий з них для вирішення конкретної задачі [9].

На сьогоднішній день не існує спеціальних засобів автоматизації процесу кластеризації Орг.С. Проте, існують програмні засоби, які дозволяють

проводити кластеризацію об'єктів будь-якої природи, в тому числі і Орг.С. Такі засоби не враховують особливості Орг.С. для виділення кластерів.

Тому актуальною є задача розробки моделі кластеризації Орг.С., яка дозволить автоматизувати процес виділення класів та видів Орг.С. на основі обраних ознак.



1 ОГЛЯД ІСНУЮЧИХ ВІДІВ, КЛАСІВ ОРГ.С., МЕТОДІВ ТА ЗАСОБІВ ЇХ ВІДЛЕННЯ

1.1 Види Орг.С.

Орг.С. виникли давно, та з розвитком суспільства вони розростались, ускладнювались, набували більшого значення в житті людини. За найпростішим формулюванням, Орг.С. – це група людей, що діють разом для досягнення певної мети. Для успішного досягнення цієї мети, діяльність повинна координуватися. Тому Орг.С. можна розглядати як групу людей, діяльність яких свідомо координується для досягнення певної мети.

Існують важливі фундаментальні відмінності, які призводять до виділення двох видів Орг.С (рисунок 1.1).



Рисунок 1.1 – Види Орг.С.

Неформальна Орг.С. – група людей, які регулярно взаємодіють один з одним, що виникла спонтанно.

Формальна Орг.С. – Орг.С., що є юридичною особою, мета діяльності якої закріплена в установчих документах, а функціонування – в нормативних актах, угодах та положеннях, що регламентують права та відповідальність кожного з її учасників.

Формальні Орг.С. поділяються на комерційні та некомерційні (Рис.1).

Комерційні Орг.С. - Орг.С., діяльність яких направлена на систематичне отримання прибутку від виконання робіт, продажу товарів, використання майна чи надання послуг.

Некомерційні Орг.С. - Орг.С., основною метою діяльності яких не є отримання та розподіл прибутку між учасниками.

Неформальні Орг.С. також існують в межах формальних Орг.С., за виключенням маленьких (у формі міжособистих неформальних груп). Вони суттєво впливають на різні сторони управління формальними Орг.С. Кажучи про неформальні Орг.С., їх, зазвичай, так і називають. Надалі термін "Орг.С." буде використовуватися по відношенню до формальних Орг.С.

1.2 Класифікація Орг.С.



Основним методом систематизації Орг.С. є класифікація за одним або декількома критеріями.

Існують наступні класифікації Орг.С. за різними критеріями [10]:

- за формою власності
 - приватні
 - державні
 - муніципальні
- за способом фінансування
 - бюджетні
 - позабюджетні
- за цільовим призначенням
 - виробництво продукції
 - виконання робіт
 - надання послуг

- за широтою виробничого профілю
 - спеціалізовані
 - диверсифіковані
- за органами правління
 - урядові
 - неурядові
- за цільовою аудиторією
 - громадські
 - господарські
- за поєднанням науки та виробництва
 - наукові
 - виробничі
 - науково-виробничі
- за кількістю стадій виробництва
 - одноетапні
 - багатоетапні
- за розміщенням
 - на одній території
 - на одній географічній точці
 - на різних географічних точках

Усі вищеперелічені класифікації поділяють усю множину Орг.С. на дві або три групи. Найповнішою є класифікація Орг.С. за організаційно-правовою формою.

1.2.1 Класифікація за організаційно-правовою формою

Організаційно-правова форма господарювання — форма здійснення господарської (зокрема, підприємницької) діяльності з відповідною правовою основою, яка визначає характер відносин між засновниками (учасниками), режим майнової відповідальності по зобов'язаннях підприємства (організації), порядок створення, реорганізації, ліквідації, управління, розподілу одержаних прибутків, можливі джерела фінансування діяльності, тощо [8].

Нормативним документом, що затверджує класифікацію Орг.С. за організаційно-правовою формою є КОПФГ.

Даний класифікатор призначено для використовування органами державного управління та іншими користувачами для обліку, збирання й обробляння статистичної та адміністративної інформації щодо державної реєстрації, аналізування та узагальнювання результатів економічної діяльності суб'єктів господарської (зокрема підприємницької) діяльності, ведення державних реєстрів: Єдиного державного реєстру підприємств та організацій України (ЄДРПОУ), Єдиного державного реєстру юридичних осіб та фізичних осіб-підприємців, Реєстру корпоративних прав держави та інших [8].

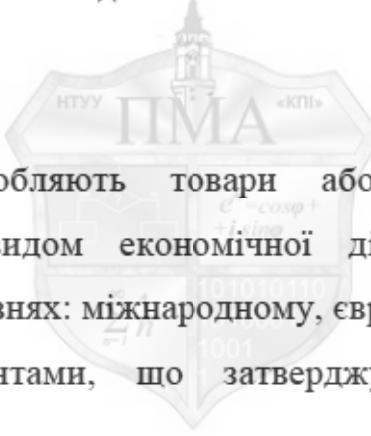
Кожна позиція класифікатора структурно складається з тризначного ідентифікаційного коду та назви організаційно-правової форми господарювання. Ідентифікаційний код будується, використовуючи серійно-порядковий метод кодування. Початкові коди серій (100, 200, 300, 400, 500, 600, 800 та 900) та їхні назви, використовують для узагальнювання об'єктів класифікації і не призначені для безпосереднього кодування суб'єктів господарської та іншої діяльності [8]. Усього в класифікаторі виділяють 8 серій:

1. підприємства;
2. господарські товариства;

3. кооперативи;
4. організації (установи, заклади);
5. об'єднання підприємств (юридичних осіб);
6. відокремлені підрозділи без статусу юридичної особи;
7. об'єднання громадян, профспілки, благодійні організації та інші подібні організації;
8. інші організаційно-правові форми.

До складу 8 серій входить 66 класів Орг.С.

1.2.2 Класифікація за видом економічної діяльності



Орг.С., які виробляють товари або надають послуги можна прокласифікувати за видом економічної діяльності. Така класифікація здійснюється на трьох рівнях: міжнародному, європейському та національному. Нормативними документами, що затверджують таку класифікацію є відповідно:

1. Міжнародна стандартна галузева класифікація всіх видів економічної діяльності (ISIC).
2. Класифікація видів економічної діяльності Європейського Спітовариства (NACE).
3. Національний класифікатор України. Класифікація видів економічної діяльності (КВЕД). Інші національні класифікатори країн.

Дані класифікації співвідносяться за своїм змістом між собою: європейська з міжнародною, а національна – з європейською та міжнародною. NACE — похідна класифікація від ISIC i, задля задоволення конкретних потреб користувачів у ЄС, вона більш деталізована на нижчому рівні. Український

КВЕД створено на основі NACE (Rev. 2) для адаптації статистичних даних з даними ЄС та ООН.

Порівняння класифікацій Орг.С. наведено у таблиці 1.1.

Таблиця 1.1 – Порівняння класифікацій Орг.С. за видом економічної діяльності

	Секцій	Розділів	Класів	Співвідноситься з	Рівень
ISIC	21	88	419	-	Міжнародна
NACE	21	88	615	ISIC	Європейска
КВЕД	21	88	615	ISIC, NACE	Українська

Кожна з цих класифікацій поділяє усю множину Орг.С. на секції, розділи, групи та класи. Кількість секцій та розділів в усіх трьох класифікаціях є однаковою. Кількість груп і класів в класифікації NACE та КВЕД є більшою у порівнянні з ISIC, адже вони є її розширенням.

1.3 Кластеризація Орг.С.

Одним зі способів виділення класів Орг.С. є кластерний аналіз. У [11] зазначено, що: “Кластерный анализ – один из способов классификации объектов по их признакам”. Кластерний аналіз є багатовимірною статистичною процедурою, яка “... передбачає збір даних, що зберігають інформацію про вибірку об’єктів, та упорядкування об’єктів у порівнянно однорідні групи” [12].

Методи кластерного аналізу використовуються як інструментарій для: визначення пріоритетних напрямів розвитку регіонів; прогнозування соціально-економічних явищ; класифікації емітентів цінних паперів; вибору локальних стратегій; типологізації інвестиційного потенціалу лісозаготівельних підприємств; дослідження рівня вмотивованості працівників; прогнозування банкрутства промислових підприємств; вирішення проблем розвитку лісового господарства; сегментації ринку; аналізу фінансово-економічного стану підприємств будівельної галузі тощо [13].

Слід зауважити, що на сьогоднішній день не існує уніфікованої моделі кластеризації Орг.С. Проте, у літературі описують застосування методів кластерного аналізу для вирішення конкретних практичних проблем, які виникають в межах Орг.С. Результати дослідження використання кластерного аналізу Орг.С. для вирішення різних задач наведені у таблиці 1.2.

Таблиця 1.2 – Порівняння використання кластерного аналізу для вирішення різних задач

Мета дослідження	Об'єкт кластеризації	Методи кластерного аналізу	Результати досліджень	Літ. поси-лання
Аналіз витрат на створення нової продукції (СНП) на машинобудівних підприємствах.	20 машинобудівних підприємств Харківської області.	Ієрархічний агломеративний метод Уорда з використанням евклідової метрики.	Було отримано три групи: з високим, середнім та низьким рівнем витрат на СНП. У межах кожного кластера був оцінений інтегральний показник рівня інноваційної та фінансово-господарської діяльності підприємств.	[14]

Продовження таблиці 1.2

Мета дослідження	Об'єкт кластеризації	Методи кластерного аналізу	Результати дослідень	Літ. посилання
Оцінка розподілу пріоритетів сфер діяльності машинобудівних підприємств з метою визначення найоптимальнішого варіанта розподілу коштів між цими сферами для ефективного проведення диверсифікації діяльності товаровиробників.	10 провідних підприємств машинобудування Харківської області.	Метод аналізу ієрархій.	Отримано 2 кластери. Зроблені рекомендації для правильного розподілу коштів за сферами, на основі досвіду підприємств що потрапили в один кластер.	[15]
Спрощення процесу асортиментного планування мережі магазинів, для зменшення витрат на реалізацію процесу управління.	Магазини роздрібних мереж.	Модифікований метод k-середніх з урахуванням експертних оцінок можливості групування об'єктів в один кластер.	Запропоновано вирішення завдання кластеризації за наявності кількох метрик і суперечливих критеріїв оцінки ефективності використання різних розподілів під час асортиментного планування в торгівельних мережах.	[16]
Оцінка стратегічного протистояння підприємств-конкурентів на вітчизняному фармацевтичному ринку.	10 провідних підприємств фармацевтичного ринку.	Метод ближнього сусіда з використанням евклідової метрики.	Отримано три кластери, на основі яких зроблено висновки щодо гостроти конкуренції між підприємствами.	[17]

Продовження таблиці 1.2

Дослідження використання методу кластерного аналізу, що об'єднує різні процедури, які використовуються для проведення групування схожих між собою об'єктів – малих промислових підприємств однієї галузі.	Малі промислові підприємства однієї галузі.	Метод k-середніх.	Виявлено основних пріоритетних конкурентів для досліджуваного підприємства, що сприяє спрощенню проведення заходів щодо аналізу та діагностування стану підприємств-конкурентів.	[18]
Сегментування ринку за групами споживачів, за основними конкурентами та за параметрами продукції.	Підприємства шкіряної промисловості.	Агломера-тивний ієрархічний алгоритм класифікації з використанням евклідової відстані.	Встановлено, що на ринку шкіряних матеріалів можна виділити два сегменти, з дешевою та дорогою ціною сировини.	[12]
Дослідження питання адаптації підприємств машинобудування до умов Світової організації торгівлі (СОТ).	Підприємства машинобудування України.	Метод незваженого попарного середнього.	Запропоновано методичний підхід до розробки політики підвищення конкурентоспроможності підприємств машинобудування України, який ґрунтуються на використанні кластерного аналізу та дає змогу визначити однорідні групи підприємств, для кожної з яких запропоновані уніфіковані стратегії економічних перетворень.	[19]

1.4 Засоби автоматизації кластеризації Орг.С.

На сьогоднішній день не існує уніфікованих програмних засобів для класетризації Орг.С. Проте, в багатьох системах є модулі чи пакети з реалізацією методів кластерного аналізу. Порівняльні характеристики найпопулярніших з них, наведені у таблиці 1.3.

Таблиця 1.3 – Порівняльний аналіз існуючих засобів кластеризації

Метрики	Методи зв'язку	Графічне представлення результатів	Простота використання	Ціна
MATLAB. Statistics and Machine Learning Toolbox				
Евклідова	Близнього сусіда	У вигляді дендрограм, 2D та 3D графіків	Для досвідчених користувачів	Від 1250\$
Нормалізована евклідова	Дальнього сусіда			
Махalanобіса	Попарного середнього			
Манхеттеновська	Центроїдний			
Мінковського	Уарда			
Косинусна	k-середніх			
Кореляційна				
Хеммінга				
Джаккарда				
Інші				

Продовження таблиці 1.3

STATISTICA. Multivariate Exploratory Techniques				
Евклідова	Близнього сусіда	Дендограми, двовходові діаграми об'єднань, графічне представлення схеми об'єднання, діаграма середніх при використанні методу k-середніх	Для початківців та досвідчених користувачів.	Від 1910\$
SPSS Statistics				
Евклідова	Близнього сусіда	У вигляді дендограми	Для початківців та досвідчених користувачів.	Від 2530\$

Як видно з таблиці 1.3, сьогодні існують програмні засоби, у яких реалізована більшість методів кластерного аналізу. Проте, такі програмні засоби відрізняються дорогою ціною, а значить вони не доступні широкому колу користувачів. Також, деякі з них потребують спеціальних навиків та знань для їх використання.

Висновки до розділу

У розділі проведено огляд існуючих видів, класів, методів, та засобів для вирішення задачі виділення груп Орг.С., проаналізовано літературні джерела та нормативні документи. Проаналізовано застосування кластерного аналізу для вирішення різних задач, розглянуто можливості та цінові пропозиції програмних засобів, що включають в себе модулі чи пакети з реалізацією методів кластерного аналізу.

Встановлено:

1. Одним з основних підходів виділення груп Орг.С. є класифікація. Більшість існуючих класифікацій поділяють множину Орг.С. на 2-3 групи за одним або двома критеріями. У нормативному документі [8] наведена класифікація Орг.С. за організаційно-правовою формою, виділено 8 секцій та 66 класів. У нормативних документах [1,4,5] затверджено поділ Орг.С. за видом економічної діяльності. Усього виділено 21 секція 88 розділів та від 419 до 615 класів.
2. Кластеризація Орг.С. на сьогоднішній день є задачею, що досліджена недостатньо. В літературних джерелах знайдено приклади використання кластерного аналізу для вирішення різноманітних задач. Проте, уніфікованої моделі кластеризації Орг.С. на сьогоднішній день не існує.

3. Основним призначенням існуючих класифікацій є облік, збирання й обробляння статистичної та адміністративної інформації. Методи кластерного аналізу використовують для вирішення задач ефективного управління Орг.С., прогнозування соціально-економічних явищ, вибору локальних стратегій та багато інших. Отже, кластерний аналіз, на відміну від класифікації, дозволяє вирішувати значно ширше коло задач.
4. На сьогоднішній день не існує універсальних засобів, які вирішують задачу кластеризації Орг.С. Проте, існуючі системи мають модулі або пакети з реалізацією більшості методів кластерного аналізу. Такі системи є занадто дорогими, що робить їх недоступними для широкого кола користувачів.



2 МОДЕЛЬ КЛАСТЕРИЗАЦІЇ ОРГ.С.

2.1 Структурне представлення та формалізація поняття Орг.С.

Для створення концептуальної моделі кластеризації Орг.С. необхідно чітко розуміти, що собою представляє сутність Орг.С. Структурне представлення сутності “Орг.С.” і рівнів її моделей пропонується на рисунку 2.1 [20].

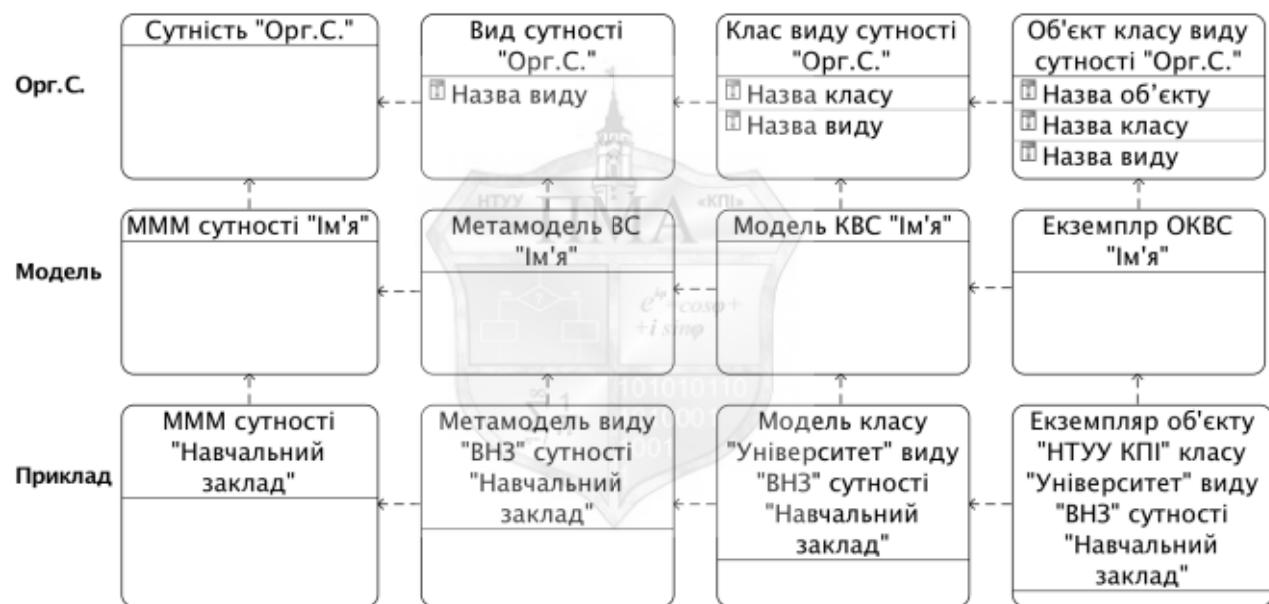


Рисунок 2.1 – Структурне представлення сутності “Орг.С.” і рівнів її моделей

Орг.С. представлені чотирма рівнями моделювання. Рівень метамоделі є сутністю “Організаційна система”, рівень метамоделі - видом сутності “Організаційна система”, рівень моделі – класом виду сутності “Організаційна система”, рівень екземпляру – об’єктом класу виду сутності “Організаційна система”.

Наприклад, для сутності “Навчальний заклад” одним з видів є “Вищий навчальний заклад”, одним з класів – “Університет”, а реальним об’єктом є “НТУУ КПІ”.

Усі сутності структурного визначення поняття “Організаційна система” поєднані відношенням конкретизації [21]. Таке представлення відображає вже існуючу класифікацію Орг.С. згідно [1,4,5,8].

Для формалізації поняття Орг.С. використаємо аналітичне визначення поняття “система” $S = (E, R)$, як множини сутностей E та відношень між ними $R \subseteq E^n$, які необхідні і достатні для існування системи [22]. Для системи та/або її елементів можна визначити множину ознак. Ознака $f = (k, v)$ це те, що описує елемент, та складається з двох частин: визначення ознаки k та значення ознаки v для даного елементу, тобто $f : E \cup R \rightarrow \{f = (k, v) | k \in K, v \in V\}$. Множина K визначає всі можливі назви ознак. Множина значень V може об’єднувати в собі елементи різної природи: скаляр, вектор, формулу або деякий формальний вираз, та ін.

Об’єкт, як останній рівень представлення сутності “Організаційна система” визначається множиною ознак $F(S)$, які притаманні сутності. Проте, також характеризується множиною своїх ознак $F(O)$, тобто $F(S) \subseteq F(O)$.

Таке структурне формальне представлення поняття Орг.С. дозволяє розробити концептуальну та уніфіковану модель кластеризації Орг.С.

2.2 Концептуальна модель кластеризації Орг.С.

Одним із способів виділення класів об’єктів та видів класів об’єктів є кластерний аналіз. Структурне представлення організаційних систем дозволяє виділити концептуальну модель кластеризації Орг.С. (рисунок 2.2).



Рисунок 2.2 – Концептуальна модель кластеризації Org.C.

Дана концептуальна модель дозволяє виділяти класи та види Org.C., у відповідності з вищеведеним структурним представленням Org.C.

Порядок проведення кластеризації наступний:

- об'єднання об'єктів у класи на основі їх ознак;
- формалізація класів, тобто виділення ознак отриманих класів;
- об'єднання класів у види на основі їх ознак.

Наступним рівнем представлення процесу кластеризації Org.C. є уніфікована модель.

2.3 Уніфікована модель кластеризації Org.C.

Уніфікована модель кластеризації Org.C. складається зі структурного та динамічного представлення, які графічно представляються діаграмою компонентів та діаграмою діяльності. Поєднанням цих двох представлень є діаграма “Плавальні доріжки”, яка найкраще відображає увесь процес кластеризації Org.C.

Для реалізації уніфікованої моделі використовується уніфікована мова моделювання UML версії 2.4.1, а в якості засобу – Visual Paradigm 12.1.

2.3.1 Уніфікована мова моделювання UML

UML - це мова для візуалізації, специфікації, конструювання та документування артефактів програмних систем.

Словник мови UML включає три види будівельних блоків:

1. сутності;
2. відносини;
3. діаграми.

Сутності - це абстракції, що є основними елементами моделі. Відносини зв'язують різні сутності; діаграми групують сукупності сутностей, що становлять інтерес.

В UML є чотири типи сутностей:

1. структурні;
2. поведінкові;
3. що групують;
4. анотаційні.

Сутності є основними об'єктно-орієнтованими блоками мови. З їхньою допомогою можна створювати коректні моделі. Нижче детально розглянуті сутності, які будуть використовуватись при розробці уніфікованої моделі.

Структурні сутності - це іменники в моделях, записаних мовою UML. Як правило, вони являють собою статичні частини моделі, що відповідають концептуальним або фізичним елементам системи. Існує сім різновидів структурних сутностей.

Клас (Class) - це опис сукупності об'єктів із загальними атрибутами, операціями, відношеннями та семантикою. Клас реалізує один або кілька інтерфейсів. Графічно клас зображується у вигляді прямокутника, у якому звичайно записані його ім'я, атрибути й операції.

Інтерфейс (Interface) - це сукупність операцій, які визначають сервіс (набір послуг), надаваний класом або компонентом. Таким чином, інтерфейс описує видиму ззовні поведінку елемента. Інтерфейс може представляти поведінку класу чи компонента повністю або частково; він визначає тільки специфікації операцій (сигнатури), але ніколи - їхньої реалізації. Графічно інтерфейс зображується у вигляді кола, під яким пишеться його ім'я. Інтерфейс рідко існує сам по собі - звичайно він приєднується до реалізуючого його класу або компонента.

Компонент (Component) - це фізична замінна частина системи, що відповідає деякому набору інтерфейсів і забезпечує його реалізацію. У системі можна зустріти різні види встановлюваних компонентів, такі як COM+ або JavaBeans, а також компоненти, що є артефактами процесу розробки, наприклад, файли вихідного коду. Компонент, як правило, являє собою фізичне поєднання логічних елементів, таких як класи, інтерфейси й кооперації. Графічно компонент зображується у вигляді прямокутника із вкладками, що містить звичайно тільки ім'я.

У мові UML визначено чотири типи відношень:

1. залежність;
2. асоціація;
3. узагальнення;
4. реалізація.

Ці відношення є основними з'єднувальними будівельними блоками в UML і застосовуються для створення коректних моделей.

Залежність (Dependency) - це семантичне відношення між двома сущностями, при якому зміна однієї з них, незалежної, може вплинути на семантику іншої, залежної. Графічно залежність зображується у вигляді прямої пунктирної лінії, часто зі стрілкою, що може містити мітку.

Асоціація (Association) - структурне відношення, що описує сукупність зв'язків; зв'язок - це з'єднання між об'єктами. Різновидом асоціації

є агрегування (Aggregation) - так називають структурне відношення між цілим і його частинами. Графічно асоціація зображується у вигляді прямої лінії (іноді, що завершується стрілкою, або утримуючою міткою), поруч із якої можуть бути присутнім додаткові позначення, наприклад кратність і імена ролей.

Узагальнення (Generalization) - це відношення "спеціалізація/узагальнення", при якому об'єкт спеціалізованого елемента (нащадок) може бути підставлений замість об'єкта узагальненого елемента (батька або предка). Таким чином, нащадок (Child) успадковує структуру й поведінку свого батька (Parent). Графічно відношення узагальнення зображується у вигляді лінії з незафарбованою стрілкою, що вказує на батька. Нарешті, реалізація (Realization) - це семантичне відношення між класифікаторами, при якому один класифікатор визначає "контракт", а інший гарантує його виконання.

Відношення реалізації зустрічаються між інтерфейсами й реалізуючими їх класами або компонентами. Відношення реалізації зображується у вигляді пунктирної лінії з незафарбованою стрілкою, як щось середнє між відношеннями узагальнення та залежності. Чотири описаних елементи є основними типами відношень, які можна включати в моделі UML.

Діаграма в UML - це графічне подання набору елементів, зображене найчастіше у вигляді зв'язного графа з вершинами (сущностями) і ребрами (відношеннями). Діаграми малюють для візуалізації системи з різних точок зору. Діаграма - у деякому сенсі, одна з проекцій системи. Як правило, за винятком найбільш тривіальних випадків, діаграми дають згорнуте подання елементів, з яких складена система. Той самий елемент може бути присутнім у всіх діаграмах, або тільки в декількох (найпоширеніший варіант), або не бути присутнім у жодній (дуже рідко). Теоретично діаграми можуть містити будь-які комбінації сущностей і відносин. На практиці, однак, застосовується порівняно невелика кількість типових комбінацій, що відповідають п'яти найбільш уживаним видам, які становлять архітектуру програмної системи [3].

Таким чином, в UML виділяють дев'ять типів діаграм:

1. діаграми класів;
2. діаграми об'єктів;
3. діаграми прецедентів;
4. діаграми послідовностей;
5. діаграми кооперації;
6. діаграми станів;
7. діаграми дій;
8. діаграми компонентів;
9. діаграми розгортання.

Вище наведено неповний список діаграм, застосовуваних в UML. Інструментальні засоби дозволяють генерувати і інші діаграми, але дев'ять перерахованих зустрічаються на практиці найчастіше [3].

Нижче детально описано діаграми, які будуть використовуватися для реалізації уніфікованої моделі кластеризації Орг.С.

Діаграма діяльності – візуальне представлення графу діяльностей. Граф діяльностей є різновидом графу станів скінченного автомата, вершинами якого є певні дії, а переходи відбуваються по завершенню дій [3].

Компонентна діаграма – діаграма реалізації, що моделює фізичні аспекти об'єктно-орієнтованих систем. Компонентна діаграма показує організацію набору компонентів і залежності між ними.

Діаграма класів (class diagram) служить для представлення статичної структури моделі системи в термінології класів об'єктно-орієнтованого програмування. На цій діаграмі показують класи, інтерфейси, об'єкти й кооперації, а також їхні відносини.

2.3.2 Структурне представлення моделі кластеризації Орг.С.

Розроблена компонентна модель кластерного аналізу Орг.С. складається з наступних компонентів (рисунок 2.3): інтерфейс користувача, база даних, система управління, розрахунок вагових коефіцієнтів ознак, розрахунок міри подібності об'єктів, виділення кластерів на основі правила зв'язку, аналіз результатів [23].

Компоненти представляють собою частини системи, які можуть бути куплені, розроблені, модифіковані, оновлені незалежно один від іншого. Компоненти зв'язані між собою за допомогою наданих або необхідних інтерфейсів.

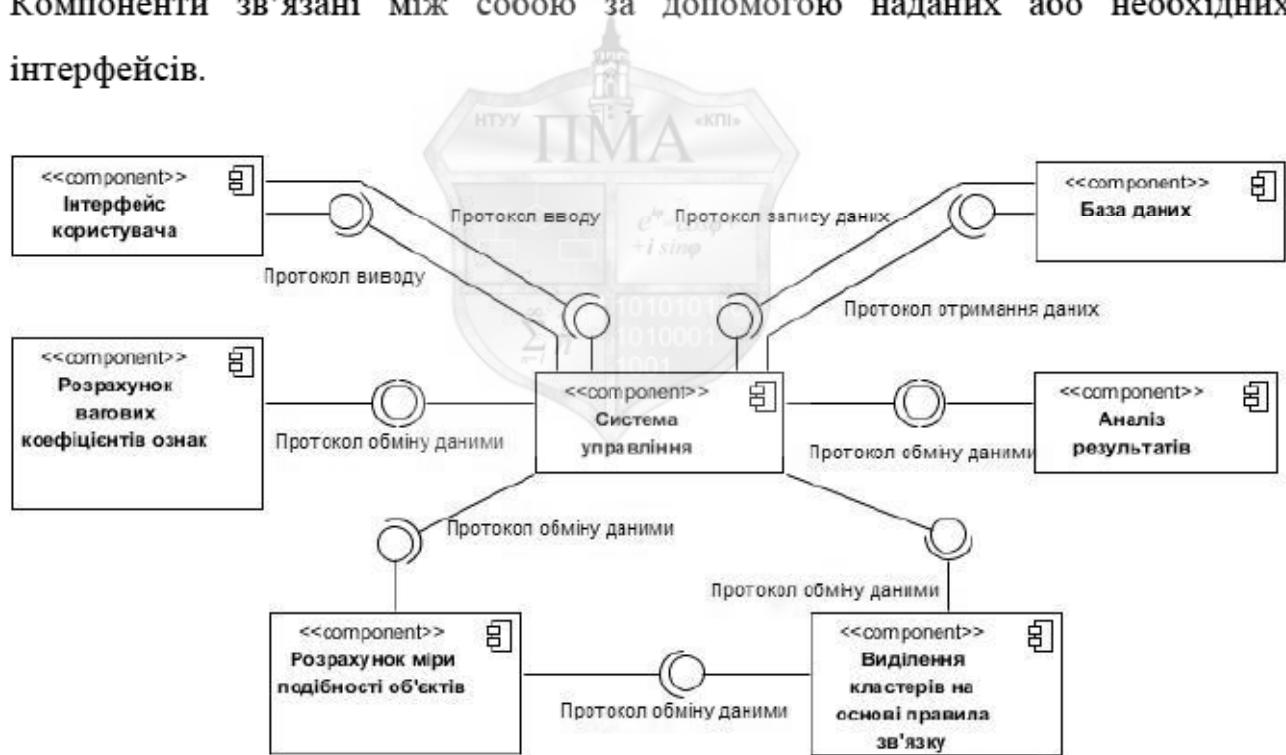


Рисунок 2.3 – Структурне представлення моделі кластеризації Орг.С. Діаграма компонентів

Специфікація діаграми компонентів:

- компонент «Інтерфейс користувача» – призначений для взаємодії з користувачем. Забезпечує введення вхідних параметрів задачі і виведення результатів;
- компонент «База даних» – призначений для зберігання даних про Орг.С., а саме ознак Орг.С. та їх значень для відповідних об'єктів кластеризації а також результатів роботи програми;
- компонент «Система управління» – призначений для організації взаємодії між модулями, зокрема для передачі даних між ними;
- компонент «Розрахунок вагових коефіцієнтів ознак» – призначений для обчислення або задання вагових коефіцієнтів ознак;
- компонент «Розрахунок міри подібності об'єктів» – призначений для нормалізації значень ознак та обчислення «відстані» між об'єктами за обраною метрикою. У результаті формується матриця відстаней;
- компонент «Виділення кластерів на основі правила зв'язку» – формує кластери на основі одного з правил зв'язку. Вибір правила виконується в залежності від задачі, яка вирішується.
- компонент «Аналіз результатів» – призначений для вибору необхідного рівня кластеризації, аналізу та корегування отриманих результатів.

2.3.3 Динамічне представлення моделі кластеризації Орг.С.

Розроблена діаграма діяльності (рисунок 2.4) відображає послідовність виконання кластерного аналізу Орг.С.

Кластерний аналіз відбувається за наступним алгоритмом:

- з бази даних завантажуються дані для користувальського інтерфейсу, а саме інформація про об'єкти кластеризації, їх ознаки та значення а також інша додаткова інформація;
- завантажується інтерфейс користувача;
- користувач обирає дані для початку процесу кластеризації. Ними можуть бути самі об'єкти кластеризації, їх ознаки та значення, а також, інші дані, в залежності від задачі;
- користувач задає вагові коефіцієнти ознак, або дані, на основі яких проводиться розрахунок вагових коефіцієнтів. Даний крок є необов'язковим та може бути пропущений;
- обчислюється відстань між об'єктами, на основі обраної метрики;
- виконується алгоритм кластеризації, що виконує розбиття об'єктів на класи. Зупинка виконується в залежності від обраного правила зупинки;
- результати виводяться користувачу;
- виконується аналіз отриманих результатів. Якщо розбиття є недостатнім або не задовольняє користувача, процедура повторюється, починаючи з вибору ознак та значень. В іншому випадку результати зберігаються у базу даних і процес кластеризації завершується.

2.3.4 Діаграма “Плавальні доріжки” кластеризації Орг.С.

Результатом поєднання компонентної діаграми та діаграми діяльності є діаграма “Плавальні доріжки” (рисунок 2.5). Кожній доріжці відповідає окремий компонент компонентної діаграми. На доріжках показано які дії виконують компоненти в процесі кластеризації Орг.С.

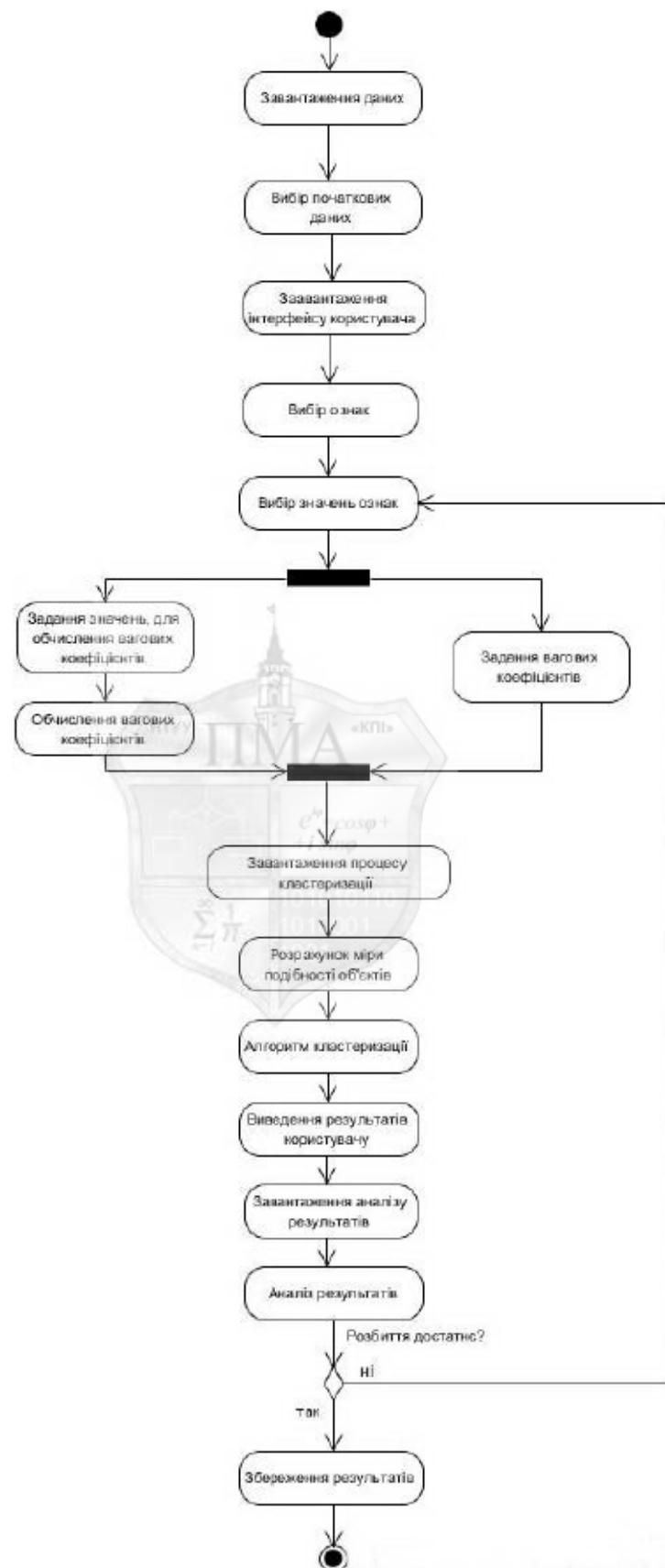


Рисунок 2.4 – Динамічне представлення моделі кластеризації Орг.С. Діаграма діяльності

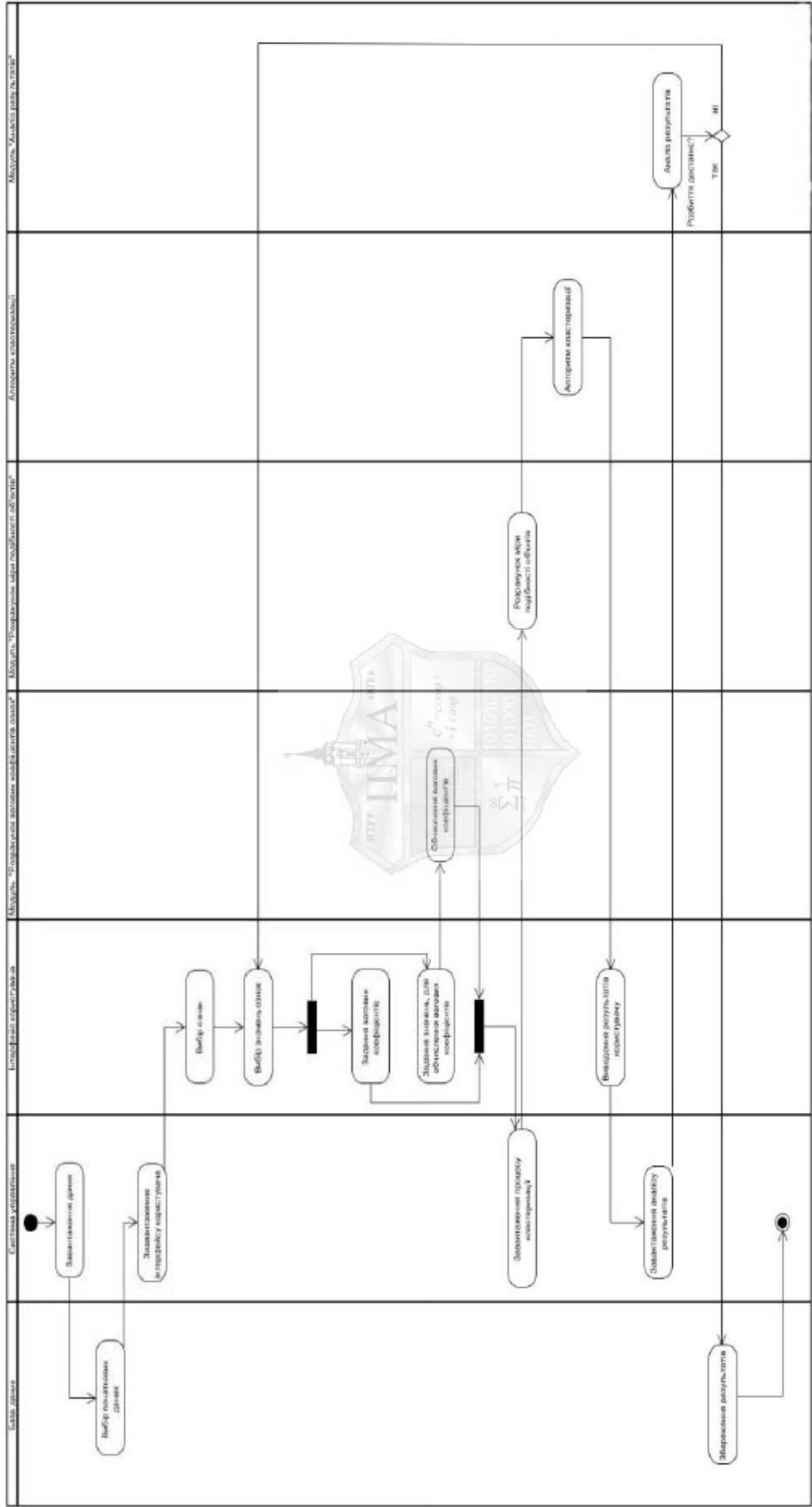


Рисунок 2.5 – Діаграма “Плавальний кластеризації” Орг.С.

Висновки до розділу

У розділі розроблене структурне представлення сутності Орг.С. та рівнів її моделей, а також виконана формалізація поняття Орг.С. На їх основі побудована концептуальна модель кластеризації Орг.С. У якості наступного рівня представлення розроблена уніфікована модель кластеризації Орг.С., що дає однозначне абстрактне представлення про складові процесу кластеризації, їх зв'язки та поведінку.

За допомогою уніфікованої мови моделювання UML побудовано наступний перелік діаграм:

- діаграма компонентів;
- діаграма діяльності;
- діаграма “Плавальні доріжки”.

3 МАТЕМАТИЧНІ МОДЕЛІ КЛАСТЕРИЗАЦІЇ ОРГ.С.

3.1 Кластерний аналіз, як метод виділення класів Орг.С.

Кластерний аналіз з'явився порівняно недавно – у 1939 р. Його запропонував вчений К. Тріон. Дослівно термін "кластер" в перекладі з англійської "cluster", означає гроно, згусток, пучок, група.

Особливо бурхливий розвиток кластерного аналізу відбувся у 60-х роках минулого століття. Передумовами цього були поява швидкісних комп'ютерів та визнання класифікації фундаментальним методом наукових досліджень.

Кластерний аналіз – це метод багатомірного статистичного дослідження, до якого належать збір даних, що містять інформацію про вибіркові об'єкти, та упорядкування їх в порівняно однорідні, схожі між собою групи.

Отже, сутність кластерного аналізу полягає у здійсненні класифікації об'єктів дослідження за допомогою численних обчислювальних процедур. В результаті цього утворюються "кластери" або групи дуже схожих об'єктів. На відміну від інших методів, цей вид аналізу дає можливість класифікувати об'єкти не за однією ознакою, а за декількома одночасно. Для цього вводяться відповідні показники, що характеризують певну міру близькості за всіма класифікаційними параметрами.

Мета кластерного аналізу полягає в пошуку наявних структур, що виражається в утворенні груп схожих між собою об'єктів – кластерів. Водночас його дія полягає й у привнесенні структури в досліджувані об'єкти. Це означає, що методи кластеризації необхідні для виявлення структури в даних, яку нелегко знайти при візуальному обстеженні або за допомогою експертів [24].

Формальне представлення поняття Орг.С., що подано вище, описує Орг.С. як сутність, якій притаманні ознаки та значення. Це дозволяє застосувати кластерний аналіз для вирішення задачі виділення класів та видів Орг.С.

3.2 Класифікація методів кластерного аналізу

Загальноприйнятої класифікації методів кластерного аналізу не існує, проте, можна виділити групи підходів [25].

- 1) Вирогіднісний підхід. Основна ідея підходу полягає в тому, що кожний об'єкт відноситься до одного класу з усієї множини. Деякі автори, (наприклад А.І. Орлов) вважають, що дана група взагалі не відноситься до кластеризації та протиставляє її під назвою “дискримінація”, тобто вибір віднесення об'єкту до одного з відомих груп.
- 2) Підходи на основі систем штучного інтелекту. До цієї групи входять такі методи як метод нечіткої кластеризації С-середніх (C-means), нейронна мережа Кохенона та генетичний алгоритм.
- 3) Логічний підхід. Побудова дендрограми відбувається за допомогою дерева розв'язків.
- 4) Теоретико-графовий підхід. До такого підходу належать методи упорядкування даних, візуалізація яких забезпечується за допомогою графів: дендрограм, кореляційних плеяд, діаграм Чекановського.
- 5) Ієрархічний підхід. В цьому випадку припускається наявність вкладених груп (кластерів різного порядку). Алгоритми в свою чергу поділяються на агломеративні та дівізівні.
- 6) Інші методи, що не увійшли в попередні групи.

Згідно вищепереліченого формального представлення Орг.С., а також моделі кластеризації Орг.С. найкращим з існуючих підходів для кластеризації Орг.С. є ієрархічний. Він дозволить виділити види та класи Орг.С. а також інші рівні їх представлення, якщо буде така необхідність.

3.2.1 Ієрархічні методи кластерного аналізу

Суть ієрархічних методів кластерного аналізу полягає в послідовному об'єднанні менших кластерів у більші (агломеративні методи) або навпаки, роз'єднанні більших кластерів на менші (дивізимні методи).

Ієрархічні агломеративні методи. Ця група методів характеризується послідовним об'єднанням вихідних елементів та відповідним зменшенням числа кластерів. На початку роботи алгоритма усі об'єкти є окремими кластерами. На першому кроці найбільш подібні об'єкти об'єднуються у кластер. На наступних кроках об'єднання продовжується доти, доки всі об'єкти не будуть становити один кластер.

Ієрархічні дивізимні методи. Ці методи є протилежністю агломеративним методам. На початку роботи алгоритму усі об'єкти належать до одного кластеру, який на наступних кроках ділиться на менші кластери. У результаті формується послідовність груп, що розщеплюються.

На рисунку 3.1 показано принцип роботи агломеративних та дивізимних методів у вигляді дендограми.

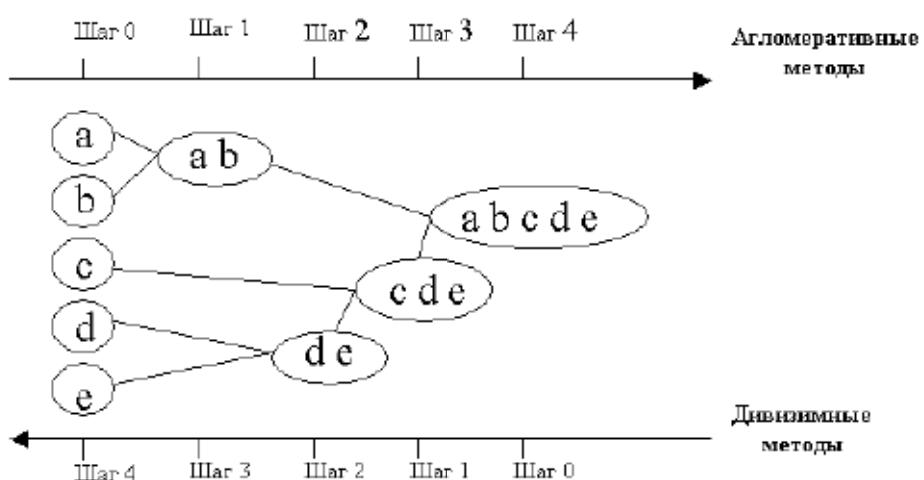


Рисунок 3.1 – Дендрограма агломеративних і дивізимних методів

Загальний алгоритм ієрархічних дивізимних методів має такі кроки:

- вибір об'єктів для кластеризації;
- вибір ознак за якими буде здійснюватись кластеризація та нормалізація їх значень за необхідності;
- вибір вагових коефіцієнтів ознак;
- вибір метрики для обчислення відстаней між об'єктами;
- обчислення матриці відстаней між об'єктами;
- вибір та застосування правила об'єднання об'єктів у кластери.

3.3 Формалізація постановки задачі кластеризації Орг.С.

Нехай $X^n = \{x_1, \dots, x_n\}$ – множина Орг.С., Y – множина номерів груп (кластерів). Задана функція відстані між об'єктами $d(x_i, x_j)$, $x_i, x_j \in X^n$. Необхідно розбити множину Орг.С. X^n на підмножини, які не перетинаються (кластери), так, щоб кожний кластер складався з об'єктів близьких за метрикою d , а об'єкти різних кластерів при цьому істотно відрізнялися. При цьому кожному об'єкту $x_i \in X^n$ присвоюється номер кластеру y .

Алгоритм кластеризації – це функція $a: X \rightarrow Y$, яка будь-якому об'єкту $x_i \in X^n$ ставить у відповідність номер кластеру $y \in Y$.

3.3.1 Метрики

Метрики для обчислення відстані між об'єктами, які використовуються на практиці відрізняються великом різноманіттям властивостей, проте, можна сформулювати загальні вимоги, яким повинна задовольняти міра подібності:

$$\begin{cases} d_{ij} \geq 0 \\ d_{ij} = d_{ji} \\ d_{ii} = 0 \end{cases} \quad (3.1)$$

Вибір метрики для кластеризації відбувається в першу чергу з урахуванням типу ознак. Ознаки можуть бути якісні та кількісні.

Метрики для якісних ознак.

Введемо ряд наступних позначень:

- Z^{00} число ознак, які приймають одночасно значення 0 для об'єктів x_i та x_j ;
- Z^{11} число ознак, які приймають одночасно значення 1 для об'єктів x_i та x_j ;
- Z^{01} число ознак, які приймають значення 0 для об'єкта x_i та 1 для об'єкта x_j ;
- Z^{10} число ознак, які приймають значення 1 для об'єкта x_i та 0 для об'єкта x_j ;
- Z загальне число ознак.

У таблиці 3.1 наведено метрики для обчислення відстаней ознак, що можуть приймати значення “так” або “ні” [26].

Таблиця 3.1 – Метрики для ознак, що приймають булеві значення

Автор	Формула
Жаккар	$1 - \frac{Z^{11}}{Z^{11} + Z^{01} + Z^{10}}$
Чекановськи	$1 - \frac{2 + Z^{11}}{2 + Z^{11} + Z^{01} + Z^{10}}$
Сокел і Міченер	$1 - \frac{Z^{11} + Z^{00}}{Z}$
Рассел і Рао	$1 - \frac{Z^{11}}{Z}$
Кульчинськи	$\frac{Z^{10} + Z^{01}}{Z^{11} + Z^{00}}$
Сокел і Сніт	$1 - \frac{Z^{11}}{Z^{11} + 2(Z^{01} + Z^{10})}$
Родджерс і Танітімо	$1 - \frac{Z^{11} + Z^{00}}{Z^{11} + Z^{00} + 2(Z^{01} + Z^{10})}$
Евклід	$\sqrt{\frac{Z^{01} + Z^{10}}{Z}}$

Найбільш пошиrenoю і простою у випадку якісних ознак є метрика Хемінга, зміст якої полягає в наступному:

$$d(x_i, x_j) = \frac{\text{число змінних в яких об'єкти } x_i \text{ та } x_j \text{ належать до різних категорій}}{z} \quad (3.2)$$

Відстань Хемінга розглядають як квадрат евклідової відстані в просторі бінарних змінних, які відповідають категоріям (класам) початкових даних, тобто:

$$d(x_i, x_j) = \frac{1}{Z} \sum_{k=1}^Z \sum_{l=1}^{p_k} (x_{il}^k - x_{jl}^k)^2 \quad (3.3)$$

де k – номер ознаки;

l – номер значення ознаки;

p_k – кількість варіантів значення k -ої ознаки.

Метрики для кількісних ознак.

В таблиці 3.2 наведено деякі метрики для обчислення відстані за кількісними ознаками.

Таблиця 3.2 - Метрики для ознак, що приймають чисельні значення

Назва	Формула
Евклідова відстань	$\sqrt{\sum_k (x_{ik} - x_{jk})^2}$
Квадрат евклідової відстані	$\sum_k (x_{ik} - x_{jk})^2$
Зважена евклідова відстань	$\sqrt{\sum_k w_k (x_{ik} - x_{jk})^2}$
Манхеттенська відстань	$\sum_k x_{ik} - x_{jk} $
Відстань Чебишева	$\max_k x_{ik} - x_{jk} $
Степенева відстань	$\left(\sum_k x_{ik} - x_{jk} ^p \right)^{\frac{1}{r}}$

3.3.2 Обчислення вагових коефіцієнтів ознак методом парних порівнянь

При обчисленні відстаней між об'єктами за зваженими метриками, є необхідність знаходження вагових коефіцієнтів ознак. Одним з алгоритмів, які дозволяють це зробити є метод парних порівнянь [27].

Метод парних порівнянь (точніше модифікація по Т. Сааті) полягає в порівнянні досліджуваних об'єктів (альтернатив, критеріїв, факторів) між собою. Об'єкти порівнюються попарно по відношенню до їх впливу («ваги», або «інтенсивності») на загальну для них характеристику.

Цей метод (метод парного порівняння альтернатив – МППА) є базовим для запропонованих Сааті методів аналітичних ієрархічних процесів і аналітичних мережевих процесів підтримки прийняття рішень. Сутність його полягає в наступному. Експертові послідовно пред'являються пари альтернатив (A_i, A_j) і пропонується визначити ступінь d_{ij} переваги альтернативи A_i над альтернативою A_j щодо деякого якісного критерію C . При цьому, якщо експерту була пред'явлена пара (A_i, A_j) і він визначив ступінь переваги, то пара (A_i, A_j) вже не пред'являється, а ступінь переваги d_{ij} визначається, виходячи з метризованих мультиплікативних відносин. Таким чином, при наявності k альтернатив експерт повинен виконати $\frac{k(k-1)}{2}$ порівнянь. Співвідношення $d_{ij} = \frac{1}{d_{ji}}$ є фундаментальним для методу Сааті для визначення відносних ваг альтернатив.

Ступені переваги надаються в фундаментальній шкалі зі значеннями. Числа показують, у скільки разів один об'єкт перевершує інший щодо загальної властивості або критерія .

Елементи d_{ij} , $i, j = 1, \dots, k$ утворюють сверхтранзитивну квадратну матрицю парних порівнянь D . При цьому елемент d_{ij} можна трактувати як відношення wag альтернатив A_i і A_{ij} , тобто $\frac{w_i}{w_j}$.

Запропоновано декілька методів обчислення відносних wag альтернатив, виходячи з матриці D .

Найбільш математично обґрунтованим є метод власного вектора, запропонований Сааті, що полягає в наступному. Позначимо через (3.4) вектор стовпець відносних wag альтернатив.

$$W = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_k \end{pmatrix} \quad (3.4)$$

Рівняння (3.5) називається характеристичним рівнянням матриці D . Йому відповідає система лінійних однорідних рівнянь (3.6).

$$|CW - \lambda W| = 0 \quad (3.5)$$

$$\begin{cases} (d_{11} - \lambda)w_1 + d_{12}w_2 + \dots + d_{1k}w_k = 0 \\ d_{21}w_1 + (d_{22} - \lambda)w_2 + \dots + d_{2k}w_k = 0 \\ \dots \dots \dots \dots \dots \dots \dots \dots \dots \\ d_{k1}w_1 + d_{k2}w_2 + \dots + (d_{kk} - \lambda)w_k = 0 \end{cases} \quad (3.6)$$

Необхідною і достатньою умовою існування нетривіального рішення цієї системи рівнянь, тобто рішення, для якого існує $i = (1, k)$, $w_i \neq 0$ є рівність нулю детермінанта цієї системи (тільки при виконанні цієї умови система має нескінченну множину рішень, серед яких обов'язково є нетривіальні). Записавши вираз детермінанта цієї системи і прирівнявши його до нуля, отримаємо рівняння k -го ступеня відносно λ , яке має k коренів (не обов'язково різних і відмінних від 0). Значення цих коренів називаються

характеристичними числами або власними значеннями матриці D . Кожному характеристичному числу відповідає своя система лінійних рівнянь, яку можна отримати, підставивши у (3.6) відповідне характеристичне число. Рішення отриманої системи щодо w_i , визначається з точністю до скалярного множника. Воно називається власним вектором матриці порівнянь.

У якості множини відносних ваг альтернатив пропонується використовувати компоненти власного вектора, що відповідає максимальному характеристичному числу λ_{max} . Таку рекомендацію можна обґрунтувати таким чином. Якщо матриця D задовольняє вимозі (3.7) то вона має лише одне характеристичне число $\lambda_{max} = k$, де k - порядок системи рівнянь (3.6), тобто кількість альтернатив.

$$\forall i, j \in h \left[d_{ij} = \left(\frac{w_i}{w_j} \right) = \left(\frac{w_i w_h}{w_j w_h} \right) = \left(\frac{w_i}{w_h} \right) \left(\frac{w_h}{w_j} \right) = d_{ih} d_{hj} \right], \quad (3.7)$$

Неважко переконатися в тому, що при виконанні згаданої умови кожне з нульових значень характеристичного числа дає тривіальний власний вектор, тобто вектор, що має всі нульові компоненти. Тому, лише максимальне для даного випадку значення характеристичного числа дає нетривіальний вектор відносних ваг альтернатив.

Виходячи з того, що для повністю узгодженої матриці D , тобто для матриці, що задовольняє умові (3.7), $\lambda_{max} = k$, а для неузгодженої матриці завжди $\lambda_{max} \geq k$, Сааті запропонував в якості показника ступеня узгодженості елементів матриці D використовувати величину індексу узгодженості (consistency index - CI) (3.8).

$$CI = \frac{\lambda_{max} - k}{k-1} \quad (3.8)$$

Для оцінки достатності ступеня узгодженості Saati пропонує використовувати ставлення узгодженості (consistency ratio - CR , яке дорівнює : $CR = \frac{CI}{CIS}$, де CIS - середнє значення CR , обчислені для великої кількості випадковим чином згенерованих матриць парних порівнянь в фундаментальній шкалі, які задовольняють умові $d_{ij} = \frac{1}{d_{ji}}$; значення CIS наведено в таблиці 3.3.

Таблиця 3.3 – Значення CIS

k	1	2	3	4	5	6	7	8	9	10
CIS	0	0	0,52	0,89	1,11	1,25	1,35	1,4	1,45	1,49

Результатуючий вектор відносних ваг вважається прийнятним, якщо CR дорівнює приблизно 0,1 (але не перевищує 0,2). Для $k = 3$, CR не повинен перевищувати 0,05, для $k = 4$, ця межа дорівнює 0,08.

Існує кілька спрощених методів обчислення відносних ваг альтернатив.

Перший з них, так званий метод ступеня, полягає в наступному. Матриця порівнянь зводиться в деяку ступінь q , після чого для обчислення w_i знаходиться сума елементів i -го рядка, яка ділиться на суму всіх елементів матриці. Обчислені значення w_i порівнюються зі значеннями, обчисленими при зведенні матриці в ступінь $q - 1$. Процес закінчується, якщо різниця не перевищує заданої величини помилки.

Суть методу середнього геометричного задається виразами (3.9) та (3.10)

$$v_i = \sqrt[k]{\prod_{j=1}^k d_{ij}} \quad (3.9)$$

$$w_i = \frac{v_i}{\sum_{h=1}^k v_h} \quad (3.10)$$

Цей метод і метод власного вектора дають майже однакові значення відносних ваг, якщо матриця парних порівнянь достатньо узгоджена.

Для отримання наступного методу обчислення відносних вагів альтернатив (3.11) звернемо увагу на те, що $d_{ij} = \frac{w_i}{w_j}$.

$$\sum_{i=1}^k d_{ij} = \sum_{i=1}^k \frac{w_i}{w_j} = \frac{\sum_{i=1}^k w_i}{w_j} \quad (3.11)$$

Оскільки $w_i, i = (1, k)$ є нормалізовані значення відносних ваг альтернатив, то маємо $\sum_{i=1}^k w_i = 1$. Тому $w_j = 1 / \sum_{i=1}^k d_{ij}$.

Даний метод, як і попередній, також дає хороші результати, якщо матриця парних порівнянь досить узгоджена.



3.3.3 Методи виділення кластерів Орг.С.

Коли кожний об'єкт представляє окремий кластер, та відстані між ними задані обраною метрикою виникає необхідність введення правила, за яким будуть об'єднуватись окремі кластери [28]. Таке правило називається правилом зв'язку. Існує багато правил зв'язку, огляд деяких з них проведено нижче.

Метод ближнього сусіда або одиночний зв'язок.

Тут відстань між двома кластерами визначається відстанню між двома найближчими об'єктами (сусідами) у різних кластерах. Цей метод дозволяє виділяти кластери якії завгодно складної форми за умови, що різні частини кластерів з'єднані ланцюжком близьких один до одного елементів. У результаті роботи методу кластери представляються довгими "ланцюжками" або "воловнистими" кластерами, з'єднані разом лише окремими елементами, які випадково виявилися ближче за інших один до одного.

Метод дальнього сусіда або повний зв'язок.

У цьому методі відстані між кластерами визначаються найбільшою відстанню між будь-якими об'єктами у різних кластерах (тобто “найбільш віддаленими сусідами”). Метод добре використовувати, коли об'єкти мають абсолютно різне походження. Якщо ж кластери мають подовжену форму або їх природнім типом є “ланцюжковий”, то цей метод використовувати не слід.

Метод Варда.

В якості відстані між кластерами береться приріст суми квадратів відстаней об'єктів до центрів кластерів, що отримуються в результаті їх об'єднання. На відміну від інших методів кластерного аналізу для оцінки відстаней між кластерами тут використовуються методи дисперсійного аналізу. На кожному кроці алгоритму об'єднуються такі два кластери, які забезпечують мінімальний приріст цільової функції, тобто сумі квадратів всередині кластеру. Цей метод націленний на об'єднання близько розміщених кластерів та створення маленьких кластерів.

Метод незваженого попарного середнього.

В якості відстані між двома кластерами обирається середня відстань між усімаарами об'єктів в них. Цей метод слід використовувати, коли об'єкти мають різне походження, у випадках наявності кластерів “ланцюжкового” типу або при припущення про нерівні розміри кластерів.

Метод зваженого попарного середнього. Цей метод схожий на метод незваженого попарного середнього, різниця полягає лише в тому, що тут в якості вагового коефіцієнта використовується розмір кластера (кількість об'єктів, що входить у кластер). Цей метод рекомендується використовувати при припущення про різні розміри кластерів.

Незважений центроїдний метод.

У якості відстані між двома кластерами в цьому методі обирається відстань між центрами їх ваги.

Зважений центроїдний метод.

Цей метод схожий на попередній, різниця полягає в тому, що тут для врахування різниці між розмірами кластерів використовуються вагові коефіцієнти. Цей метод рекомендується використовувати у випадках, якщо є припущення відносно істотних відмінностей у розмірах кластерів.

Висновки до розділу

У розділі проведено огляд математичних моделей, які можуть використовуватися для реалізації уніфікованої моделі кластеризації Орг.С.

Проаналізована існуюча класифікація методів кластерного аналізу, виконана формалізація постановки задачі кластеризації. Детально розглянуто ієрархічний кластерний аналіз, який найкраще підходить для кластеризації Орг.С.

Детально розглянуті метрики для обчислення відстаней між об'єктами, правила зв'язку, а також метод парних порівнянь для обчислення вагових коефіцієнтів ознак. Описані відмінності методів та метрик, а також рекомендації для їх використання при вирішенні різних задач.

4 ПРОГРАМНА РЕАЛІЗАЦІЯ УНІФІКОВАНОЇ МОДЕЛІ КЛАСТЕРИЗАЦІЇ ОРГ.С.

4.1 Застосування уніфікованої моделі для реалізації мобільного додатку “Your Travel Agency”

В рамках роботи була виконана програмна реалізація уніфікованої моделі Орг.С. у вигляді мобільного додатку “Your Travel Agency”. Призначення даного додатку – допомогти користувачам здійснити вибір туристичного агентства, а також надати інформацію про них.

Для розробки програмного забезпечення використовувалися наступні засоби та інструменти:

- мова програмування Objective C;
- IDE Xcode 6.3.2;
- Design and Management Tool Visual Paradigm 12.1
- IOS simulator.

Даний мобільний додаток призначений для використання на пристроях з операційною системою IOS версії не нижче 8.3., зокрема на мобільних телефонах iPhone.

Під час розробки у якості основної літератури використовувалися джерела [29-32]. Вихідний код програми наведено у Додатку А.

4.2 Вибір методів кластерного аналізу для кластеризації туристичних агентств

Вибір математичного апарату для розв'язання задачі підбору оптимального туристичного агентства здійснювався на основі вхідних та вихідних даних задачі а також її особливостей.

Вхідними параметрами задачі є:

- множина туристичних агенств;
- ознаки, які користувач обрав для пошуку.

Вихідним параметром є підмножина туристичних агенств, що найбільше задовільняють умовам запиту користувача.

Реалізація програмного засобу передбачає наявність лише тих ознак, що можуть приймати значення “так” або “ні”. Така особливість робить доцільним використання метрики Рассела і Рао, для обчислення міри подібності об'єктів:

$$d(x_i, x_j) = 1 - K \quad (4.1)$$

де x_i, x_j – об'єкти кластеризації;

K – коефіцієнт Рао.

Коефіцієнт Рао обчислюється за наступною формулою (4.2):

$$K = \frac{Z^{11}}{Z} \quad (4.2)$$

де Z – кількість ознак, за якими проводиться кластеризація;

Z^{11} – кількість співпадань значення “так”.

Для виділення кластерів використовувався метод “Найближчого сусіда”. Вибір такого методу дозволяє у майбутньому деталізувати результат користувачу, а саме додавати до множини найоптимальніших туристичних агенств ще й множину меньш оптимальних варіантів.

Для того, щоб виконати, зупинку процесу кластеризації, створюється так званий “ідеальний” об’єкт, тобто об’єкт, що повністю співпадає з критеріями, який задав користувач. Як тільки даний об’єкт об’єднується з іншими у кластер, процес зупиняється. На вихід подається кластер, у який потрапив “ідеальний” об’єкт.

4.3 Архітектура програмного забезпечення

4.3.1 Діаграма класів

Діаграма на рисунку 4.1 показує найбільш істотні зв'язки і відносини між класами, задіяними в алгоритмі вибору оптимальних туристичних агенств, на основі методів кластерного аналізу.

Специфікація діаграми класів:

- класи GLOBAL та AppDeledate контролюють правильний початок та закінчення роботи програми;
- класи MasterViewController та DetailViewController контролюють процес вибору ознак;
- класи YTAAgencyViewController та YTAInformationViewController відображають інформацію про туристичні агентства;
- клас YTAClusterAnalysis виконує кластеризацію об’єктів за ознаками, які обрав користувач;
- клас Agency та Attribute містять інформацію про туристичні агентства та ознаки відповідно;

- клас YTADatabase є базою даних, яка зберігає туристичні агентства та озаки.

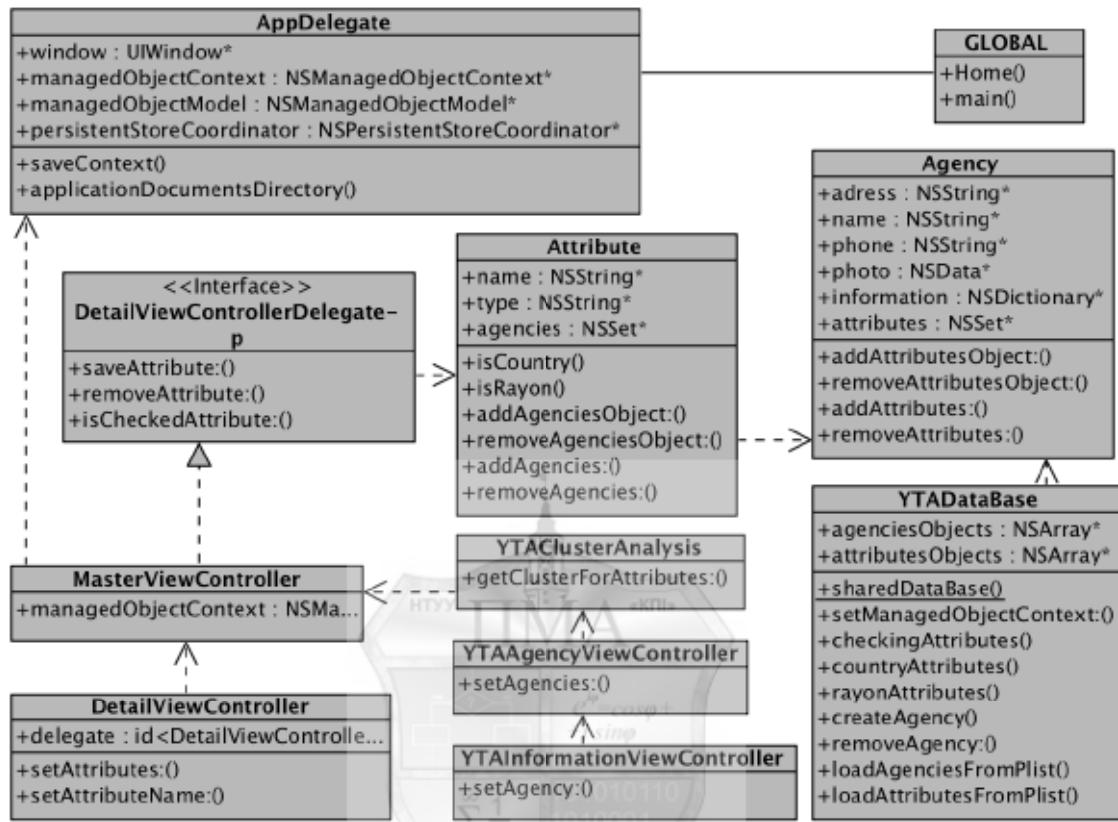


Рисунок 4.1 – Діаграма класів у нотації UML

4.3.2 Концепція “Модель – вид – контроллер”

Архітектура програми побудована з використанням концепції MVC. MVC – архітектурний шаблон, який використовується під час проектування та розробки програмного забезпечення. Цей шаблон поділяє систему на три частини: модель даних, вигляд даних та керування. Застосовується для відокремлення даних (модель) від інтерфейсу користувача (виду) так, щоб

зміни інтерфейсу користувача мінімально впливали на роботу з даними, а зміни в моделі даних могли здійснюватися без змін інтерфейсу користувача.

Компонента Вид шаблону MVC.

В архітектурі програми компонента Вид шаблону MVC представлена користувацьким інтерфейсом, що був розроблений на основі підходу розкадрування (storyboard).

Розкадрування — представлення сценарію у вигляді послідовності кадрів. Є робочою схемою майбутнього фільму, що пропонує різні схеми, навіть малюнки, креслення, із зображенням кадрів, місця, руху камери, а також вказівки щодо звуку (шуми, музика, голос за кадром, діалоги і т. п.).

Розкадрування — це продумування сюжету за допомогою ключових кадрів. А ключові кадри — це кадри, які відображають зміну плану або дії. Всі інші кадри між ними називаються проміжними.

На рисунку 4.2 представлено розкадрування інтерфейсу мобільного додатку “Your Travel Agency”.

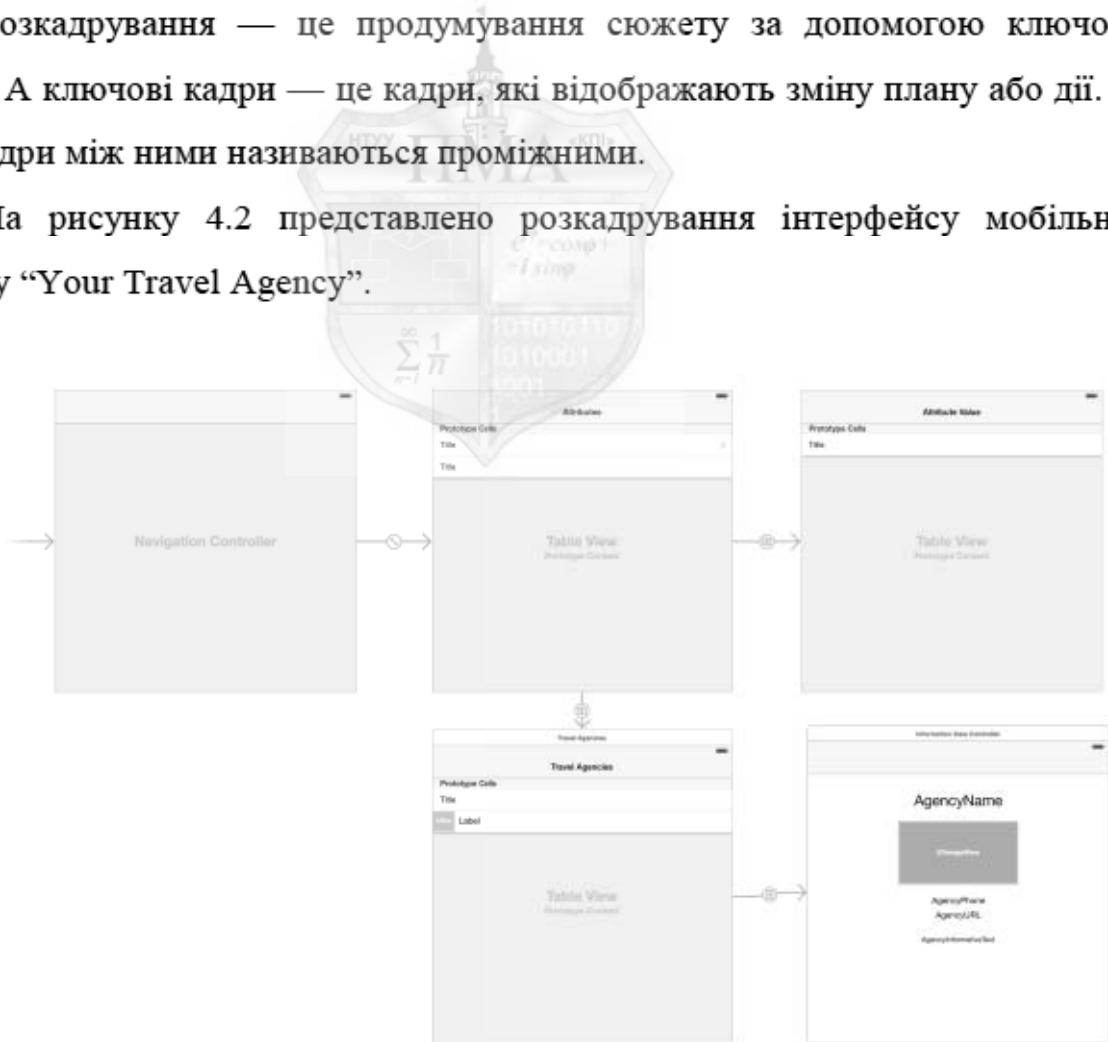


Рисунок 4.2 – Розкадрування інтерфейсу

Компонента Модель шаблону MVC.

Модель програмного забезпечення реалізована як база даних. Робота з базою даних була забезпечена використанням фреймворку Core Data, постачальником якого є компанія Apple.

База даних містить усього дві таблиці: “Туристичне агентство” та “Ознака”.

Таблиця “Туристичне агентство” містить наступні поля:

- назва;
- адреса;
- телефон;
- зображення логотипу;
- інформація (значення ознак).

Таблиця “Ознака” містить наступні поля:

- назва;
- тип.

Компонента Контроллер шаблону MVC.

В розробленому додатку можна виділити два типи контроллерів: контролери, що управляють інтерфейсом (YTAInformationViewController, YTAAgencyViewController, MasterViewController, DetailViewController) та клас який виконує кластерний аналіз (YTAClusterAnalysis).

4.4 Інтерфейс користувача

Інтерфейс мобільного додатку розроблено згідно [29]. Він складається з 4 екранів, які призначені для вибору ознак (рисунок 4.3), вибору значень ознак (рисунок 4.4), перегляду списку туристичних агентств (рисунок 4.5) та перегляду інформації про туристичне агентство (рисунок 4.6). Елемент

“Agencies” запускає процедуру кластерного аналізу, якщо користувач обрав параметри пошуку. Якщо ні – вібодражає усі туристичні агентства з бази даних.

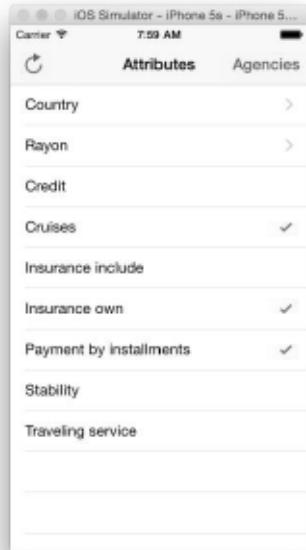


Рисунок 4.3 – Вибір ознак



Рисунок 4.4 – Вибір значень ознак

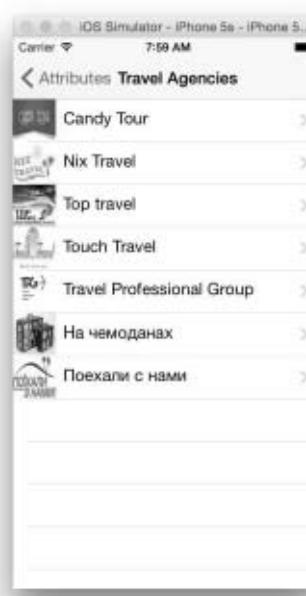


Рисунок 4.5 – Список агентств



Рисунок 4.6 – Інформація про агентство

4.5 Результати роботи програми

Перевірка роботи програми була проведена з використанням даних про 15 туристичних агентств (Додаток Б).

Результати роботи програми показали наступні результати:

- при обранні однієї ознаки додатком було відфільтровано 13 туристичних агентств;
- при обранні двох ознак додатком було відфільтровано 9 туристичних агентств;
- при обранні трьох ознак додатком було відфільтровано 3 туристичних агентства.

Подальше збільшення кількості обраних ознак збільшувало число туристичних агентств, які пропонував додаток, адже все важче було знайти об'єкт, що максимально відповідав би запиту користувача.

Висновки до розділу

У розділі проведена програмна реалізація моделі кластеризації Орг.С. у вигляді мобільного додатку “Your Travel Agency” для пошуку туристичних агентств.

У якості математичних методів кластерного аналізу обрано метод близнього сусіда і метрику Рассела та Рао.

Архітектура програмного забезпечення розроблена на основі діаграми класів з використанням концепції “Модель – вид – контроллер”.

Проведено тестування програми, яке довело дієздатність моделі а також доцільність її використання для пошуку туристичних агентств.

ВИСНОВКИ

Результатом магістерської дисертації є модель кластеризації Орг.С., яка призначена для автоматизації процесу кластеризації Орг.С.

Проведено огляд існуючих видів, класів Орг.С., методів та засобів їх виділення.

Встановлено:

- відсутність уніфікованої моделі кластеризації Орг.С.;
- відсутність спеціалізованого програмного забезпечення для кластеризації Орг.С.;
- доцільність використання кластерного аналізу для виділення видів та класів Орг.С.;

Розроблено:

- концептуальну та уніфіковану модель кластеризації Орг.С.

Реалізовано:

- програмне забезпечення для пошуку туристичних агентств за обраними критеріями, з використанням уніфікованої моделі кластеризації Орг.С.

Розроблена уніфікована модель кластеризації дозволяє формувати види та класи Орг.С. використовуючи для цього більше, ніж одну ознаку. Така модель є основою для реалізації засобів автоматизації кластеризації Орг.С. Її особливістю є те, що компоненти даної моделі можуть бути реалізовані тими методами кластерного аналізу, які найкраще підходять для вирішення конкретних задач.

Результати роботи опубліковані в матеріалах:

- сьомої наукової конференції магістрантів та аспірантів «Прикладна математика та комп'ютинг»;
- 17-ї Міжнародної науково-технічної конференції SAIT 2015.

Перспективи:

- розробка інших програмних засобів для вирішення задач, які потребують виділення класів та видів Орг.С. з використанням розробленої моделі;
- порівняння роботи різних метрик та правил зв'язку для розробленого мобільного додатку “Your Travel Agency”;
- розробка версії додатку “Your Travel Agency” для використання на ПК.



СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Національний класифікатор України. Класифікатор видів економічної діяльності. [Текст] / ДК 009:2010. – [Чинний від 2012–01–01]. — К.: Держспоживстандарт України. – 2010. – 46 с.
2. Маслянко П.П. Бізнес інжинірінг організаційних систем [Текст] / П.П. Маслянко, О.С. Майстренко // Наукові вісті НТУУ “КПІ”. – 2011. – №1. – 69–78 с.
3. Fowler M. UML Distilled [Текст] / М. Фаулер // СПб: Символ-Плюс. – 2005. – 184 с.
4. International Standard Industrial Classification of All Economic Activities [Електронний ресурс] // Режим доступу: <http://unstats.un.org/unsd/cr/registry/isic-4.asp>
5. NACE Rev 2. Statistical classification of economic activities in the European Community [Електронний ресурс] // Режим доступу: <http://www.nace.org/home.aspx>
6. Олянич Д. Б. Теория организации: учебник [Текст] / Д. Б. Олянич // Ростов н/Д: Феникс. – 2008. — 408 с.
7. Коваленко, А. В. Кластерный анализ финансово-экономического состояния предприятий строительной отрасли [Електронний ресурс] // Режим доступу: <http://ej.kubagro.ru/2010/06/pdf/37.pdf>.
8. Державний класифікатор України. Класифікація організаційно-правових форм господарювання [Текст] / ДК 002:2004. – [Чинний від 2004–06–01]. — К.: Держспоживстандарт України. – 2004. – 46 с.
9. Воронин, А.В. Использование кластерного анализа для выбора локальных стратегий [Текст] / А.В. Воронин // Проблемы и перспективы управления экономикой и маркетингом в организации. – №1. – 2001. – 23с.

10. Дроздов Н.Д. Введение в теорию организационных систем: Учеб. пособие [Текст] / Н.Д. Дроздов // Твер. гос. ун-т. Тверь. – 2001. – 172 с.
11. Волкова Н.А. Кластерный анализ результатов социологического опроса работников предприятия [Електронний ресурс] // Режим доступу: <http://ieee.tusur.ru/nauka/cluster>
12. Гаркавенко С.С. Маркетинг: підручник для вузів [Текст] / С.С. Гаркавенко // К.: Лібра. – 2002. – 712 с.
13. Соколова Л.В. Використання методів кластерного аналізу у практичній діяльності підприємств [Текст] / Л.В. Соколова, Г.М. Верясова // Харків. – 2010. – 34с.
14. Рета М.В. Управління витратами на створення нової продукції : автореф. ... канд. екон. наук за спец. 08.00.04 НТУ “ХПІ”. [Текст] / М. В. Рета. // Харків. – 2010. – 19с.
15. Попова Л.М. Методичний підхід до оцінки розподілу пріоритетів сфер діяльності з метою ефективного проведення диверсифікації діяльності підприємства [Електронний ресурс] // – Режим доступу: http://www.nbuv.gov.ua/portal/Soc_Gum/Aprer/2009_5_1
16. Wagstaff, K. Constrained K-means Clustering with Background Knowledge. Proceedings of the Eighteen International Conference on Machine Learning [Текст] / Wagstaff K., Cardie C. // 2001. – Р. 577–584.
17. Котелевская, Н.В. Экономическая диагностика предприятия: просто о сложном [Текст] / Н.В. Котелевская. // Харьков: Консульт. – 2007. – Том 10 (160). – 160 с.
18. Управління діяльністю підприємства у конкурентному середовищі: монографія [Текст] / Л.В. Соколова, Г.М. Верясова, О.В. Манакова та ін.; за заг. ред. Л.В. Соколової. // Х.: ТОВ “Компанія СМІТ”. – 2010. – 190 с.
19. Геворкян, А. Ю. Адаптація підприємств машинобудування до умов СОТ: автореф. ... канд. екон. наук за спец. 08.00.04 НТУ “ХПІ” [Текст] / А.Ю. Геворкян // Харків. – 2010.

20. Оріщенко А. О. Концептуальна модель кластеризації організаційних систем [Текст] / П.П. Маслянко, А.О. Оріщенко // Прикладна математика та комп'ютинг. ПМК, 2015 : сьома наук. Конференція магістрантів та аспірантів присвячена 25-річчю факультету прикладної математики, Київ, 15-17 квіт. 2015 р.: - К.: Просвіта, 2015. – 308 с.
21. Маслянко П. П. Система сущностей бизнес моделей организационных систем [Текст] / П. П. Маслянко, А. С. Майстренко // Кибернетика и Системный Анализ. – 2012. 118–128 с.
22. Маслянко П.П., Майстренко А.С. Интеграция разнородных моделей организационных систем. [Текст] / П.П. Маслянко, А.С. Майстренко // Проблемы управления и информатики. – 2012. – №21. – 143–156 с.
23. Оріщенко А. Компонентна модель кластерного аналізу організаційних систем [Текст] / А. Оріщенко, П. Маслянко // Системний аналіз та інформаційні технології: матеріали 17-ї Міжнародної науково-технічної конференції SAIT 2015, Київ, 22-25 червня 2015 р. / ННК “ПСА” НТУУ “КПІ”. – К.: ННК “ПСА” НТУУ “КПІ”. – 2015. – 304 с.
24. Купалова Г.І. Теорія економічного аналізу [Текст] / Купалова Г.І. // Маріуполь. – 2002. – 712 с.
25. Бериков В. С., Лбов Г. С. Современные тенденции в кластерном анализе [Текст] / В.С. Бериков, Г.С. Лбов // Всероссийский конкурсный отбор обзорно-аналитических статей по приоритетному направлению «Информационно-телекоммуникационные системы». – 2008. — 26 с.
26. Жамбю М. Иерархический кластер-анализ и соответствия [Текст] / М. Жамбю // Финансы и статистика, Москва. – 1988. – 230 с.
27. Саати Т. Принятий решений. Метод анализа иерархий [Текст] / Т. Саати // Москва – Радио и связь. – 1993. – 349 с.
28. Трухачёва Н.В. Математическая статистика в медико-биологических исследованиях с применением пакета Statistica [Текст] / Н.В. Трухачёва // Издательство: ГЭОТАР-Медиа. — 267 с.

29. IOS Human Interface Guideline [Електронний ресурс] // Режим доступу: <https://developer.apple.com/library/ios/documentation/iPhone/Conceptual/iPhoneOSProgrammingGuide/iPhoneAppProgrammingGuide>
30. CoreData guide [Електронний ресурс] // Режим доступу: <https://developer.apple.com/library/ios/documentation/iPhone/Conceptual/iPhoneOSProgrammingGuide/iPhoneAppProgrammingGuide>
31. Нахавандипур В. IOS. Приемы программирования [Текст] / В.Нахавандипур // СПБ.: Питер. – 2014. – 832 с.
32. Фримен Э. Паттерны проектирования [Текст] / Э. Фримен, Э Фримен // СПБ.: Питер. – 2014. – 656 с.

