

ДОДАТОК А


Перелік джерел посилання за науковими напрямками керівника та науковців
кафедри програмної інженерії

4. Mazurova, O. Research of ACID transaction implementation methods for distributed databases using replication technology / Mazurova, O., Naboka, A., Shirokopetleva, M. Innovative technologies and scientific solutions for industries, (2 (16), pp. 19-31. Doi: 10.30837/ITSSI.2021.16.019.

5. Analyzing and Comparison of NoSQL DBMS Kuzochkina, A., Shirokopetleva, M., Dudar, Z. 2018 International Scientific-Practical Conference on Problems of Infocommunications Science and Technology, PIC S and T 2018 - Proceedings, 2019, стр. 560–564, 8632133 DOI 10.1109/INFOCOMMST.2018.8632133.

ДОДАТОК Б

Звіт результатів перевірки на унікальність тексту в базі ХНУРЕ



Ім'я користувача: Кардаш Євген Вікторович каф.ПІ	ID перевірки: 1016331334
Дата перевірки: 07.06.2024 10:51:20 EEST	Тип перевірки: Doc vs Internet + Library
Дата звіту: 07.06.2024 11:11:13 EEST	ID користувача: 100013622

Назва документа: **2024_М_ПІ_ІПЗм-22-1_Недаєв_І_С_скорочений**

Кількість сторінок: **40** Кількість слів: **7898** Кількість символів: **57575** Розмір файлу: **875.14 KB** ID файлу: **1016130942**

Виявлено модифікації тексту (можуть впливати на відсоток схожості)

2.44%

Схожість

Найбільша схожість: **0.58%** з джерелом з Бібліотеки (ID файлу: **1016113569**)

1.8% Джерела з Інтернету	38	Сторінка 42
1.91% Джерела з Бібліотеки	71	Сторінка 42

0% Цитат

Вилучення цитат вимкнене

Вилучення списку бібліографічних посилань вимкнене

0% Вилучень

Немає вилучених джерел

Модифікації

Виявлено модифікації тексту. Детальна інформація доступна в онлайн-звіті.

Замінені символи	16
Підозріле форматування	7 сторінок

ДОДАТОК В

Слайди презентації



Дослідження залежності продуктивності реляційних та графових баз даних від даних



Недаєв Ігор Сергійович, ІПЗм-22-1
Науковий керівник: к.т.н, доц. Мазурова Оксана
Олексіївна

18 червня 2024

Актуальність дослідження

- У сучасних програмних продуктах виникає потреба не лише у зберіганні текстової або числової інформації, але й у зберіганні зображень, відео та аудіо файлів.
- У зв'язку із розвитком технологій та збільшенням обсягу даних, з якими працюють програмні продукти, все частіше виникає необхідність зберігання їх у великому обсязі.
- Це зумовлює використання різних типів БД, кожен із яких має своє призначення, переваги та недоліки.



Постановка задачі

- провести аналіз та вибір реляційних та графових для дослідження;
- провести аналіз та моделювання обраної предметної області (побудувати інфологічну модель, реляційну та графову логічні моделі);
- розробити план експериментального дослідження (визначити умови проведення експериментів, метрики для дослідження, розробити запити для досліджуваних даних);
- спроектувати та розробити програмне забезпечення для проведення дослідження;
- провести експериментальне дослідження продуктивності роботи зазначених СКБД, за його результатами сформулювати рекомендації щодо використання досліджуваних СКБД із різними даними (тип даних, їх обсяг, їх зв'язність тощо).

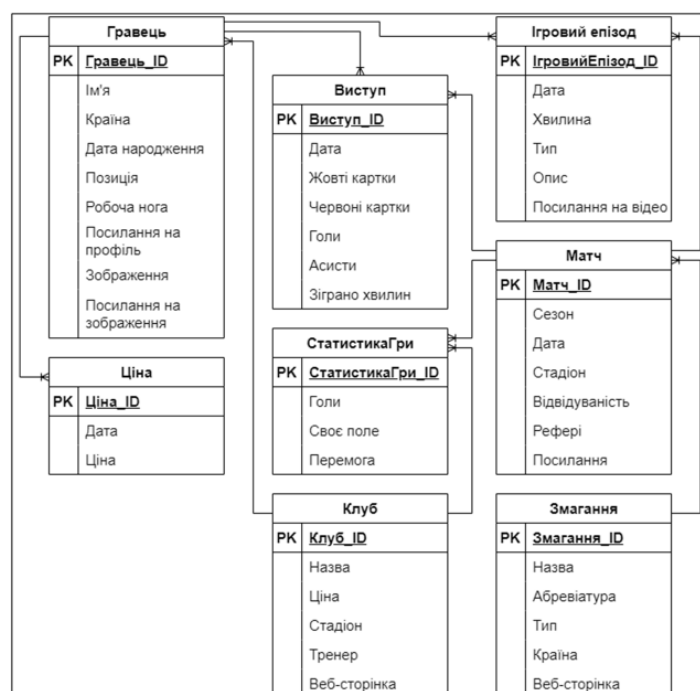


Вибір графової СКБД для дослідження

Метрики /СКБД	Час відгуку (мс)	Масштабованість	Підтримка граф. моделі	Відповідність ACID	Простота використання та розробки
OrientDB	45	Добре	Багато-модельна	Повна відповідність	Добре
Amazon Neptune	20	Відмінно	Графо-орієнтована	Повна відповідність	Добре
Azure Cosmos DB	20	Відмінно	Багато-модельна	Повна відповідність	Задовільно
Neo4j	15	Добре	Графо-орієнтована	Повна відповідність	Відмінно
ArangoDB	40	Добре	Багато-модельна	Повна відповідність	Добре



Аналіз та моделювання предметної області



ER-діаграма БД



Аналіз та моделювання предметної області

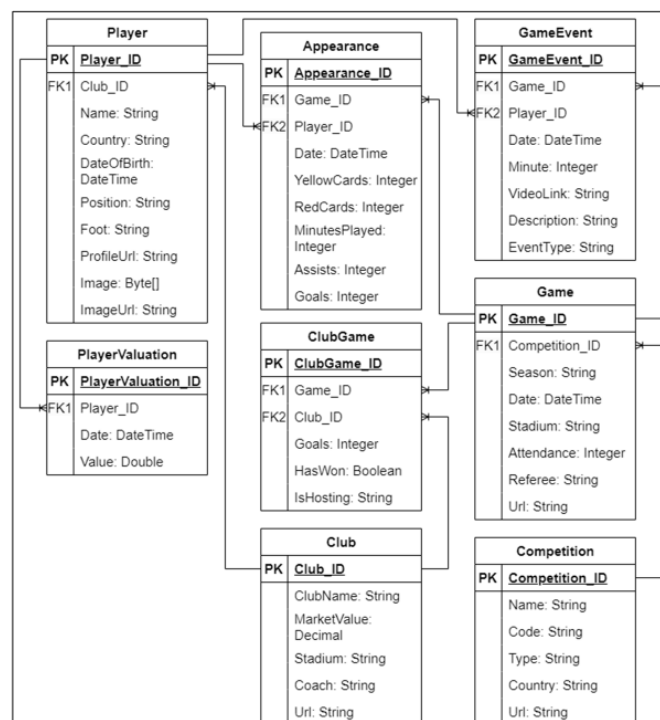
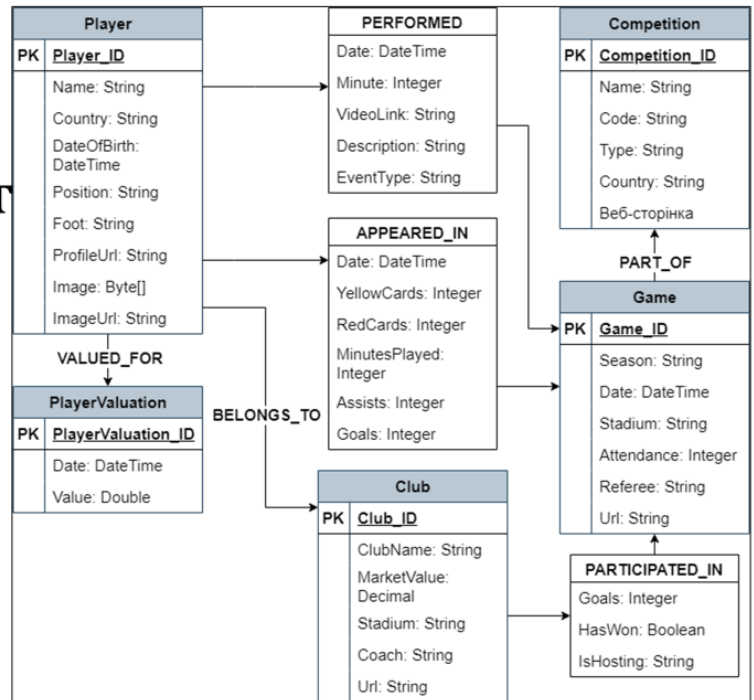


Схема реляційної БД



Аналіз та моделювання предметної області

Логічна модель графової БД



Планування експериментального дослідження

Обрано наступні метрики:

- швидкість виконання запиту (мс);
- відсоток використання процесора (%);
- витрати оперативної пам'яті (мб).



Response body

```
All done!
Memory used by SQL Server during query execution: 194,15234375 MB
CPU SQL Server during query execution: 19,51640067680618%
Execution time: 3310,4471 ms
```

Планування експериментального дослідження

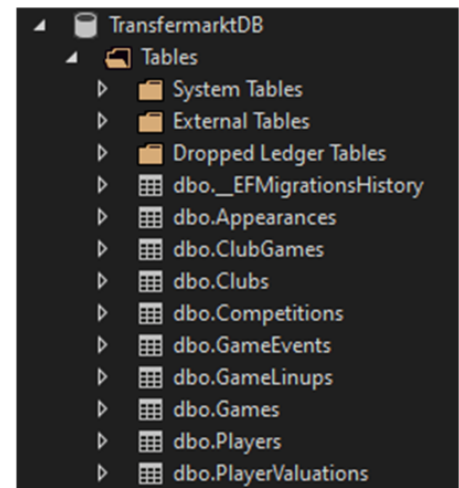
Для кожного із 4-ох типів запитів (INSERT/CREATE, UPDATE, DELETE, SELECT) було розроблено 4 варіації, де кожен запит використовує сутності, що:

- мають по 4 активних зв'язки та використовують зображення;
- мають по 4 активних зв'язки та **не** використовують зображення;
- **не** мають активних зв'язків та використовують зображення;
- **не** мають активних зв'язків та **не** використовують зображення.



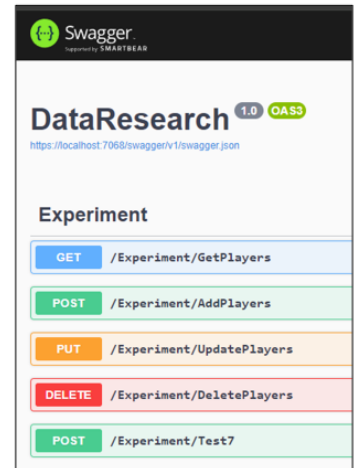
Архітектура системи для проведення експериментального дослідження

- Система для проведення експериментального дослідження складається із 2 проектів-реалізацій СКБД та бібліотеки для проведення вимірювань досліджуваних метрик.
- Усі застосунки розроблені на базі платформи .NET.



Опис програмного забезпечення, що було використано у дослідженні

- Проект для проведення замірів метрик – консольна бібліотека, створена на платформі .NET.
- Проекти-реалізації реляційної та графової СКБД – ASP.NET API, що використовують Swagger у якості свого інтерфейсу.
- Проекти написані на C#8.0.



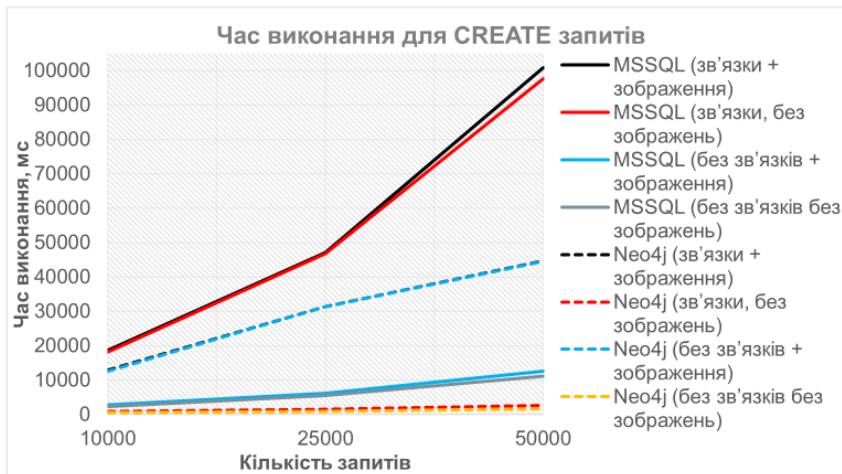
Результати експерименту

Перший експеримент:
INSERT для MSSQL та CREATE для Neo4j.

СКБД та запит /Метрики	Швидкість виконання запиту (мс)			Відсоток використання процесора (%)			Витрати оперативної пам'яті (MB)		
	10000	25000	50000	10000	25000	50000	10000	25000	50000
MSSQL (зв'язки + зображення)	18757	47170	100844	18,15	37,41	74,81	261,2	551,5	1081,3
MSSQL (зв'язки, без зображень)	18232	46835	97738	16,53	35,3	74,96	237,6	536,9	1065,5
MSSQL (без зв'язків + зображення)	2808	6259	12670	13,69	22,12	27,56	130,2	442	798,8
MSSQL (без зв'язків, без зображень)	2403	5611	11171	11,94	22,32	32,39	129,5	444,1	893,8
Neo4j (зв'язки + зображення)	12913	31433	44778	6,63	6,72	6,84	68,8	209,1	606,6
Neo4j (зв'язки, без зображень)	857	1491	2630	5,98	6,65	6,77	31,3	57	153
Neo4j (без зв'язків + зображення)	12672	31384	44484	4	3,28	4,8	69,8	205,6	610
Neo4j (без зв'язків, без зображень)	566	946	1686	4,72	3,16	4,15	29,6	71,2	145,9



Результати експерименту



- На продуктивність MSSQL найбільше впливає зв'язність даних.
- На продуктивність Neo4j впливає наявність зображень.



Результати експерименту

Другий експеримент: UPDATE.

- Neo4j виявилась продуктивнішою за MSSQL.
- MSSQL для всіх 4 варіацій запитів показала приблизно однаковий результат.

СКБД та запит /Метрики	Швидкість виконання запиту (мс)			Відсоток використання процесора (%)			Витрати оперативної пам'яті (МБ)		
	10000	25000	50000	10000	25000	50000	10000	25000	50000
MSSQL (зв'язки + зображення)	4836	8639	21679	19,96	22,29	37,60	282,51	483,88	843,61
MSSQL (зв'язки, без зображень)	3359	8213	18973	12,22	18,93	29,69	251,95	452,10	838,85
MSSQL (без зв'язків + зображення)	3444	8322	16253	13,51	24,48	33,32	192,79	399,84	774,86
MSSQL (без зв'язків, без зображень)	3346	7789	15453	14,7	23,77	27,48	158,53	370,7	741,96
Neo4j (зв'язки + зображення)	14638	38311	51910	6,61	6,8	7,1	140,16	392,94	904,61
Neo4j (зв'язки, без зображень)	374	773	1730	6,16	6,37	6,67	29,6	72,22	146,96
Neo4j (без зв'язків + зображення)	16297	37265	49492	6,42	6,54	7,29	145,63	492,66	894,16
Neo4j (без зв'язків, без зображень)	368	749	1745	6,11	6,22	6,69	25,38	68,98	145,85



Результати експерименту

Третій експеримент: DELETE.

- Крайній відсоток використання процесора знову показала Neo4j.
- Витрати оперативної пам'яті також менші у Neo4j.



СКБД та запит /Метрики	Швидкість виконання запити (мс)			Відсоток використання процесора (%)			Витрати оперативної пам'яті (МБ)		
	10000	25000	50000	10000	25000	50000	10000	25000	50000
MSSQL (зв'язки + зображення)	1549	2369	4138	8,84	10,8	8,89	23,59	23,66	23,72
MSSQL (зв'язки, без зображень)	1234	1582	3082	8,7	11,83	11,83	23,47	23,61	23,71
MSSQL (без зв'язків + зображення)	1346	2387	3643	11,76	11,75	11,95	23,44	23,62	23,72
MSSQL (без зв'язків, без зображень)	1262	1605	1998	11,44	11,56	11,71	23,42	23,62	23,68
Neo4j (зв'язки + зображення)	1077	3221	9157	6,21	6,68	7,75	2,49	2,61	2,94
Neo4j (зв'язки, без зображень)	781	2297	7558	5,15	5,35	5,68	2,44	2,5	2,64
Neo4j (без зв'язків + зображення)	793	1741	3467	6,01	6,53	7,32	2,3	2,37	2,6
Neo4j (без зв'язків, без зображень)	400	692	1218	4,38	5,07	5,35	2,28	2,39	2,43

Результати експерименту

Четвертий експеримент: SELECT.

- На продуктивність MSSQL найбільше впливає зв'язність даних.
- На продуктивність Neo4j впливає наявність зображень.



СКБД та запит /Метрики	Швидкість виконання запити (мс)			Відсоток використання процесора (%)			Витрати оперативної пам'яті (МБ)		
	10000	25000	50000	10000	25000	50000	10000	25000	50000
MSSQL (зв'язки + зображення)	3273,6	4874,9	8761,7	20,96	22,23	23,31	173,55	347,64	690,63
MSSQL (зв'язки, без зображень)	3228,7	4938,8	7925	18,85	19,75	22,73	81,33	169,93	231,61
MSSQL (без зв'язків + зображення)	1722,4	2051,7	2561,5	15,25	19,15	21,51	125,48	334,99	613,27
MSSQL (без зв'язків, без зображень)	1586	1844	2047,9	14,84	20,96	21,56	49,12	71,46	82,42
Neo4j (зв'язки + зображення)	651,4	975,4	1889,4	17,86	20,41	28,60	174,84	375,47	697,73
Neo4j (зв'язки, без зображень)	513,7	674,7	1073,3	7,28	9,97	13,64	52,87	144,52	309,16
Neo4j (без зв'язків + зображення)	574,6	790,1	1638,3	3,07	8,16	18,34	159,95	342,18	495,24
Neo4j (без зв'язків, без зображень)	444,1	483,8	923,2	3,96	5,50	15,12	35,36	65,55	182,23

Аналіз отриманих результатів

- У всіх досліджах Neo4j показала найкращий показник процесорного відсотка та майже у всіх – найкращий показник використання оперативної пам'яті.
- MSSQL виявилась продуктивнішою за показником «час виконання запиту» майже у всіх досліджах із використанням зображень, крім сценарію.



Висновки

- проведено аналіз та вибір реляційних та графових БД для дослідження;
- проведено аналіз та моделювання обраної предметної галузі футбольної статистики;
- розроблено план експериментального дослідження;
- спроектовано та розроблено програмне забезпечення для проведення дослідження;
- проведено експериментальне дослідження продуктивності роботи зазначених СКБД, за його результатами сформовано рекомендації щодо використання досліджуваних СКБД із різними даними (тип даних, їх обсяг, їх зв'язність тощо).
- За результатами дослідження опубліковано тези «Investigation of the Efficiency Dependence of Relational and Graph Databases on Data» на 12-тій Міжнародній науковій та технічній конференції «Інформаційні системи та технології» (IST-2023)



ДОДАТОК Г
Апробація результатів роботи



Investigation of the Efficiency Dependence of Relational and Graph Databases on Data

Ihor Nedaiev^a, Oksana Mazurova^a and Marija Shirokopetleva^a

^a *Kharkiv National University of Radio Electronics, 14 Nauky Avenue, Kharkiv, 61166, Ukraine*

Abstract

The purpose of the study is to analyze the methods of working with relational and graph databases, to develop recommendations for their use depending on the input data.

Keywords ¹

Data, dependence, efficiency, graph database, metric, relational database

1. Introduction

Nowadays, databases are an integral part of most application software systems under development. During development, different types of databases are used, where the logic of each of them has its own purpose, strengths and weaknesses. Accordingly, each type of DBMS (database management system) has its own specialized field of use and was created to solve its own range of tasks. In order to model complex, loosely structured subject areas with different types of data, relationships, etc., developers face an important task of determining which type of DBMS is best suited in this context.

2. Problem statement

The purpose of this paper is to formulate more effective recommendations for choosing a data model and a suitable DBMS taking into account the context of the data that will form the basis of the respective databases based on the analysis of the results of their experimental studies.

To achieve this goal, the following tasks need to be solved:

- to select the best database systems for further research based on the analysis of existing DBMSs and multi-criteria decision-making;
- plan an experimental study (select a subject area and design a database, develop a query system for the study and select metrics for the study);
- develop software solutions for each of the studied DBMSs;
- conduct experiments and formulate recommendations for choosing a DBMS in accordance with the context of the subject area data.

3. Description of the study

It was decided to focus on the study of the efficiency of relational and graph [1,2] databases and their dependence on the data context. The object of the study is the process of working with graph and relational databases, and the subject of the study is the methods of storing, processing and accessing data in relational and graph database management systems, in particular, lazy loading using the Entity Framework.

There are some general recommendations in the literature on choosing a database, but there is no clear division - the choice is up to the developers. Many studies show that graph databases are much

Proceedings of the *Information Systems and Technology (IST-2023)*, November 28-December 01, 2023, Kharkiv, Ukraine

EMAIL: ihor.nedaiev@nure.ua (A. 1); oksana.mazurova@nure.ua (A. 2); marija.shirokopetleva@nure.ua (A. 3)

ORCID: 0009-0000-6194-8143 (A. 1); 0000-0003-3715-3476 (A. 2); 0000-0002-7472-6045 (A. 3)



Information Systems and Technology (IST-2023) Part 2. Youth section

more efficient than relational databases if the database contains many relationships between tables [2], while with a large amount of data and a small number of relationships, their performance is much lower [1, 3].

We analyzed such popular graph DBMSs as: Neo4j, ArangoDB, OrientDB-X, Amazon Neptune, and Azure Cosmos DB. In order to select a DBMS for further research, the task of multicriteria selection was solved, taking into account such criteria as response time, scalability, data consistency and support for ACID properties during transaction processing, support for a graphical data model, ease of use and development.

The qualitative assessments obtained during the analysis were converted to quantitative scales, normalized, and analyzed using the Pareto principle. A convolutional model based on linear additive convolution with weighting coefficients was chosen for the multicriteria selection:

Based on the results of solving the multicriteria decision-making problem, the Neo4j graph DBMS was chosen for further research, since Amazon Neptune is focused on larger projects and has limited capabilities when used for free.

As a subject area for modeling and studying the dependence on the data context, the field of football statistics was chosen, which is described by data of various types, including media data, and requires modeling various types of relationships, for example, the history of encounters between two players, etc.

Researching the process of working with graph and relational databases involves studying the ways of storing, processing, and accessing data in these systems, as well as their differences in terms of efficiency.

As methods of storing, processing and accessing data in relational and graph database management systems, it was decided to consider the option of lazy loading using the Entity Framework as well. Lazy loading is a technique used to postpone the loading of related data until the moment when it is really needed.

4. Conclusion

At this stage, the most suitable database systems were selected for further research and an experimental study was planned. In the future, the obtained results will be used to write a master's thesis, in particular, to develop a software solution and develop recommendations for the choice of a DBMS depending on the subject area based on the experiments.

5. References

- [1] Kotiranta P., Junkkari M., Nummenmaa J. Performance of graph and relational databases in complex queries. *Applied sciences*. 2022. Т. 12, № 13. С. 6490. URL: <https://doi.org/10.3390/app12136490> (дата звернення: 03.04.2023).
- [2] Lazarska M., Siedlecka-Lamch O. Comparative study of relational and graph databases. 2019 IEEE 15th international scientific conference on informatics, м. Poprad, Slovakia, 20–22 листоп. 2019 р. 2019. URL: <https://doi.org/10.1109/informatics47936.2019.9119303> (дата звернення: 03.04.2023).
- [3] Mazurova, O. NOSQL database logic design methods for MONGODB and NEO4J / Mazurova, O., Syvolovskyi, I., Syvolovska, O. *Innovative technologies and scientific solutions for industries*, № 2 (20), pp. 52-63.

ДОДАТОК Д

Експертний висновок результатів перевірки кваліфікаційної роботи на
відповідність оформлення вимогам ДСТУ 3008: 2015

Експертний висновок результатів перевірки кваліфікаційної роботи

студент
(посада)

програмної інженерії
(кафедра)

ІІЗМ-22-1
(група)

Недаєв І.С.

(прізвище, ім'я, по батькові)

Зауваження

Пункт ДСТУ 3008-2015	Зміст пункту	Сторінка кваліфікаційної роботи
1	2	3
	7.1 Загальні положення	
	7.3 Нумерація сторінок звіту	
	7.4 Нумерація розділів, підрозділів, пунктів, підпунктів	
	7.5 Рисунки	
	7.6 Таблиці	
	7.7 Переліки	
	7.8 Примітки	
	7.9 Виноски	
	7.10 Формули та рівняння	
	7.11 Посилання	
	7.13 Список авторів	
	7.14 Скорочення та умовні позначки	
	7.15 Додатки	

Експерт

(підпис)

Вадим НЕЧВОЛОД

(прізвище, ініціали)

13.06.2024