

АНОТАЦІЯ

Волох Б.Ю. Інтелектуальне кешування даних в системі підтримки процесу відновлення пошкоджених об'єктів нерухомості. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 126 «Інформаційні системи та технології». – Київський національний університет будівництва і архітектури. – Київ, 2025.

Дисертацію присвячено вирішенню науково-практичної задачі підвищення ефективності обробки і зберігання даних в інформаційній інфраструктурі галузі архітектури і будівництва.

Наукова новизна одержаних результатів полягає в розробці методу інтелектуального управління кешем в інформаційній системі підтримки процесу відновлення пошкоджених об'єктів нерухомості, що вирішує ключові задачі кешування, базується на моделях машинного навчання і здатен забезпечити ефективний доступ до актуально важливих даних з урахуванням суттєвих ознак користувачів і пошкоджених об'єктів нерухомості.

вперше розроблено:

– гібридний метод інтелектуального управління кешем, що вирішує ключові задачі кешування, базується на моделях машинного навчання та враховує суттєві ознаки користувачів і пошкоджених об'єктів нерухомості;

удосконалено:

– модель інформаційної системи підтримки процесу відновлення пошкоджених об'єктів нерухомості шляхом інтеграції модуля інтелектуального кешування, що підвищує продуктивність системи і забезпечує швидкий доступ до актуально важливих даних;

– процес інформаційної підтримки користувачів системи підтримки процесу відновлення пошкоджених об'єктів нерухомості шляхом розширення колекцій документів для нереляційної бази даних, що призначена для зберігання даних про

пошкоджені об'єкти нерухомості, історії запитів користувачів і статистику пошкоджень;

набули подальшого розвитку:

- підходи до обробки і зберігання даних в інформаційній інфраструктурі галузі архітектури і будівництва через поєднання технологій інтелектуального кешування з нереляційними базами даних.

Основний зміст дисертаційної роботи

Запропоновано гібридний метод інтелектуального управління кешем, орієнтований на інформаційні високонавантажені системи з динамічним характером запитів, зокрема на систему підтримки процесу відновлення пошкоджених об'єктів нерухомості. Метод охоплює три ключові задачі: оцінка доцільності додавання об'єкта до кешу, вибір об'єкта для видалення у разі заповнення кешу та періодичну інвалідацію неактуальних об'єктів у кеші. Для вирішення цих задач використано поєднання моделей машинного навчання: табличного Q-Learning з механізмом відкладеного асинхронного оновлення значень моделі та логістичної регресії з механізмом відкладеного асинхронного накопичення буфера навчальних прикладів для подальшого навчання моделі та періодичного оновлення її параметрів. Проведено експериментальну оцінку ефективності методу, яка підтвердила перевагу інтелектуального підходу порівняно з існуючими методами кешування LRU та LFU, зокрема в умовах обмеженого розміру вхідних даних. Запропонований метод має високий потенціал для використання в інших інформаційних системах завдяки можливості адаптації вектора ознак моделей машинного навчання відповідно до характеристик об'єктів і контексту їх використання.

За результатами дослідження розроблено:

- гібридний метод інтелектуального управління кешем, що вирішує ключові задачі кешування, базується на моделях машинного навчання та враховує суттєві ознаки користувачів і пошкоджених об'єктів нерухомості;
- систему експериментального тестування ефективності гібридного методу інтелектуального управління кешем, що включає модулі для симуляції запитів,

реалізації логіки роботи методу, а також взаємодії з моделями машинного навчання, технологією кешування та нереляційною базою даних.

У *першому розділі* здійснено аналіз основних особливостей організації та виконання процесу будівельно-технічної експертизи в сучасних умовах, зокрема після початку масштабних руйнувань інфраструктури України внаслідок бойових дій; окреслено основні етапи роботи експертів; визначено та описано проблеми, пов'язані з неефективною організацією зберігання, обробки й повторного використання даних у цьому процесі; сформульовано та описано ключові проблеми роботи з даними в інформаційних високонавантажених системах, визначено вплив цих проблем на продуктивність і складність їх вирішення; запропоновано підходи до вирішення виявлених проблем роботи з даними шляхом впровадження нереляційних баз даних і технології кешування як ефективних рішень, що здатні забезпечити гнучкість, масштабованість і швидкий доступ до даних; наведено порівняльний аналіз реляційних та нереляційних баз даних для кращого обґрунтування доцільності використання нереляційних баз даних; наведено загальну інформацію про технологію кешування; описано потенційні переваги для роботи з даними в разі використання цієї технології для будівельно-технічної експертизи та інформаційних високонавантажених систем; проведений аналіз слугує основою для формування вимог до роботи з даними в інформаційній системі підтримки процесу відновлення пошкоджених об'єктів нерухомості; виконано постановку задачі.

У *другому розділі* визначено основні функції, структурні та технічні вимоги для бази даних системи підтримки процесу відновлення пошкоджених об'єктів нерухомості; описано основні типи запитів на читання та запис, які виконуватимуть користувачі цієї системи з різними рівнями доступу; обґрунтовано пріоритетні властивості бази даних системи відповідно до CAP-теореми; визначено ключові критерії, за якими оцінюється доцільність використання різних типів і технологій нереляційних баз даних для потреб системи; обґрунтовано доцільність використання СУБД MongoDB як бази даних для системи підтримки процесу відновлення пошкоджених об'єктів нерухомості на основі порівняльного аналізу з іншими типами та технологіями нереляційних баз даних; спроектовано структуру основних

колекцій документів бази даних MongoDB для системи, зокрема: «Пошкоджені будівлі», «Міська статистика», «Користувачі» та «Історія запитів»; побудовано схему логічних взаємозв'язків між цими колекціями як основу для розробки гібридного методу інтелектуального управління кешем, що дозволяє суттєво підвищити ефективність роботи з даними; обґрунтовано доцільність використання інтелектуального кешування як засіб подолання обмежень існуючих методів кешування; описано роль кешування в сучасних інформаційних високонавантажених системах.

У *третьому розділі* розроблено гібридний метод інтелектуального управління кешем, який вирішує три ключові задачі: оцінка доцільності додавання об'єкта до кешу, вибір об'єкта для видалення з заповненого кешу та періодична інвалідація неактуальних об'єктів у кеші; подано визначення методу управління кешем як основи для реалізації гібридного підходу; обґрунтовано застосування моделей машинного навчання – табличного Q-Learning та логістичної регресії для вирішення відповідних задач запропонованого методу інтелектуального управління кешем; запропоновано представлення стану об'єкта у вигляді вектора ознак; спроектовано додаткові колекції документів для бази даних системи підтримки процесу відновлення пошкоджених об'єктів нерухомості MongoDB, необхідні для коректної роботи запропонованих моделей машинного навчання; обґрунтовано доцільність використання Redis як основної технології кешування для системи; описано алгоритм роботи гібридного методу інтелектуального управління кешем; показано фінальну блок-схему його роботи із застосуванням моделей машинного навчання табличного Q-Learning і логістичної регресії, технології нереляційної бази даних MongoDB і технології кешування Redis; представлено оновлену модель системи з інтегрованим модулем інтелектуального кешування, що реалізовує розроблений гібридний метод.

У *четвертому розділі* описано структуру та основні функції системи експериментального тестування ефективності запропонованого гібридного методу інтелектуального управління кешем; наведено детальний опис компонентів системи, зокрема бази даних MongoDB, кеш Redis, двох моделей машинного навчання –

табличного Q-Learning і логістичної регресії та модуля керування кешем, що реалізовує алгоритм роботи гібридного методу; реалізовано повноцінну програмну архітектуру системи, включно з агентами, буферами, оновленням моделей у фоновому асинхронному режимі та синхронізацією з базою даних MongoDB і кешем Redis; продемонстровано результати роботи системи у двох типах експериментів: аналіз частки кеш-хітів після одного запуску та серії запусків з накопичувальним навчанням моделей з фіксованим набором даних; у кожному експерименті проведено порівняння ефективності запропонованого гібридного методу управління кешем з найпоширенішими існуючими методами кешування: LRU та LFU; візуалізовано результати у вигляді графіків і рисунків, що підтверджують перевагу гібридного методу за умов обмеженого розміру кешу та кількості запитів.

Ключові слова: база даних, будівельно-технічна експертиза, інтелектуальне кешування, інформаційна високонавантажена система, машинне навчання, нереляційна база даних, технологія кешування, штучний інтелект.

ABSTRACT

Volokh B. Intelligent Data Caching for the System Supporting the Process of Damaged Buildings Restoration. – Qualifying scientific work presented as a manuscript.

Dissertation for obtaining the scientific degree of Doctor of Philosophy in specialty 126 "Information systems and technologies". - Kyiv National University of Construction and Architecture. - Kyiv, 2025.

The dissertation is devoted to solving the scientific and practical problem of improving the efficiency of data processing and storage in the information infrastructure of the architecture and construction industry.

The scientific novelty of the obtained results. The scientific novelty of the dissertation consists in the fact that:

developed for the first time:

- the hybrid method of intelligent cache management that solves the key caching tasks, is based on machine learning models, and takes into account the essential characteristics of users and damaged real estate objects;

improved:

- the model of the information system supporting the process of damaged buildings restoration by integrating an intelligent caching module, which improves a performance of the system and provides efficient access to high-priority data;

- the process of providing information support to the users of the system supporting the process of damaged buildings restoration by expanding the collections of documents for a non-relational database designed to store data about damaged real estate objects, history of user requests, and damage statistics;

further advance:

- approaches to data processing and storage in the information infrastructure of the architecture and construction industry by combining intelligent caching technologies with non-relational databases.

Main content of the dissertation

The hybrid method of intelligent cache management is presented, focused on highly loaded information systems with dynamically changing requests, in particular on the system supporting the process of damaged buildings restoration. The method covers three key tasks: assessing the expediency of adding an object to the cache, selecting an object for deletion when the cache is full, and periodically invalidating obsolete objects in the cache. To solve these tasks, a combination of machine learning models is used: table-based Q-Learning with a mechanism for deferred asynchronous updating of model values and logistic regression with a mechanism for deferred asynchronous accumulation of a buffer of training examples for further training of the model and periodic updating of its parameters. An experimental evaluation of the method's effectiveness was performed, which confirmed the advantage of the intelligent approach over existing LRU and LFU caching methods, particularly in conditions of limited input data size. The proposed method has high potential for usage in other information systems due to the ability to adapt the feature vector of machine learning models to the features of objects and the context of their usage.

Based on the results of the study, the following have been developed:

- the hybrid method of intelligent cache management that solves the key caching tasks, is based on machine learning models, and takes into account the essential characteristics of users and damaged real estate objects;
- the system for experimental testing of the effectiveness of the hybrid intelligent cache management method, which includes the modules for simulating requests, implementing logic of the method, and interacting with the machine learning models, the caching technology, and the non-relational database.

In the first chapter, the main features of organizing and performing the process of building-technical expertise in modern conditions are analyzed, with particular attention to the challenges that have emerged after the beginning of large-scale infrastructure destruction in Ukraine as a result of military operations; the key stages of expert activities are outlined; problems related to inefficient organization of data storage, processing, and reuse in this process are identified and described; the key problems of working with data in high-load information systems are formulated and described, and the impact of these

problems on performance and the complexity of their solution are determined; approaches to solving identified data processing issues have been proposed through the integration of non-relational databases and caching technology as effective solutions capable of ensuring flexibility, scalability, and efficient access to data; the comparative analysis of relational and non-relational databases is provided to better substantiate the appropriateness of using non-relational databases; general information about caching technology is provided; the potential advantages for working with data when using this technology for construction and technical expertise and high-load information systems are described; the conducted analysis serves as the basis for forming requirements for data processing in the information system supporting the process of damaged buildings restoration; the task is set.

In the second chapter, the main functions, structural and technical requirements for the database of the system supporting the process of damaged buildings restoration are defined; the primary types of read and write queries expected to be performed by users with varying levels of access are described; the priority properties of the database are substantiated in accordance with the CAP theorem; the key criteria for assessment the feasibility of using different types and technologies of non-relational databases for the needs of the system are defined; the expediency of using MongoDB as database for the system supporting the process of damaged buildings restoration based on a comparative analysis with other types and technologies of non-relational databases is substantiated; the structure of the main MongoDB document collections for the system is designed, including “Damaged Buildings,” “Urban Statistics,” “Users,” and “Query History”; the scheme of logical relationships between these collections is built as a basis for the development of a hybrid method of intelligent cache management, which can significantly increase the efficiency of data processing; the expediency of using intelligent caching as a means of overcoming the limitations of existing caching methods is substantiated; the role of caching in modern information high-load systems is described.

In the third chapter, the hybrid method of intelligent cache management is developed, which solves three key tasks: assessing the feasibility of adding an object to the cache, selecting an object to remove from the full cache, and periodically invalidating

obsolete objects in the cache; the definition of the cache management method as a basis for the implementation of a hybrid approach is presented; the application of machine learning models - table-based Q-Learning and logistic regression for solving the relevant tasks of the proposed method of intelligent cache management is substantiated; a feature vector representation of the object state is proposed; additional collections of documents for the MongoDB database of the system supporting the process of damaged buildings restoration were designed, which are necessary for the correct functioning of the proposed machine learning models; the expediency of using Redis as the main caching technology for the system is substantiated; the algorithm of the hybrid method of intelligent cache management is described in detail; a final block diagram of its operation is presented, illustrating the use of table-based Q-Learning and logistic regression machine learning models, the non-relational MongoDB database, and the Redis caching technology; an updated model of the system with the integrated module of intelligent caching that implements the developed hybrid method is presented.

In the fourth chapter, the structure and core functionalities of the experimental testing system for evaluating the effectiveness of the proposed hybrid method of intelligent cache management are described; a detailed description of the system components, including the MongoDB database, the Redis cache, two machine learning models - table-based Q-Learning and logistic regression, and the cache management module that implements the hybrid method algorithm is presented; a complete software architecture of the system was implemented, including agents, buffers, updating models in the background asynchronously and synchronization with the MongoDB database and Redis cache; the results of the system are demonstrated in two types of experiments: analysis of the proportion of cache hits after one run and a series of runs with accumulative training of models with a fixed data set; in each experiment, the comparison of the efficiency of the proposed hybrid cache management method with the most common existing caching methods LRU and LFU is made; the results are visualized in the form of graphs and figures, which confirm the advantage of the hybrid method under conditions of limited cache size and number of requests.

Keywords: database, building-technical expertise, intelligent caching, high-load information system, machine learning, non-relational database, caching technology, artificial intelligence.