

Облікова картка дисертації

I. Загальні відомості

Державний обліковий номер: 0826U005127

Особливі позначки: відкрита

Дата реєстрації: 04-09-2026

Статус: Запланована

Реквізити наказу МОН / наказу закладу:



II. Відомості про здобувача

Власне Прізвище Ім'я По-батькові:

1. Зважій Дмитро Володимирович

2. Dmytro Zvazhii

Кваліфікація:

Ідентифікатор ORCID ID: Не застосовується

Вид дисертації: доктор філософії

Аспірантура/Докторантура: так

Шифр наукової спеціальності: 122

Назва наукової спеціальності: Комп'ютерні науки

Галузь / галузі знань: інформаційні технології

Освітньо-наукова програма зі спеціальності: Комп'ютерні науки

Дата захисту: 08-09-2026

Спеціальність за освітою: Комп'ютерні науки та інформаційні технології

Місце роботи здобувача: Національний університет "Киево-Могилянська академія"

Код за ЄДРПОУ: 16459396

Місцезнаходження: вул. Г. Сковороди, Київ, 04070, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: 03wfca816

III. Відомості про організацію, де відбувся захист

Шифр спеціалізованої вченої ради (разової спеціалізованої вченої ради): PhD 16437

Повне найменування юридичної особи: Національний університет "Кієво-Могилянська академія"

Код за ЄДРПОУ: 16459396

Місцезнаходження: вул. Г. Сковороди, Київ, 04070, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: 03wfca816

IV. Відомості про підприємство, установу, організацію, в якій було виконано дисертацію

Повне найменування юридичної особи: Національний університет "Кієво-Могилянська академія"

Код за ЄДРПОУ: 16459396

Місцезнаходження: вул. Г. Сковороди, Київ, 04070, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: 03wfca816

V. Відомості про дисертацію

Мова дисертації: Українська

Коди тематичних рубрик: 20.19.17, 20.23.17, 20.53.19

Тема дисертації:

1. Індексний доступ на основі суфіксних масивів для підрядкового пошуку в PostgreSQL
2. Suffix-array-based index access for substring search in PostgreSQL

Реферат/анотація:

1. Зважій Д. В. Індексний доступ на основі суфіксних масивів для підрядкового пошуку в PostgreSQL. — Кваліфікаційна наукова праця на правах рукопису. Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 122 «Комп'ютерні науки» (12 — Інформаційні технології). — Національний університет «Кієво-Могилянська академія», Київ, 2026. Дисертацію присвячено розробці методу індексного доступу для підрядкового пошуку, безпосередньо вбудованого в реляційну систему керування базами даних (СКБД). Підрядковий пошук, тобто знаходження усіх рядків бази даних, що містять заданий шаблон як префікс, як суфікс або як довільне входження, є фундаментальною операцією інформаційно-пошукових систем, проте у наявних реляційних СКБД він підтримується лише фрагментарно: вбудовані механізми PostgreSQL (Б-дерево з text_pattern_ops, GIN із pg_trgm, повнотекстовий пошук) охоплюють лише часткові випадки задачі, а зовнішні рішення на основі суфіксного індексу (зокрема SAES над Elasticsearch) не дають внутрішньої підтримки в самій реляційній СКБД. Це робить актуальною розробку методу індексного доступу, безпосередньо вбудованого в реляційну СКБД, який єдиною індексною структурою підтримує усі три

операції підрядкового пошуку, визначений у термінах інваріантів цієї платформи та апробований у PostgreSQL. Об'єктом дослідження є процес виконання операцій підрядкового пошуку в реляційних системах керування базами даних. Предметом дослідження є моделі, методи та програмні засоби індексного доступу на основі суфіксних структур даних, що підтримують виконання операцій підрядкового пошуку у середовищі реляційної СКБД-платформи. Метою дослідження є побудова теоретичної моделі, розробка та експериментальне обґрунтування методу індексного доступу на основі суфіксного масиву для підтримки операцій підрядкового пошуку в PostgreSQL. Метод має гарантувати логарифмічну складність пошуку $O(m \log N)$, однакову для трьох операцій (за префіксом, суфіксом та входженням), повну інтеграцію з менеджером буферів, журналюванням із записом наперед та програмним інтерфейсом методів доступу PostgreSQL і коректну взаємодію з моделлю багатоверсійного керування паралельністю на шляху читання. Для досягнення мети виконано аналіз наявних суфіксних структур даних і механізмів PostgreSQL, побудовано теоретичну модель сторінкового суфіксного масиву в моделі блочного зовнішнього зберігання, розроблено метод індексного доступу SAindex та його програмну реалізацію в ядрі PostgreSQL, проведено порівняльне експериментальне оцінювання на реальних і синтетичних даних. Застосовано методи аналізу алгоритмів у цій моделі, системної інтеграції через Access Method API PostgreSQL і системного програмування мовою C, а також непараметричний критерій Манна-Вітні з контролем розміру ефекту через дельту Кліффа. Наукова новизна одержаних результатів полягає в такому. Уперше розроблено теоретичну модель сторінкового суфіксного масиву в системі параметрів pB, e, e_k, Np , яка зводить три операції підрядкового пошуку до спільного діапазонного примітива з позиційною фільтрацією. Уперше доведено межі I/O-складності методу: верхню межу пошукової фази $\leq 2p \log p + O(1)$ блочних доступів незалежно від типу операції, лінійні нижні межі фази збирання результатів і повноту покриття множин результатів. Уперше встановлено просторову складність методу як $N(e+8,125)+B$ байтів. Уперше запропоновано метод індексного доступу, повний цикл функціонування якого специфіковано в термінах загальних інваріантів реляційної СКБД, що уможливорює його перенесення на будь-яку СКБД, яка задовольняє ці інваріанти. Дістала подальшого розвитку інтеграція суфіксних структур у систему типів і операторів реляційної СКБД через клас операторів `text_sa_ops` і два нові оператори підрядкового пошуку. Практичне значення та апробація методу. Метод апробовано програмною реалізацією SAindex (2782 рядки коду мовою C), інтегрованою в ядро PostgreSQL 18devel. На основі критерію Манна-Вітні з контролем розміру ефекту через дельту Кліффа встановлено перевагу SAindex над триграмним методом GIN_TRGM із прискоренням 6,3× у медіані для запитів за входженням і 46,5× для префіксних запитів, перевагу над спеціалізованими B-деревами в режимах низької селективності та коректність результатів у 160 з 160 тестових випадків. Дисертація складається з п'яти розділів: аналітичний огляд методів пошуку підрядків; експериментальна оцінка наявних механізмів підрядкового пошуку в PostgreSQL; теоретична модель індексного доступу на основі суфіксного масиву в зовнішній пам'яті; проектування методу SAindex для PostgreSQL; опис реалізації SAindex та її експериментальної оцінки. Ключові слова: суфіксний масив, дерева, B-дерево, підрядковий пошук, обробка природної мови, метод індексного доступу, модель зовнішньої пам'яті, PostgreSQL, попарні порівняння, булева алгебра, прийняття рішень, оцінювання, архітектура системи, комп'ютерна система, інформаційна технологія.

2. Zvazhii D. V. Suffix-array-based index access for substring search in PostgreSQL. — Qualification scientific work submitted as a manuscript. PhD thesis for the degree of Doctor of Philosophy in speciality 122 “Computer Science” (field of knowledge 12 “Information Technologies”). — National University of Kyiv-Mohyla Academy, Kyiv, 2026. The dissertation is devoted to the development of an index access method for substring search natively integrated into a relational database management system. Substring search — locating all database strings that contain a given pattern as a prefix, as a suffix, or as an arbitrary occurrence — is a foundational operation of information-retrieval systems, yet existing relational DBMSs support it only fragmentarily: the built-in PostgreSQL mechanisms (the B-tree with `text_pattern_ops`, GIN with `pg_trgm`, full-text search) cover only partial cases of the problem, and external solutions based on a suffix index (notably SAES over Elasticsearch) do not provide in-DBMS support inside the relational system itself. This motivates the development of an index access method, natively integrated into a relational DBMS, that supports all three substring operations through a single index structure, is specified in terms

of the invariants of that platform, and is validated on PostgreSQL. The object of the research is the process of executing substring-search operations in relational database management systems. The subject of the research is the models, methods, and software facilities of index access based on suffix data structures that support the execution of substring-search operations within a relational DBMS platform. The aim of the research is to construct a theoretical model, design, and experimentally substantiate a suffix-array-based index access method supporting substring-search operations in PostgreSQL. The method is required to deliver logarithmic search complexity $O(m \log N)$, uniform across the three operations (prefix, suffix, and contains), full integration with the buffer manager, write-ahead log, and access-method interface of PostgreSQL, and correct interoperation with multi-version concurrency control on the read path. To attain this aim, the existing suffix data structures and PostgreSQL mechanisms were analysed, a theoretical model of a paged suffix array in the external-memory block model was constructed, the SAindex method and its implementation in the PostgreSQL core were developed, and a comparative experimental evaluation on real and synthetic data was carried out. The work applies algorithm analysis in that model, systems integration via the PostgreSQL Access Method API, systems programming in C, and the non-parametric Mann-Whitney U test with effect-size control via Cliff's delta. Scientific novelty of the obtained results. For the first time, a theoretical model of a paged suffix array has been developed within the parameter system $\alpha B, \epsilon, \epsilon_k, N, \beta$, which reduces the three substring-search operations to a single range-search primitive with positional filtering. For the first time, the I/O complexity bounds of the method have been proven: an upper bound on the search phase of $\leq 2\alpha \log \beta + O(1)$ block accesses regardless of the operation type, linear lower bounds on the result-collection phase, and completeness of result-set coverage, together with a space complexity of $N(\epsilon + 8.125) + B$ bytes. For the first time, an index access method is proposed whose complete lifecycle is specified in terms of the general invariants of a relational DBMS, transferable to any DBMS satisfying these invariants. Further developed is the integration of suffix structures into the type-and-operator system through the `text_sa_ops` operator class and two new substring-search operators. Practical value and validation of the method. The method has been validated through a software implementation, SAindex (2782 lines of C code), integrated into the PostgreSQL 18-devel core. A Mann-Whitney U test with effect-size control via Cliff's delta established the dominance of SAindex over the trigram method GIN_TRGM with a median speedup of $6.3\times$ for contains queries and $46.5\times$ for prefix queries, an advantage over specialised B-trees under low-selectivity regimes, and correctness on 160 of 160 test cases. The dissertation consists of five chapters: an analytical review of substring-search methods; an experimental evaluation of the existing substring-search mechanisms in PostgreSQL; a theoretical model of index access based on a suffix array in external memory; the design of the SAindex method for PostgreSQL; and a description of the SAindex implementation and its experimental evaluation. Keywords: suffix array, trees, B-tree, substring search, natural language processing, index access method, external-memory model, PostgreSQL, pairwise comparisons, Boolean algebra, decision making, evaluation, system architecture, computer system, information technology.

Державний реєстраційний номер ДіР:

Пріоритетний напрям розвитку науки і техніки: Фундаментальні наукові дослідження з найбільш важливих проблем розвитку науково-технічного, соціально-економічного, суспільно-політичного, людського потенціалу для забезпечення конкурентоспроможності України у світі та сталого розвитку суспільства і держави

Стратегічний пріоритетний напрям інноваційної діяльності: Не застосовується

Підсумки дослідження: Нове вирішення актуального наукового завдання

Публікації:

- A. Hlybovets and D. Zvazhii, "Implementation of a suffix tree-based index for searching for substrings in a large DBMS," *Cybernetics and Systems Analysis*, vol. 61, no. 2, pp. 265–275, Mar. 2025.

- Д. В. Зважій, «Особливості індексації у PostgreSQL», Наукові записки НаУКМА. Комп'ютерні науки, т. 8, с. 113–117, лист. 2025.
- Д. В. Зважій та С. С. Гороховський, «Впровадження індексів у PostgreSQL», Проблеми програмування, № 1, с. 14–24, 2026.
- Д. В. Зважій та О. М. Малій, «Побудова розподілених суфіксних дерев», Технічна інженерія, № 1(97), с. 210–216, лют. 2026.
- Д.В. Зважій, «Використання суфіксних дерев для покращення операцій по шукувсистемахуправліннябазамиданих», уМатеріалиXIIВсеукраїнської наукової конференції молодих математиків, Київ, 9–11 трав. 2024.
- Д.В.Зважій,«Перспектививпровадженняіндексунабазісуфіксногодерева для пошуку підрядків у СУБД великого розміру», у Матеріали XVI Міжна родної науково-практичної конференції «Теоретичні та прикладні аспекти побудови програмних систем» TAAPSD'2025, Київ, 24–25 лист. 2025.
- Д.В. Зважій, «Застосування суфіксних масивів для розробки індексів у PostgreSQL», у Матеріали Міжнародної науково-практичної конференції здобувачів вищої освіти і молодих вчених «Інформаційні технології: теорія і практика», Харків, 25–27 бер. 2026.

Наукова (науково-технічна) продукція:

Соціально-економічна спрямованість:

Охоронні документи на ОПІВ:

Впровадження результатів дисертації: Впроваджено

Зв'язок з науковими темами: 0118U000647

VI. Відомості про наукового керівника/керівників (консультанта)

Власне Прізвище Ім'я По-батькові:

1. Глибовець Андрій Миколайович
2. Andrii Glybovets

Кваліфікація: д. т. н., професор, 01.05.03

Ідентифікатор ORCID ID: 0000-0003-4282-481X

Додаткова інформація:

Повне найменування юридичної особи: Національний університет "Кієво-Могилянська академія"

Код за ЄДРПОУ: 16459396

Місцезнаходження: вул. Г. Сковороди, Київ, 04070, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: 03wfca816

VII. Відомості про офіційних опонентів та рецензентів

Офіційні опоненти

Власне Прізвище Ім'я По-батькові:

1. Пасічник Володимир Володимирович
2. Volodymyr Pasichnyk

Кваліфікація: д. т. н., професор, 01.05.01

Ідентифікатор ORCID ID: 0000-0002-5231-6395

Додаткова інформація:

Повне найменування юридичної особи: Національний університет "Львівська політехніка"

Код за ЄДРПОУ: 02071010

Місцезнаходження: вул. Степана Бандери, Львів, 79013, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

Власне Прізвище Ім'я По-батькові:

1. Погорілий Сергій Дем'янович
2. Sergiy Pogorilyy

Кваліфікація: д. т. н., професор, 05.13.05

Ідентифікатор ORCID ID: 0000-0002-6497-5056

Додаткова інформація:

Повне найменування юридичної особи: Київський національний університет імені Тараса Шевченка

Код за ЄДРПОУ: 02070944

Місцезнаходження: вул. Володимирська, Київ, 01033, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

Рецензенти**Власне Прізвище Ім'я По-батькові:**

1. Франчук Олег Васильович
2. Oleg Franchuk

Кваліфікація: к. т. н., доц., 05.13.06

Ідентифікатор ORCID ID: 0000-0002-1122-4689

Додаткова інформація:

Повне найменування юридичної особи: Національний університет "Кієво-Могилянська академія"

Код за ЄДРПОУ: 16459396

Місцезнаходження: вул. Г. Сковороди, Київ, 04070, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: 03wfca816

Власне Прізвище Ім'я По-батькові:

1. Олецкий Олексій Віталійович

2. Oleksiy Oletsky

Кваліфікація: к. т. н., доц., 05.13.06

Ідентифікатор ORCID ID: 0000-0002-0553-5915

Додаткова інформація:

Повне найменування юридичної особи: Національний університет "Кієво-Могилянська академія"

Код за ЄДРПОУ: 16459396

Місцезнаходження: вул. Г. Сковороди, Київ, 04070, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: 03wfca816

VIII. Заключні відомості

**Власне Прізвище Ім'я По-батькові
голови ради**

Олійник Богдана Віталіївна

**Власне Прізвище Ім'я По-батькові
головуючого на засіданні**

Олійник Богдана Віталіївна

**Відповідальний за підготовку
облікових документів**

Басенко Олена Едуардівна

Реєстратор

Юрченко Тетяна Анатоліївна

**Керівник відділу УкрІНТЕІ, що є
відповідальним за реєстрацію наукової
діяльності**



Юрченко Тетяна Анатоліївна