

УДК: 004.738.5:001.102

Кузьмич Анна Сергіївна
(*наук. керівник – старший викладач Стаценко Д. В.*)
Національний університет «Полтавська політехніка
імені Юрія Кондратюка», м. Полтава

ПРОБЛЕМА ДУБЛЮВАННЯ ДАНИХ У ЦИФРОВОМУ ПРОСТОРІ ТА ШЛЯХИ ЇЇ ВИРІШЕННЯ

Анотація. Досліджено проблему дублювання даних у цифровому просторі та її вплив на якість інформаційних ресурсів. Проаналізовано причини та наслідки виникнення дублікатів. Обґрунтовано необхідність комплексного підходу до вирішення проблеми, що передбачає впровадження сучасних технологій обробки та управління даними.

Ключові слова: дублювання даних, інформаційні ресурси, цифровий простір, якість даних, інформаційні системи.

Однією з характерних рис цифрового простору є його постійне наповнення різноманітною інформацією, що зумовлює розширення накопичувальних можливостей баз даних та вдосконалення платформ для зберігання інформації. У зв'язку з цим постає актуальне питання пошуку ефективного збереження значних обсягів інформації за умов її безперервного надходження та перевищення темпів видалення застарілих даних.

Сьогодні інформація охоплює всі сфери суспільного життя, від загальнодержавних новин до міжособистісного спілкування. Водночас у пересічного користувача часто виникають певні труднощі під час пошуку необхідної інформації, пов'язані з отриманням великої кількості дублікатів або перекручених даних. Це пов'язано з тим, що інформація проходить складний шлях від першоджерела до кінцевого споживача. У цьому процесі доволі складно зберегти точність і об'єктивність викладу, натомість легко виникають спотворення, подібні до ефекту «зіпсованого телефону». Саме в таких умовах дублювання даних стає поширеним явищем.

Проблема дублювання даних полягає у неодноразовому збереженні ідентичної або подібної інформації в різних інформаційних системах, базах даних чи навіть у межах однієї системи. Це явище зумовлене відсутністю єдиних стандартів обліку даних, недостатньою координацією між підрозділами, використанням програмних продуктів, які не є інтегровані між собою [3]. На практиці це проявляється у повторному введенні однакових даних, створенні кількох записів одного й того самого об'єкта, зберіганні застарілих копій інформації.

Явище дублювання даних негативно впливає на якість інформаційних ресурсів. По-перше, воно ускладнює процеси пошуку та обробки інформації, змушуючи користувача працювати зі значною кількістю однотипних записів. По-друге, виникає ризик спотворення інформації, внаслідок чого різні копії одного й того самого повідомлення містять певні розбіжності. Як наслідок – усе це призводить до неефективного використання ресурсів та ускладнення прийняття обґрунтованих рішень.

Проблема дублювання даних має не лише теоретичний, а й практичний характер. Зокрема, результати досліджень великих медичних реєстрів свідчать про

те, що приблизно 30 % записів можуть містити дублікати, що суттєво впливає на точність обробки інформації [4]. До того ж до 92 % дублікатів виникають на етапі введення даних, що свідчить про значну роль людського чинника. У сфері інформаційних технологій ситуація є ще більш критичною, адже у деяких наборах даних частка дубльованої інформації може досягати показника 99 %, що значно ускладнює процес аналізу та прийняття рішень.

На нашу думку, проблема дублювання даних у цифровому просторі має системний характер і потребує комплексного підходу до її вирішення. Насамперед важливим є впровадження єдиних стандартів ведення даних та уніфікації форматів їх зберігання, що забезпечить узгодженість інформації між різними системами. Також доцільним є використання інтегрованих інформаційних систем, які забезпечують централізоване зберігання та синхронізацію даних. Важливу роль також відіграє впровадження технологій автоматичного виявлення та усунення дублікатів [2].

Можна виокремити чотири ефективні способи боротьби з дублюванням даних:

1. Fuzzy Matching (нечітке зіставлення), що передбачає використання алгоритмів (наприклад, відстань Левенштейна або Soundex) для виявлення схожих записів навіть за відсутності повного збігу.

2. Зв'язування записів (record linkage), що застосовується для об'єднання даних із різних джерел у єдиний узгоджений запис.

3. Управління та стандартизація даних, що передбачають впровадження чітких правил введення, уніфікації форматів та контролю якості даних, внаслідок чого досягається мінімізація ймовірності створення дублікатів.

4. Використання сучасних інструментів (зокрема AWS Glue), що забезпечує автоматизовану обробку, трансформацію та дедуплікацію даних із використанням алгоритмів і сценаріїв ETL. За допомогою спеціальних скриптів Glue можна визначити логіку дедуплікації на основі конкретних атрибутів або унікальних ідентифікаторів. Це гарантує, що в озері даних або сховищі даних зберігаються лише правильні та унікальні записи [5].

Сучасні програмні рішення дають змогу ефективно аналізувати великі масиви даних, виявляти збіги за ключовими параметрами та пропонувати варіанти їх об'єднання. Водночас не менш важливим є підвищення рівня інформаційної культури користувачів, оскільки значна частина дублювання зумовлена людським чинником [1].

Отже, проблема дублювання даних у сучасному цифровому просторі є наслідком стрімкого зростання обсягів інформації, недосконалості процесів її обробки та значного впливу людського чинника. Вона має системний характер і проявляється як на рівні окремих інформаційних систем, так і в масштабах усього цифрового середовища, ускладнюючи доступ до якісної та достовірної інформації. Ефективне подолання дублювання даних можливе лише за умов поєднання технологічних та організаційних підходів. Упровадження сучасних інструментів обробки та дедуплікації даних, стандартизація процесів їх введення, підвищення інформаційної культури користувачів сприятимуть зменшенню кіль-

кості дублікатів і підвищенню якості інформаційних ресурсів. Це забезпечить ефективне функціонування цифрових платформ та покращить процеси прийняття рішень у різних сферах діяльності.

Список використаних джерел

1. Практичні поради щодо уникнення помилок та дублювання при електронному документообігу. *iFin Бухгалтерія*. URL: <https://www.ifin.in.ua/edm-articles> (дата звернення: 05.04.2026).
2. Руденко Г. Методи нормалізації даних при впровадженні ERP-систем: реальні кейси, технічні аспекти та ризики. *SPEKA – інтернет видання про підприємництво та технології, бізнес-тренди, інновації та розвиток IT*. URL: <https://speka.ua/business/metodi-normalizaciyi-danix-pri-vprovadzenni-erp-sistem-realni-keisi-texnicni-aspekti-ta-riziki-rbmqgw> (дата звернення: 02.04.2026).
3. Що таке дедуплікація даних – терміни та визначення у сфері кібербезпеки. *Keepsolid VPN Unlimited*. URL: <https://surl.li/ycrujq> (дата звернення: 04.04.2026).
4. Duplicate patient records in large healthcare databases: challenges and solutions. *PubMed Central*. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC12550146/> (дата звернення: 01.04.2026).
5. Проблема дублювання даних. *LinkedIn Україна*. URL: <https://ua.linkedin.com/pulse/data-duplication-predicament-kshitija-kj-gupte?tl=uk> (дата звернення: 02.04.2026).



УДК: 004.9:378

Левченко Марина Русланівна

**(наук. керівник – канд. техн. наук, старший викладач Поремський Ю. В.)
Донецький національний університет імені Василя Стуса, м. Вінниця**

ІНФОРМАЦІЙНА СИСТЕМА ЯК ЗАСІБ ПІДТРИМКИ ПІДГОТОВКИ ДО КВАЛІФІКАЦІЙНОГО ІСПИТУ З КОМП'ЮТЕРНИХ НАУК

Анотація. У тезах розглянуто інформаційну систему як засіб підтримки підготовки до кваліфікаційного іспиту з комп'ютерних наук. Охарактеризовано її призначення, основні функції та практичну цінність. Показано, що використання системи сприяє систематизації знань, самоконтролю та підвищенню ефективності підготовки студентів.

Ключові слова: інформаційна система, кваліфікаційний іспит, комп'ютерні науки, тестування, підготовка студентів.

Нині підготовка здобувачів вищої освіти до кваліфікаційного іспиту з комп'ютерних наук вимагає не лише опрацювання значного обсягу теоретичного матеріалу, а й використання сучасних цифрових засобів, здатних забезпечити системність, послідовність і зручність навчального процесу. Умови сучасної освіти характеризуються постійним зростанням обсягу навчальної інформації, потребою в оперативному повторенні ключових тем та необхідністю регулярного самоконтролю знань [1]. За таких обставин застосування спеціалізованої інформаційної системи, орієнтованої на підтримку підготовки до кваліфікаційного іспиту, є доцільним і актуальним. Традиційний підхід до підготовки, що базується на використанні розрізнених конспектів, електронних матеріалів, тестів і сторонніх ресурсів, не завжди забезпечує належний рівень упорядкованості

XI Всеукраїнська наукова студентська конференція

«Інформаційні технології і системи в документознавчій сфері» (м. Вінниця, 09 квітня 2026 р.)