

**Гліб Олександрович Шевченко<sup>1</sup>, Інна Олександрівна Саяпіна<sup>2</sup>**

<sup>1</sup>магістрант кафедри програмного забезпечення комп'ютерних систем  
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського» (Київ, Україна)  
E-mail: [shevchenko.h.o.-kp41mn@edu.kpi.ua](mailto:shevchenko.h.o.-kp41mn@edu.kpi.ua). ORCID: <https://orcid.org/0009-0005-5733-5826>

<sup>2</sup>кандидат технічних наук, доцент, доцент кафедри програмного забезпечення комп'ютерних систем  
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського» (Київ, Україна)  
E-mail: [saiapina@pzks.fpm.kpi.ua](mailto:saiapina@pzks.fpm.kpi.ua). ORCID: <https://orcid.org/0000-0003-1541-1681>  
Researcher ID: [P-6639-2018](https://orcid.org/P-6639-2018). SCOPUS Author ID: [57221620283](https://orcid.org/57221620283)

**МЕТОД ТА ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ АВТОМАТИЗАЦІЇ ПРОЦЕСІВ  
УПРАВЛІННЯ ДАНИМИ В ГЕТЕРОГЕННИХ БАЗАХ ДАНИХ**

Зростання обсягів і різноманіття форматів даних у сучасних інформаційних системах призводить до необхідності інтеграції гетерогенних джерел, представлених реляційними, документоорієнтованими та іншими моделями зберігання. У таких умовах ускладнюється забезпечення узгодженості, якості та актуальності інформації, що вимагає автоматизації процесів управління даними. У роботі досліджується задача побудови методу автоматизованого управління гетерогенними даними на основі формалізованої моделі інтеграції. Запропонований метод описує процес інтеграції як композицію перетворень над множиною джерел даних, що включає синтаксичну гармонізацію структур, нормалізацію значень, семантичне зіставлення атрибутів на основі правил відповідності та дедуплікацію записів із використанням функцій подібності. Реалізація методу передбачає використання модульної архітектури програмного забезпечення, у якій етапи обробки даних виконуються послідовно та незалежно від типів джерел. Для зберігання інтегрованих даних застосовується реляційна база даних із підтримкою напівструктурованих форматів, що забезпечує гнучкість моделі та масштабованість системи. У межах підходу також визначено критерії оцінювання якості даних та механізми їх актуалізації у динамічних середовищах. Запропонований метод дозволяє підвищити рівень узгодженості та достовірності інтегрованих даних, скоротити час їх обробки та забезпечити ефективне функціонування систем у розподілених середовищах.

**Ключові слова:** гетерогенні бази даних; інтеграція даних; автоматизоване управління даними; семантичне узгодження; дедуплікація; якість даних; формалізована модель; розподілені системи.

Рис.: 3. Табл.: 1. Бібл.: 10.

**Актуальність теми дослідження** зумовлена стрімким зростанням обсягів даних у сучасних інформаційних системах та необхідністю їх ефективного опрацювання в умовах використання різнорідних джерел, форматів і моделей зберігання. Гетерогенність баз даних ускладнює забезпечення узгодженості, достовірності, актуальності та якості інформації, а також потребує значних витрат часу на ручне налаштування процесів інтеграції, очищення й оновлення даних. У зв'язку з цим важливим науково-практичним завданням є розроблення методів і програмних засобів автоматизації управління даними в гетерогенних середовищах, які дозволяють зменшити дублювання, усунути семантичні та структурні невідповідності, підвищити якість інтегрованих даних і забезпечити масштабованість інформаційних систем.

**Постановка проблеми.** У сучасних інформаційних системах дані формуються з великої кількості різнорідних джерел, що відрізняються за структурою, форматом і змістом. У результаті виникають гетерогенні середовища, в яких поєднуються різні моделі зберігання та представлення інформації. Об'єднання таких даних у єдину систему супроводжується труднощами, пов'язаними з невідповідністю структур, розбіжностями у семантиці, дублюванням записів і різною актуальністю інформації. Існуючі підходи до інтеграції даних часто потребують значного ручного налаштування та не забезпечують достатньої гнучкості при зміні джерел. Це зумовлює необхідність створення методів автоматизованого управління гетерогенними даними, здатних забезпечити їх узгоджене та ефективне використання.

**Аналіз останніх досліджень і публікацій.** Проблема інтеграції гетерогенних даних є однією з ключових у сучасних інформаційних системах, де обсяги інформації постійно зростають. За оцінками сучасних досліджень, значна частина корпоративних даних має неструктурований або напівструктурований характер, що ускладнює їх обробку та інтеграцію. У таких умовах класичні підходи, зокрема ETL-процеси (Extract–Transform–

Load), залишаються поширеним рішенням для централізованого управління даними, проте вони вимагають значних витрат на налаштування та підтримку, особливо при зміні структури джерел [1].

З метою підвищення масштабованості та гнучкості були запропоновані концепції Data Lake, Data Fabric та Data Mesh. Data Lake дозволяє зберігати великі обсяги різномірних даних у вихідному вигляді, що спрощує їх накопичення, однак часто призводить до зниження якості та ускладнення подальшого аналізу. Data Fabric забезпечує єдиний рівень доступу до розподілених джерел, тоді як Data Mesh переносить відповідальність за дані на рівень окремих доменів, що покращує масштабованість, але ускладнює координацію. Додатково застосовуються підходи на основі MDM, графів знань та федеративних запитів, які дозволяють об'єднувати дані без їх фізичного переміщення [2].

Останнім часом також з'являються підходи, орієнтовані на використання великих мовних моделей у задачах інтеграції даних (LLM-driven Data Engineering) [3]. Вони застосовуються для автоматизації семантичного зіставлення атрибутів, генерації правил трансформації та обробки неструктурованих даних. Разом із тим такі підходи знаходяться на стадії активного розвитку та мають обмеження, пов'язані з контрольованістю результатів, інтерпретованістю та вимогами до обчислювальних ресурсів.

Важливим аспектом є забезпечення якості інтегрованих даних. У практичних системах до 20–30 % даних можуть містити дублікати або суперечливу інформацію, що вимагає застосування методів дедуплікації та узгодження. При цьому витрати на очищення та підготовку даних можуть становити значну частину загального часу обробки, що підтверджує необхідність автоматизації цих процесів.

Виділення недосліджених частин загальної проблеми. Незважаючи на різноманіття існуючих підходів, більшість із них не забезпечують повної автоматизації інтеграції та управління гетерогенними даними. Вони або орієнтовані на окремі етапи обробки, або потребують значного ручного налаштування, що обмежує їхню ефективність у динамічних середовищах. Це визначає актуальність розробки методів, які поєднують формалізований підхід із можливістю автоматизованої обробки та забезпечення якості даних [4].

**Мета дослідження.** Метою дослідження є розробка методу автоматизованої інтеграції гетерогенних баз даних на основі формальної композиції послідовних трансформаційних відображень, що забезпечує синтаксичне узгодження форматів, нормалізацію одиниць вимірювання, семантичне зіставлення атрибутів та автоматизовану дедуплікацію записів. Застосування розробленого методу дає змогу автоматизувати усунення дублювання та конфліктів записів, знижуючи частоту ручних втручань на 10–12 %, підвищуючи комплексний показник якості інтегрованих даних на 3–5 % та скорочуючи час оновлення даних на 5–8 % порівняно з традиційними підходами.

**Виклад основного матеріалу.** У роботі розглядається задача автоматизованого управління даними в гетерогенних базах, яка полягає в інтеграції інформації з різномірних джерел у єдину узгоджену модель. У сучасних інформаційних системах дані формуються з багатьох незалежних джерел, що відрізняються за структурою, форматом і семантикою. Це призводить до виникнення дублювань, конфліктів значень та втрати узгодженості, що ускладнює їх подальше використання. Для усунення цих проблем запропоновано метод, який формалізує процес інтеграції та дозволяє розглядати його як керований і відтворюваний процес. Нехай множина джерел даних визначається як  $S = \{s_1, s_2, \dots, s_n\}$ , де кожне джерело має власну структуру представлення  $F_i$ . Процес інтеграції полягає в побудові узагальненої моделі  $U$ , яка виступає єдиним логічним простором даних. Для кожного джерела інтеграція описується як відображення

$$\tau_i : F_i \rightarrow U, \quad (1)$$

що реалізується через композицію перетворень. Такий підхід дозволяє забезпечити незалежність інтеграції від конкретних форматів джерел і спрощує масштабування системи при їх розширенні.

**Загальна схема процесу інтеграції.** Інтеграція реалізується як послідовний процес обробки, у якому кожен етап спрямований на усунення певного типу невідповідностей між даними (рис. 1). Послідовність перетворень формує конвєр, у межах якого дані по-ступово переходять від початкового стану до узгодженої моделі.

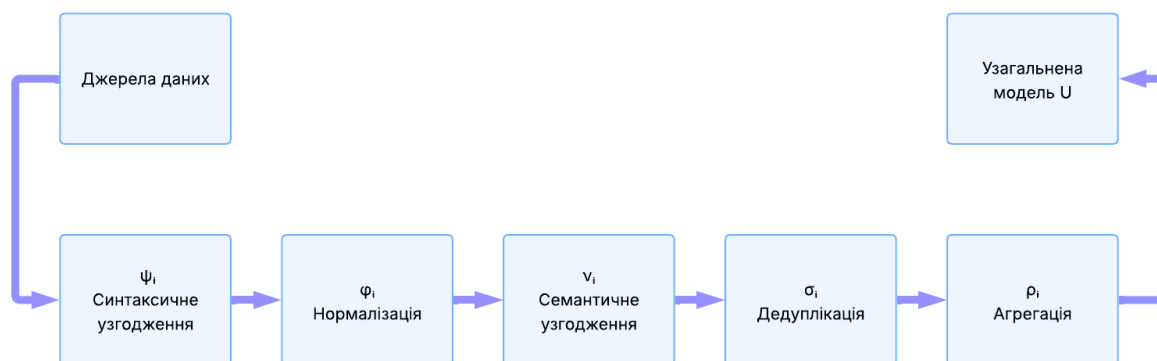


Рис. 1. Узагальнена схема процесу інтеграції гетерогенних даних

Джерело: розроблено авторами.

Формально інтеграцію для кожного джерела можна подати у вигляді композиції функцій

$$U_i = \rho_i \circ \sigma_i \circ \nu_i \circ \phi_i \circ \psi_i, \quad (2)$$

де  $\rho_i$  — функція агрегації та формування узагальненої моделі даних;

$\sigma_i$  — функція нормалізації значень;

$\nu_i$  — функція семантичного зіставлення атрибутів;

$\phi_i$  — функція синтаксичного узгодження структур;

$\psi_i$  — функція попереднього перетворення даних джерела;

$i$  — індекс джерела даних,  $i = 1, \dots, n$ .

Така формалізація дозволяє описати процес інтеграції як набір незалежних, але узгоджених операцій, що можуть бути реалізовані у вигляді окремих модулів системи.

**Синтаксичне узгодження даних.** Початковий етап інтеграції пов'язаний з усуненням структурних відмінностей між джерелами. Дані, що надходять у різних форматах, приводяться до уніфікованого внутрішнього представлення за допомогою правил трансформації та парсерів. Це дозволяє перетворити ієрархічні або вкладені структури у форму, придатну для подальшої обробки, зберігаючи при цьому їх змістову цілісність. Важливим є те, що правила трансформації задаються незалежно від конкретного джерела, що дозволяє повторно використовувати їх при підключенні нових даних [5].

**Нормалізація значень.** Після усунення структурних відмінностей виконується нормалізація значень, яка забезпечує узгодженість даних на рівні їх змісту. У процесі нормалізації значення приводяться до єдиних форматів та одиниць вимірювання, що дозволяє уникнути логічних помилок при їх подальшій обробці. Наприклад, числові значення можуть бути приведені до однакових одиниць вимірювання, а часові дані — до єдиного формату. Такий підхід дозволяє підвищити точність результатів інтеграції та створює основу для коректного порівняння записів [6].

**Семантичне узгодження даних.** Семантичне узгодження є одним із ключових етапів інтеграції, оскільки саме на цьому рівні встановлюється відповідність між атрибутами різних джерел (рис. 2). На відміну від синтаксичного узгодження, цей етап враховує зміст даних і дозволяє об'єднувати інформацію навіть за відсутності повної структурної відповідності. Реалізація цього етапу базується на використанні правил відповідності, які визначають, які атрибути різних джерел відносяться до однієї сутності [7].

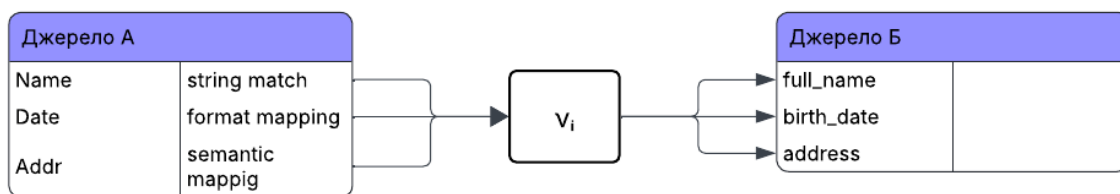


Рис. 2. Приклад семантичного зіставлення атрибутів

Джерело: розроблено авторами.

Такий підхід дозволяє створити єдину логічну модель даних, у якій різні представлення одних і тих самих об'єктів інтегруються в узгоджену структуру.

Дедуплікація записів. У процесі інтеграції значна частина даних може містити дублікати або записи, що частково збігаються. Для їх виявлення використовується функція подібності  $\delta(di, dj)$ , яка оцінює ступінь відповідності між записами. Значення цієї функції визначається на основі порівняння атрибутів записів, що дозволяє враховувати як точні, так і часткові збіги [8].

Якщо значення подібності перевищує заданий поріг  $\epsilon$ , записи вважаються такими, що описують одну сутність, і підлягають об'єднанню. Такий підхід дозволяє зменшити надлишковість даних і підвищити їхню достовірність, що є критично важливим для подальшого аналізу.

Агрегація та формування узагальненої моделі. Завершальним етапом інтеграції є агрегація, у результаті якої формується узагальнена модель  $U$ . У цій моделі об'єднуються дані з різних джерел, усуваються конфлікти значень і встановлюються зв'язки між сутностями. Важливим аспектом є те, що при агрегації враховується актуальність даних, що дозволяє обрати найбільш релевантні значення у випадку конфліктів [9].

Контроль якості даних. Якість інтегрованих даних оцінюється за сукупністю показників (рис. 3), що відображають їх повноту, точність, актуальність та унікальність. Значення цих показників аналізуються після виконання кожного етапу інтеграції, що дозволяє визначити вплив окремих перетворень на кінцевий результат. Такий підхід забезпечує можливість адаптації процесу інтеграції шляхом зміни правил обробки та підвищує надійність системи [10].

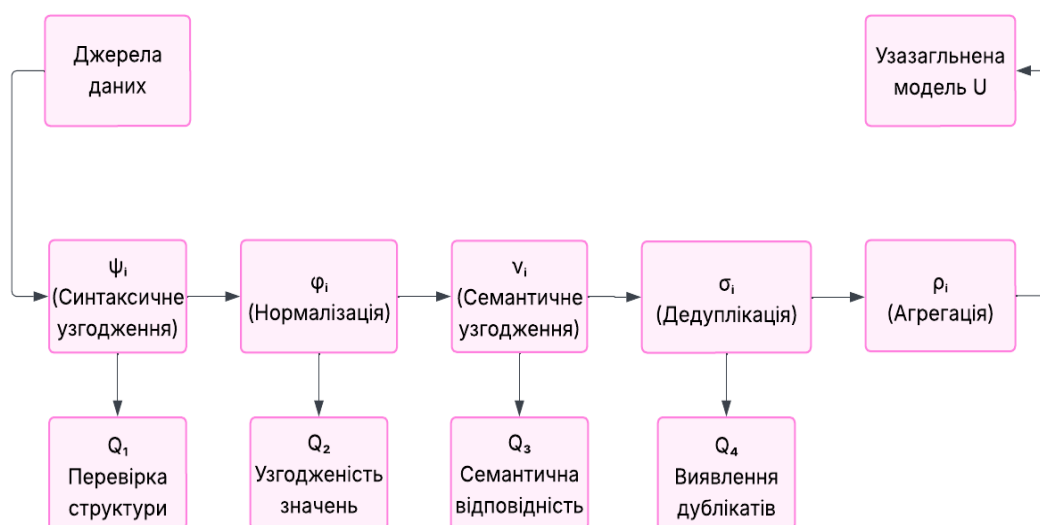


Рис. 3. Схема контролю якості даних на етапах інтеграції

Джерело: розроблено авторами.

Архітектура програмної системи. Запропонований метод реалізується у вигляді програмної системи з модульною архітектурою, яка забезпечує поетапну обробку гетерогенних даних та їх інтеграцію у єдину узагальнену модель. Архітектура побудована за принципом розподілу відповідальності між окремими рівнями, що дозволяє ізолювати функціональні компоненти системи та забезпечити їх незалежне масштабування.

На верхньому рівні архітектури розташовано рівень прийому даних, який забезпечує взаємодію з зовнішніми джерелами. Джерела можуть представляти дані в різних форматах, включаючи структуровані (реляційні таблиці), напівструктуровані (JSON, XML) та неструктуровані дані. На цьому етапі виконується прийом даних через API або інші інтерфейси, їх первинна валідація та перевірка на відповідність базовим вимогам коректності. Також здійснюється попереднє розділення потоків даних залежно від їх типу та джерела, що дозволяє оптимізувати подальшу обробку.

Середній рівень архітектури відповідає за основну обробку даних та реалізує запропонований метод інтеграції. Він організований у вигляді конвеєра обробки, в якому кожен модуль виконує окремий етап трансформації. Передача даних між модулями здійснюється у стандартизованому внутрішньому форматі, що дозволяє уникнути повторних перетворень. На цьому рівні реалізуються процедури синтаксичного узгодження, нормалізації значень, семантичного зіставлення та дедуплікації. Важливою особливістю є можливість незалежного налаштування кожного модуля, що дозволяє адаптувати систему до специфіки конкретних джерел даних. Крім того, така організація забезпечує можливість паралельної обробки даних, що підвищує продуктивність системи при великих обсягах інформації.

Нижній рівень архітектури представлений підсистемою зберігання даних, яка відповідає за формування та підтримку узагальненої моделі U. На цьому рівні реалізується ефективно зберігання інтегрованих даних, організація індексів та підтримка механізмів доступу до інформації. Структура зберігання орієнтована на забезпечення швидкого виконання запитів та можливість оновлення даних без порушення їх узгодженості. Важливим аспектом є підтримка актуальності даних, що досягається за рахунок періодичного оновлення та обробки нових записів.

Взаємодія між рівнями організована таким чином, що кожен наступний рівень отримує вже підготовлені дані, що зменшує навантаження на систему та підвищує ефективність обробки. Така архітектура дозволяє масштабувати окремі компоненти незалежно, наприклад, збільшувати ресурси рівня обробки при зростанні обсягів даних або розширювати підсистему зберігання при збільшенні кількості інтегрованої інформації.

Крім того, модульна побудова системи дозволяє інтегрувати додаткові компоненти, зокрема механізми моніторингу, логування та контролю якості даних. Це забезпечує можливість відстеження стану системи та оперативного реагування на зміни у потоках даних. У результаті запропонована архітектура забезпечує гнучкість, масштабованість та надійність функціонування системи в умовах роботи з гетерогенними даними.

Для оцінювання ефективності запропонованого методу було розглянуто два умовні сценарії інтеграції гетерогенних даних, що поєднують реляційні таблиці та напівструктуровані JSON-документи. У межах оцінювання аналізувалися обсяг записів, рівень дублювання до і після інтеграції, а також час обробки даних. Оціночні результати наведено у таблиці 1.

Таблиця 1 – Результат експериментальної перевірки методу.

Набір даних	Тип джерел	Кількість записів	Рівень дублюв. до, %	Рівень дублюв. після, %	Час інтеграції до, хв	Час інтеграції після, хв
DS1 (тестовий)	SQL + JSON	20 000	14,2	4,1	12,5	11,3
DS2 (CRM)	SQL + JSON/API	65 000	17,8	5,6	31,0	28,4

Джерело: розроблено авторами.

Оціночні результати, наведені в таблиці 1, свідчать про зменшення рівня дублювання записів та скорочення часу інтеграції при застосуванні запропонованого методу. Найбільший ефект досягається за рахунок поєднання етапів семантичного зіставлення, дедуплікації та поетапної обробки даних.

**Висновки.** У роботі розглянуто задачу інтеграції гетерогенних даних та запропоновано формалізований підхід до її вирішення, який базується на поетапній обробці інформації з використанням композиції функцій перетворення. Запропонований метод дозволяє представити процес інтеграції у вигляді послідовності узгоджених операцій, що забезпечує його відтворюваність та масштабованість.

У ході дослідження було визначено ключові етапи інтеграції, зокрема синтаксичне узгодження, нормалізацію, семантичне зіставлення, дедуплікацію та агрегацію, кожен з яких спрямований на усунення окремого класу невідповідностей у даних. Запропонована структура процесу дозволяє систематизувати обробку даних та забезпечити контроль їхньої якості на кожному етапі.

Особливістю роботи є введення механізму контролю якості даних, який дозволяє оцінювати коректність результатів інтеграції після виконання кожного етапу обробки. Це забезпечує підвищення достовірності узагальненої моделі та зменшення кількості помилок, пов'язаних із дублюванням та семантичними невідповідностями.

Запропонований підхід дозволяє підвищити ефективність інтеграції даних за рахунок формалізації процесу, зменшення надлишковості інформації та забезпечення узгодженості даних. Отримані результати можуть бути використані при проектуванні інформаційних систем, що працюють з великими обсягами гетерогенних даних.

Перспективи подальших досліджень пов'язані з удосконаленням методів семантичного зіставлення даних, зокрема із застосуванням інтелектуальних підходів до автоматичного формування правил відповідності, а також з розширенням механізмів оцінювання якості інтегрованих даних.

#### Список використаних джерел

1. Dong, X. L., & Srivastava, D. (2015). *Big data integration*. Springer. <https://doi.org/10.1007/978-3-031-01853-4>.
2. Blohm, I., Wortmann, F., Legner, C., & Köbler, F. (2024). Data products, data mesh, and data fabric. *Business & Information Systems Engineering*. <https://doi.org/10.1007/s12599-024-00876-5>
3. Remadi, A., El Hage, K., Hobeika, Y., & Bugiotti, F. (2024). To prompt or not to prompt: Navigating the use of large language models for integrating and modeling heterogeneous data. *Data & Knowledge Engineering*, 102313. <https://doi.org/10.1016/j.datak.2024.102313>.
4. Buono, F. D., Faggioli, G., Paganelli, M., Baraldi, A., Guerra, F., & Ferro, N. (2022). A framework to evaluate the quality of integrated datasets. *ACM SIGAPP Applied Computing Review*, 22(4), 5–23. <https://doi.org/10.1145/3584014.3584015>.
5. Rahm, E., & Bernstein, P. A. (2001). A survey of approaches to automatic schema matching. *The VLDB Journal*, 10(4), 334–350. <https://doi.org/10.1007/s007780100057>.
6. Kimball, R., & Ross, M. (2013). *The data warehouse toolkit: The definitive guide to dimensional modeling*. John Wiley & Sons.
7. Paganelli, M., Buono, F. D., Guerra, F., & Ferro, N. (2020). Unsupervised evaluation of data integration processes. У *IiWAS '20: The 22nd international conference on information integration and web-based applications & services*. ACM. <https://doi.org/10.1145/3428757.3429129>.
8. Barlaug, N., & Gulla, J. A. (2021). Neural networks for entity matching: A survey. *ACM Transactions on Knowledge Discovery From Data*, 15(3), 1–37. <https://doi.org/10.1145/3442200>.
9. Masmoudi, M., Lamine, S. B. A. B., Karray, M. H., Archimede, B., & Zghal, H. B. (2024). Semantic data integration and querying: A survey and challenges. *ACM Computing Surveys*. <https://doi.org/10.1145/3653317>.
10. Batini, C., & Scannapieco, M. (2016). *Data and information quality*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-24106-7>.

### References

1. Dong, X. L., & Srivastava, D. (2015). *Big data integration*. Springer. <https://doi.org/10.1007/978-3-031-01853-4>.
2. Blohm, I., Wortmann, F., Legner, C., & Köbler, F. (2024). Data products, data mesh, and data fabric. *Business & Information Systems Engineering*. <https://doi.org/10.1007/s12599-024-00876-5>
3. Remadi, A., El Hage, K., Hobeika, Y., & Bugiotti, F. (2024). To prompt or not to prompt: Navigating the use of large language models for integrating and modeling heterogeneous data. *Data & Knowledge Engineering*, 102313. <https://doi.org/10.1016/j.datak.2024.102313>.
4. Buono, F. D., Faggioli, G., Paganelli, M., Baraldi, A., Guerra, F., & Ferro, N. (2022). A framework to evaluate the quality of integrated datasets. *ACM SIGAPP Applied Computing Review*, 22(4), 5–23. <https://doi.org/10.1145/3584014.3584015>.
5. Rahm, E., & Bernstein, P. A. (2001). A survey of approaches to automatic schema matching. *The VLDB Journal*, 10(4), 334–350. <https://doi.org/10.1007/s007780100057>.
6. Kimball, R., & Ross, M. (2013). *The data warehouse toolkit: The definitive guide to dimensional modeling*. John Wiley & Sons.
7. Paganelli, M., Buono, F. D., Guerra, F., & Ferro, N. (2020). Unsupervised evaluation of data integration processes. *Y IiWAS '20: The 22nd international conference on information integration and web-based applications & services*. ACM. <https://doi.org/10.1145/3428757.3429129>.
8. Barlaug, N., & Gulla, J. A. (2021). Neural networks for entity matching: A survey. *ACM Transactions on Knowledge Discovery From Data*, 15(3), 1–37. <https://doi.org/10.1145/3442200>.
9. Masmoudi, M., Lamine, S. B. A. B., Karray, M. H., Archimede, B., & Zghal, H. B. (2024). Semantic data integration and querying: A survey and challenges. *ACM Computing Surveys*. <https://doi.org/10.1145/3653317>.
10. Batini, C., & Scannapieco, M. (2016). *Data and information quality*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-24106-7>.

Дата першого надходження статті до видання: 11.03.2026

Дата прийняття статті до друку після рецензування: 30.03.2026

UDC 004.65:004.678

*Hlib Shevchenko<sup>1</sup>, Inna Saiapina<sup>2</sup>*

<sup>1</sup>Master's student of the Computer Systems Software Department  
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute" (Kyiv, Ukraine)  
E-mail: [shevchenko.h.o.-kp41mn@edu.kpi.ua](mailto:shevchenko.h.o.-kp41mn@edu.kpi.ua). ORCID: <https://orcid.org/0009-0005-5733-5826>

<sup>2</sup>PhD in Technical Sciences, Associate Professor, Associate Professor of the Computer Systems Software Department,  
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute" (Kyiv, Ukraine)  
E-mail: [saiapina@pzks.fpm.kpi.ua](mailto:saiapina@pzks.fpm.kpi.ua). ORCID: <https://orcid.org/0000-0003-1541-1681>  
Researcher ID: [P-6639-2018](https://orcid.org/0000-0003-1541-1681). SCOPUS Author ID: [57221620283](https://orcid.org/0000-0003-1541-1681)

### METHOD AND SOFTWARE FOR AUTOMATING DATA MANAGEMENT PROCESSES IN HETEROGENEOUS DATABASES

*The growth in the volume and diversity of data formats in modern information systems leads to the need for integrating heterogeneous data sources represented by relational, document-oriented, and other storage models. Under such conditions, ensuring data consistency, quality, and timeliness becomes increasingly complex, which necessitates the automation of data management processes.*

*This paper addresses the problem of developing a method for automated management of heterogeneous data based on a formalized integration model. The proposed method represents the integration process as a composition of transformations applied to a set of data sources, including syntactic harmonization of structures, normalization of values, semantic matching of attributes based on correspondence rules, and record deduplication using similarity functions.*

*The implementation of the method is based on a modular software architecture, in which data processing stages are executed sequentially and independently of the types of data sources. A relational database with support for semi-structured data formats is used for storing integrated data, ensuring model flexibility and system scalability. The approach also defines criteria for data quality evaluation and mechanisms for maintaining data relevance in dynamic environments.*

*The proposed method improves data consistency and reliability, reduces processing time, and enables efficient operation of systems in distributed environments.*

**Keywords:** heterogeneous databases; data integration; automated data management; semantic matching; deduplication; data quality; formalized model; distributed systems.

Fig.: 3. Table: 1. References: 10.