

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ, ОБЧИСЛЮВАЛЬНА ТЕХНІКА І АВТОМАТИКА

УДК 004.75

М. В. Двоєглазова, к.т.н., доцент

Чернігівський національний технологічний університет,
вул. Шевченка, 95, м. Чернігів, Україна
marinitta_1905@mail.ru

АНАЛІЗ МЕТОДІВ ТА ТЕХНОЛОГІЙ ІНТЕГРАЦІЇ ІНФОРМАЦІЙНИХ СИСТЕМ

Проведений аналіз існуючих методів інтеграції дає уявлення про сучасні тенденції та технології інтеграції, систематизує існуючі знання про інтеграційні процеси інформаційних систем. Інформаційні ресурси інформаційних систем передбачають застосування значної кількості різнорідних даних. Проблемою інтеграції даних є створення та використання методик представлення даних. Встановлено, що процеси інтеграції даних мають широку сферу застосування.

Ключові слова: інтеграція, інформаційні системи, інтероперабельність, методи, технології.

Постановка проблеми. Проблема інтеграції інформаційних систем зумовлена рядом факторів, пов'язаних зі зростанням обсягів даних, які застосовують автоматизовані інформаційні системи. Необхідність інтеграції різнорідних інформаційних систем зумовлює створення множини методів та технологій інтеграції, які здійснюються за допомогою різних процедур з використанням різних моделей даних.

Аналіз останніх досліджень та публікацій. Процес інтеграції інформаційних систем (ІС) полягає в формуванні глобального інформаційного ресурсу з множини локальних, таким чином щоб результат його застосування перевищував суму локальних. Актуальність цього питання пов'язана з постійним збільшенням обсягів інформації, різноманітністю її форматів та способів опрацювання, з розвитком інформаційних ресурсів і сервісів, які призначені для колективного використання, зі зростанням частки слабко структурованих даних та даних без детермінованої структури в інформаційних ресурсах [1–5].

Проте наявні знання про існуючі методи інтеграції інформаційних є несистематизованими. Систематизація знань дозволить використовувати комбінації цих методів в моделюванні процесів інтеграції інформаційних систем.

Мета статті. Головною метою дослідження є проведення аналізу існуючих мето-

дів інтеграції та інтероперабельності інформаційних систем.

Виклад основного матеріалу. На сьогоднішній день на ринку інформаційних технологій представлено багато інструментаріїв вирішення проблеми інтеграції на різних рівнях, таких як інтеграція бізнес-процесів, інтеграція корпоративних додатків, інтеграція корпоративних платформ та інтеграція даних.

Інтеграція бізнес-процесів базується на визначенні, реалізації та управлінні процесами обміну інформацією між різними системами. Застосування цього підходу дозволяє удосконалювати операції та оптимізувати витрати в процесі використання інформаційних систем.

Інтеграція додатків здійснюється за рахунок об'єднання даних або функцій однієї системи з іншою. Передача функцій або даних, властивих будь-якому додатку, у розпорядження іншого додатку використовується з метою, щоб їх взаємодія на стадії реалізації забезпечила виконання певної прикладної функції ІС.

Інтеграція на основі стандартів ґрунтується на використанні стандартних форматів даних. Для забезпечення інтеграції даних необхідно вибрати стандартні формати для даних. Стандартами для інтеграції вважають ті формати, які можуть підтримувати використання та поширення інформації та бізнес-даних.

Інтеграція платформ стосується процесів та інструментів, за допомогою яких може бути здійснений безпечний та оптимальний обмін інформацією. В результаті дані можуть безперешкодно передаватись по різних додатках [6].

Інтеграція даних являє собою процес об'єднання даних, отриманих з різних джерел. Процеси інтеграції даних передбачають розв'язання таких проблем, як інтеграція синтаксису, інтеграція структури та інтеграція структури семантики даних [2–5].

Семантика є властивістю даних, що забезпечує їх змістовність та можливість застосування даних за призначенням. Семантику визначає множина відповідностей між формальними позначеннями та реальними поняття-

ми предметної області, що дає можливість інтерпретувати дані на різних стадіях роботи з ними. В сучасних інформаційних технологіях застосовують багато різноманітних засобів визначення семантики даних.

Інтеграція семантики даних передбачає формування єдиного змістового простору для сприйняття, інтерпретації та застосування даних незалежно від формату їх подання та структури. Однією з основних проблем у цьому процесі є визначення критеріїв даних, за допомогою яких можна визначити можливість чи неможливість їх об'єднання [2–5].

Найвідоміші методи семантичної інтеграції представлені на рис. 1.

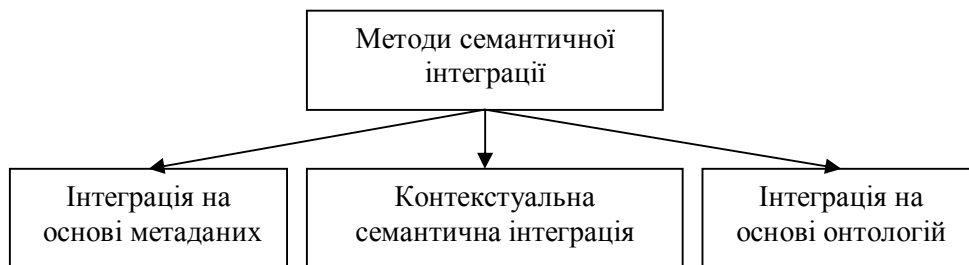


Рис. 1. Методи семантичної інтеграції

При інтеграції інформаційних систем (ІС) важливе місце посідає інтероперабельність ІС, здатність інформаційної системи взаємодіяти з іншими інформаційними системами. Така взаємодія може бути виражена у вигляді обміну даними, розподіленого виконання пошукових запитів, узгодженої зміни баз даних тощо.

Необхідність забезпечення інтероперабельності виникає при взаємодії бізнес-процесів підприємств-партнерів, узгодженості роботи існуючої ІС зі стандартними рішеннями [7].

Також властивість інтероперабельності використовується при інтеграції декількох ІС, включенні в створювану базу даних (БД) ра-

ніше використовуваних сховищ даних, розробці комплексних автоматизованих систем управління, побудові мереж інформаційних сховищ. Проблема забезпечення є дуже актуальною.

Розрізняють два аспекти інтероперабельності: структурний та семантичний. Структурний аспект інтероперабельності систем означає здатність до структурної узгодженості існуючих систем. Семантичний аспект означає можливість встановлення відповідності між змістами інформаційних систем.

Модель компонентів інтероперабельності, яку можна застосовувати в обох випадках, представлена на рис. 2 [6].



Рис. 2. Модель компонентів інтероперабельності

Існуючі методи забезпечення інтероперабельності орієнтовані на синтаксичні (структурні) аспекти, тобто направлені на узгодження та перетворення структур даних за рахунок стандартизації їх форматів та використання спеціалізованих мов (метамов) [7].

Метод інтеграції на основі метаданих передбачає порівняння складу та змісту метаданих двох наборів з метою визначення можливості їх семантичної інтеграції. Метадані забезпечують формування та застосування опису основних властивостей деякого набору даних (інформаційного ресурсу), зокрема, таких, що визначають його семантичні показники.

Найпоширенішою структурою метаданих є вимірна схема Захмана, яка передбачає застосування шести категорій метаданих. Ці категорії описують такі властивості інформаційного ресурсу (ІР):

– *об'єкти даних* – опис сутностей, які асоціюють з значенням набору даних;

– *суб'єкти даних* – опис осіб, які створюють чи застосовують дані;

– *часові показники* – опис часових моментів чи інтервалів, що характеризують створення, підтримання та застосування даних;

– *розміщення даних* – опис місцезнаходження даних та способів та порядку доступу до ресурсів;

– *призначення даних* – опис функцій та завдань, які застосовують інформаційний ресурс;

– *порядок застосування даних* – правила та обмеження на роботу з інформаційним ресурсом.

Загальну структуру метаданих інформаційного ресурсу побудованих за схемою Захмана представлено на рис. 3. Кожен з вимірів метаданих – це деяка множина значень, що характеризують один з аспектів організації, сприйняття та застосування деякого набору даних.

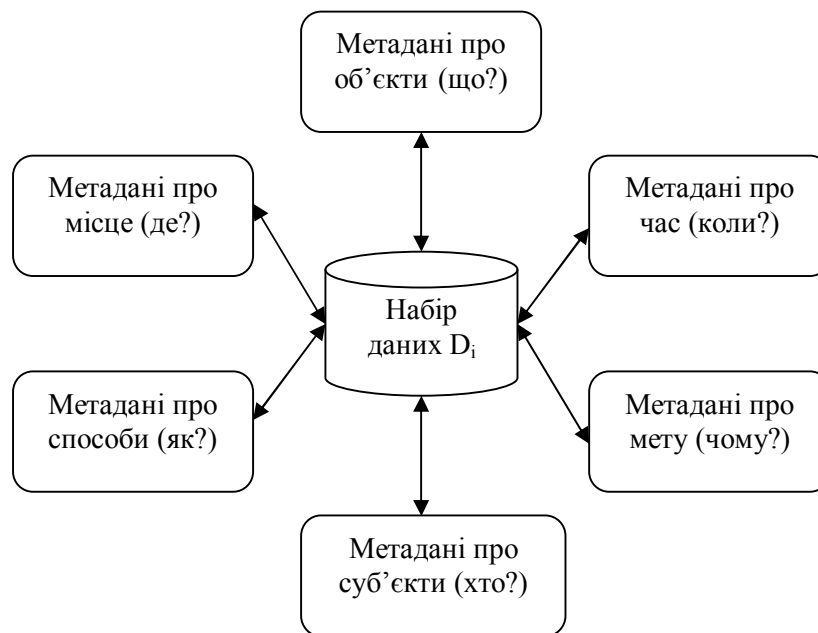


Рис. 3. Схема Захмана організації метаданих

Множину метаданих M_i для деякого набору даних D_i можна представити у вигляді:

$$M_i = \langle M_i^o, M_i^s, M_i^p, M_i^t, M_i^g, M_i^n \rangle, \quad (1)$$

де M_i^o – метадані про об'єкти даних; M_i^s – метадані про суб'єкти даних; M_i^p – метадані про розміщення даних; M_i^t – метадані про часові показники даних; M_i^g – метадані

про мету застосування даних; M_i^n – метадані про порядок використання даних.

Збіг значень метаданих двох наборів за одним або більше вимірами застосовують як критерій семантичної інтеграції. Міра збігу значень метаданих певної категорії залежить від конкретних завдань, але її числовий вираз не повинен бути меншим 85%. Альтернативним до схеми Захмана способом до організації метаданих є Дублінське ядро [2–5].

Основними принципами побудови системи метаданих на основі дублінського ядра є простота, зрозуміла семантика, інтернаціоналізація, здатність розширення, однозначність, спрощення зображення метаданих, коректність значень, зв'язок з синтаксисом, використання стандартних просторів імен.

Застосування такого переліку властивостей інформаційного ресурсу дозволяє формалізувати та уніфікувати опис даних, які підлягають інтеграції. Метадані, побудовані за принципом Дублінського ядра, мають наступні властивості:

- дозволяють побудувати опис інформаційного ресурсу будь-якого виду;
- забезпечують повний та всебічний опис всіх властивостей даних, особливо тих, що характеризують їх семантику;
- є можливість зображення метаданих у форматі XML, що спрощує процеси їх формального опрацювання.

Формально множину метаданих деякого інформаційного ресурсу D , побудовану за принципом Дублінського ядра, можна представити:

$$M^{DC}(D) = \{M_1^{DC}(D), M_2^{DC}(D), \dots, M_{15}^{DC}(D)\}, \quad (2)$$

де $M^{DC}(D)$ – множина метаданих; $M_i^{DC}(D), i = 1, 2, \dots, 15$ значення відповідного елемента базового набору Дублінського ядра.

Критерій семантичної інтеграції даних у такому випадку формулюють за допомогою функцій збігу елементів опису двох інформаційних ресурсів D_1 та D_2 , які визначено важливими для їх поєднання:

$$\text{Map}(M_i^{DC}(D_1), (M_i^{DC}(D_2))) = \text{true}, \quad (3)$$

де $i \in (1 - 15)$. Залежно від значень $M_i^{DC}(D_1)$ та $M_i^{DC}(D_2)$ функція збігу може бути істинною або хибною.

Алгоритм інтеграції ІС на основі метамodelей має містити наступні етапи:

- аналіз існуючих баз даних, їх атрибутів та зв'язків між ними;
- аналіз семантичного змісту сутностей та атрибутів;
- уточнення семантичних відповідностей;

- побудова єдиної метамodelі;
- представлення результируючих відображень між сутностями та атрибутами інформаційних систем.

На рис. 4 представлено схему інтеграції двох ІС на основі метамodelей. Метою інтеграції є забезпечення взаємодії між двома ІС. Для цього необхідно визначити відповідність сутностей ІС підприємства сутностям ІС проекту та правила їх перетворення [8].

Спочатку з інформаційних систем вилучають схеми даних. За допомогою аналізу схем даних можна визначити лише структурну інтероперабельність. Для забезпечення семантичної інтероперабельності необхідно розуміння елементів ІС. Тому необхідно використовувати другу складову частину метаданих – концептуальну модель предметної області (ПО).

На базі концептуальних моделей ІС підприємства та ІС проекту здійснюється побудова метамodelі предметної області інтегрованих ІС.

Ця метамodelь об'єднує та узгоджує в собі обидві концептуальні метамodelі. Метамodelь визначає відповідності сутностей однієї ІС сутностям іншої ІС та правила їх перетворення [8].

ІС включаються в єдиний інформаційний простір, що задається онтологією і отримують можливість взаємодіяти з іншими ІС з даної предметної області.

Метод контекстуальної інтеграції ґрунтується на змістовому порівнянні інформаційного наповнення наборів даних, що підлягають інтеграції. Цей метод дає змогу оцінити можливості інтегрування як структурованих (реляційних) даних, так і слабкоструктурованих – поданих у довільних форматах. Істотним аспектом визначення семантики слабкоструктурованих є їх контекст. Контекст може мати різні форми, такі як текст та гіперпосилання на Web-сторінці, ім'я каталогу, в якому зберігають дані, супутні анотації та коментарі до даних, перелік ключових слів тощо. В більшості випадків семантичний аналіз повного вмісту ІР є складним, тому його замінюють контекстуальним аналізом тезаурус них термінів наборів даних [2; 5].

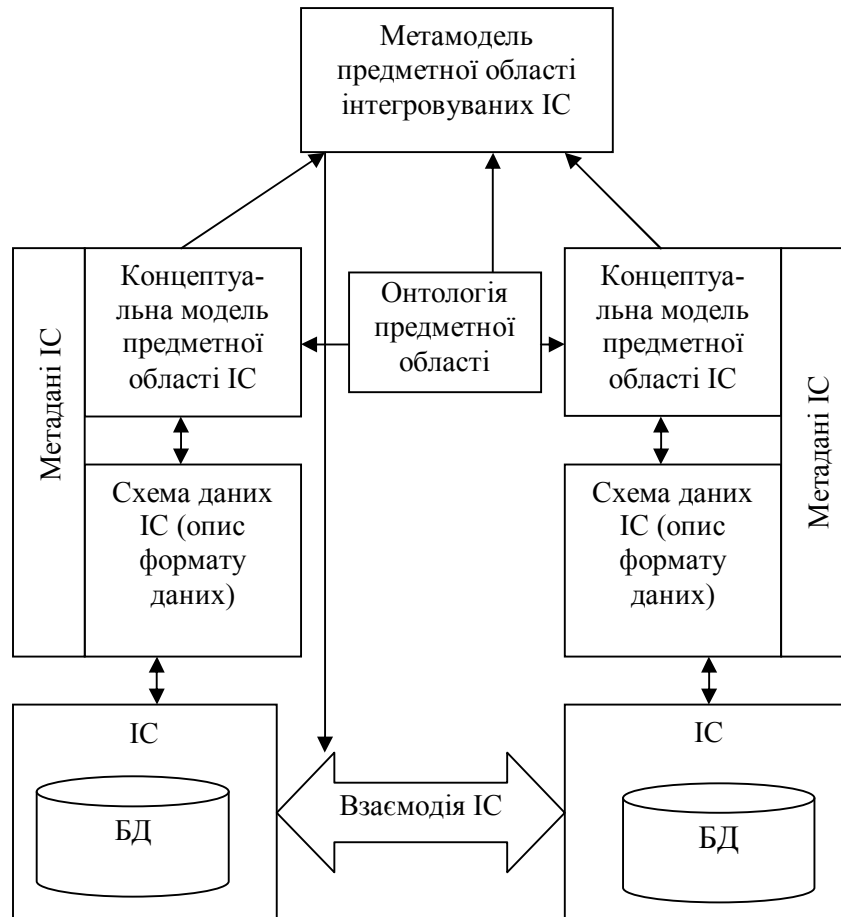


Рис. 4. Схема інтеграції двох ІС на основі метамоделей

Критерієм семантичної інтегрованості двох наборів даних в цьому випадку є функція контекстуальної семантичної віддалі між ними (CSD-функція). Значення функції контекстуальної семантичної віддалі між двома наборами даних визначають за формулою:

$$CSD(D_i, D_j) = \frac{|\Omega_{ij}|}{\text{Min}(|\Omega_i|, |\Omega_j|)}, \quad (4)$$

де Ω_i , Ω_j – множини тезаурус них термінів наборів даних D_i та D_j відповідно; Ω_{ij} – множина тезаурус них термінів, які є семантично спільними для двох наборів; $|\Omega_i|$, $|\Omega_j|$, $|\Omega_{ij}|$ – потужності відповідних множин.

Семантична інтеграція двох наборів даних є можливою, якщо значення функції контекстуальної семантичної віддалі за умови: $CSD(D_i, D_j) \geq 0,8$.

На рис. 5 представлено загальну схему процесу семантичної інтеграції двох наборів

різномірних даних з застосуванням тезаурус них термінів та функцій семантичної віддалі.

Метод контекстуального аналізу дозволяє перевірити критерії семантичної інтеграції на формальному рівні та не потребує участі експерта. Істотним недоліком цього методу є формування тезаурусу та набору даних, що містить певний інформаційний ресурс. Неоднорідність форматів та структур даних, які підлягають інтеграції, призводить до трудомісткості та малої ефективності створення наборів ключових термінів. Також, формування тезаурусу залежить від людського фактору, що в свою чергу впливає на універсальність методу.

Інтеграція на основі онтологій передбачає використання основних елементів двох попередніх методів – тезаурусу та метаданих, але є значно загальнішим за них та враховує більше аспектів семантики даних [2–5].

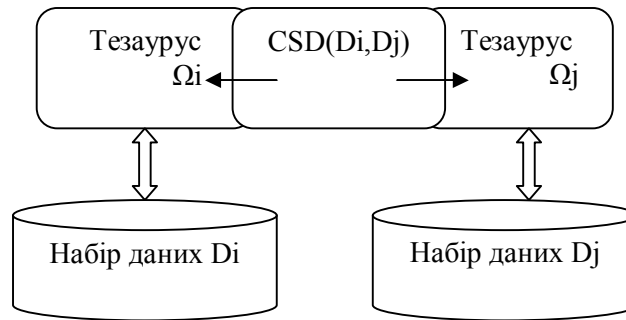


Рис. 5. Загальна схема контекстуальної семантичної інтеграції

Онтологію розглядають як цілісну формалізовану специфіку предметної області (ПО), яка повинна забезпечувати однакову інтерпретацію знань про ПО на людському та комп'ютерному рівні. У випадку інтеграції даних об'єктом опису у вигляді онтології виступає інформаційний ресурс. У загальному випадку формальним зображення онтологій є:

$$O = \langle X, R, F \rangle, \quad (5)$$

де X – скінчена множина понять (класів, концептів) ПО з їх властивостями (атрибутами); R – скінчена множина відношень (зв'язків, відповідностей) між поняттями; F – скінчена множина функцій інтерпретації (обмежень, аксіом).

Процеси семантичної інтеграції даних передбачають створення для кожного вхідного набору даних D_i власної онтології $O(D_i)$, яка формує однозначний опис семантики як всього IP, так і його елементів:

$$O(D_i) = \langle X(D_i), R(D_i), F(D_i) \rangle, \quad (6)$$

де $X(D_i)$ – множина концептів, які описують одиниці даних, їх зміст, властивості та належність до певних класів чи категорій; $R(D_i)$ – множина зв'язків та відношень між одиницями даних, що визначають порядок їх взаємодії та їх взаємного застосування; $F(D_i)$ – множина семантичних обмежень та функцій інтерпретації даних, які пов'язують їх з реальними поняттями та об'єктами ПО, а також регламентують порядок таких відповідностей.

Критерії семантичної інтеграції в цьому випадку визначають як послідовність вимог щодо елементів двох онтологій: два набори даних D_i та D_j вважають придатними до семантичної інтеграції, якщо для двох онтологій

O_i та O_j , що відповідають цим наборам даних, виконуються умови:

1. У множинах концептів $X(D_i)$ та $X(D_j)$:

- немає однакових понять, що описані порізнному;
- немає понять різного змісту, описаних однаково.

2. У множинах зв'язків $R(D_i)$ та $R(D_j)$:

- відсутні зв'язки протилежного напрямку та змісту між однаковими концептами;
- відсутні однотипні зв'язки, що не можуть бути реалізовані одночасно.

3. У множинах функцій інтерпретацій $F(D_i)$ та $F(D_j)$:

- немає функцій, одночасна реакція яких призведе до неоднозначності інтерпретацій;
- з однотипними концептами різних онтологій не пов'язано обмежень, які не можуть бути виконані одночасно.

Розрізняють наступні технології інтеграції:

- системи інтеграції корпоративних додатків
- технології, орієнтовані на розв'язання проблем інтеграції різних систем, додатків та даних всередині окремої організації;
- системи інтеграції між організаціями – технології, орієнтовані на забезпечення безпечного, надійного інформаційного обміну між різними організаціями та їх інформаційними системами;
- технології управління бізнес-процесами, які є результатом еволюції класичних систем документообігу та діловодства [6].

Висновки. Запропонований аналіз дає уявлення про сучасні тенденції та технології інтеграції, систематизує існуючі знання про інтеграційні процеси інформаційних систем.

Встановлено, що процеси інтеграції даних мають широку сферу застосування. Інформаційні ресурси цих систем передбачають застосування значної кількості даних. Тому, найважливішою проблемою інтеграції даних є створення та використання методик представлення різномірних даних. Систематизація знань про існуючі методи інтеграції дозволить використовувати комбінації даних методів в моделюванні процесів інтеграції інформаційних систем.

Список літератури

1. Абрамов В. В. Механизмы интеграции систем [Электронный ресурс] / В. В. Абрамов. – Режим доступа : <http://fetmag.mrsu.ru/2010-2/pdf/Integration.pdf>.
2. Берко А. Ю. Методи та засоби семантичної інтеграції даних / А. Ю. Берко // Вісник Національного університету «Львівська політехніка». – 2008. – № 610. – С. 190–199.
3. Берко А. Ю. Інформаційна система керування автодилерською мережею на основі технологій інтеграції / А. Ю. Берко, І. М. Матрофайло // Вісник Національного університету «Львівська політехніка». – 2008. – № 631. – С. 19–28.
4. Берко А. Ю. Структурно-семантична інтеграція даних на основі фактологічної реляційної моделі / А. Ю. Берко // Вісник Національного університету «Львівська політехніка». – 2010. – № 663. – С. 60–70.
5. Берко А. Ю. Моделі інтеграції даних у відкритих інформаційних системах / А. Ю. Берко // Актуальні проблеми економіки. – 2010. – № 10. – С. 147–152.
6. Франгулова Е. В. Классификация подходов к интеграции и интероперабельности информационных систем / Е. В. Франгулова // Вестник АГТУ. – 2010. – № 2. – С. 176–180. – (Серия : Управление, вычислительная техника и информатика).
7. Зінькович В. М. Концепції та моделі інтеграції об'єктів метатехнології в середовищі METACASE / В. М. Зінькович, Є. І. Моренцов // Проблеми програмування. – 2006. – № 2–3. – С. 635–640.
8. Михайлов И. С. Исследование и разработка методов и программных средств обеспечения структурной и семантической интероперабельности информационных систем на основе метамodelей / И. С. Михайлов // Тезисы 1-й национальной конференции по искусственному интеллекту. – Т. 2. – 2008. – С. 207–209.

References

1. Abramov, V. V Mechanisms of integration systems [Internet]. Available from: <<http://fetmag.mrsu.ru/2010-2/pdf/Integration.pdf>>.
2. Berko, A. Yu. (2008). Methods and tools for semantic data integration. *Visnyk Natsyonal'noho unyversytetu «L'vivs'ka politekhnika»*, (610), pp. 190–199 [in Ukrainian].
3. Berko, A. Yu. and Matrofoylo, I. M. (2008). Information system of autodealer network management based on technologies of integration. *Visnyk Natsyonal'noho unyversytetu «L'vivs'ka politekhnika»*, (631), pp. 19–28 [in Ukrainian].
4. Berko, A. Yu. (2010) Structural and semantic data integration based on factual relational model. *Visnyk Natsyonal'noho unyversytetu «L'vivs'ka politekhnika»*, (663), pp. 60–70 [in Ukrainian].
5. Berko, A. Yu. (2010) Models of data integration in open information systems. *Aktual'ni problemy ekonomiky*, (10), pp. 147–152 [in Ukrainian].
6. Frangulova, E. V. (2010) Classification of approaches to integration and interoperability of information systems. *Vestnik AGTU*, (2), pp. 176–180 [in Russian].
7. Zin'kovich, V. M. (2006) Concepts and models for integration of metatechnology objects in METACASE environment. *Problemy programuvannya*, (2–3), pp. 635–640 [in Ukrainian].
8. Mikhailov, Y. S. (2008) Research and development of methods and software for providing structural and semantic interoperability of information systems based on metamodels. *Pervaya natsionalnaya konferentsiya po iskusstvennomu intellektu: theses*, pp. 134–137 [in Russian].

M. V. Dvoieglazova, PhD., associate professor
Chernigiv National Technological University
Shevchenko str., 95, Chernigiv, Ukraine
marinitta_1905@mail.ru

ANALYSIS OF METHODS AND TECHNOLOGIES FOR INTEGRATION OF INFORMATION SYSTEMS

The analysis of existing methods of integration provides an overview of current trends and technologies of integration, codifies the existing knowledge on integration processes of information systems. Information resources of information systems involve the use of large amounts of heterogeneous data. The problem of data integration consists in creating and using of data presentation methods. It is found that the processes of data integration have a broad sphere of application. Today many tools for a solution of the problem of integration at various levels are presented. Integration of business processes, integration of corporate appendices, integration of corporate platforms and integration of data are the basic of them. At integration of information systems the important place is taken by interoperability, that is ability of information system to interact with other information systems. Two aspects of interoperability: structural and semantic are distinguished. Structural aspect of systems interoperability means the ability to structural coherence of existing systems. Semantic aspect means the possibility of the establishment of compliance between the maintenance of information systems. Systematization of knowledge of existing methods of integration will allow to use combinations of these methods in modeling of processes of information systems integration.

Keywords: *integration, information systems, interoperability, methods, technologies.*

*Рецензенти: Є. Ю. Сахно, д.т.н., професор,
С. В. Цюцюра, д.т.н., професор*