

Міністерство освіти і науки України
Національний університет «Львівська політехніка»

Кушнірецька Ірина Ігорівна



УДК 004.738.5 + 004.62

**ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ДИНАМІЧНОЇ ІНТЕГРАЦІЇ
СЛАБОСТРУКТУРОВАНИХ ДАНИХ У WEB-СИСТЕМАХ**

05.13.06 - інформаційні технології

Автореферат
дисертації на здобуття наукового ступеня
кандидата технічних наук

Львів – 2017

Дисертацією є рукопис

Робота виконана у Національному університеті «Львівська політехніка»
Міністерства освіти і науки України

Науковий керівник:

доктор технічних наук, професор
Берко Андрій Юліанович,
Національний університет «Львівська політехніка»,
завідувач кафедри загальної екології та
екоінформаційних систем.

Офіційні опоненти:

доктор технічних наук, професор
Бісікало Олег Володимирович,
Вінницький національний технічний університет,
декан факультету комп'ютерних систем і автоматики.

кандидат технічних наук, доцент
Ковалюк Тетяна Володимирівна,
Національний технічний університет України
“Київський політехнічний інститут імені Ігоря Сікорського”,
доцент кафедри автоматизованих систем обробки
інформації та управління.

Захист відбудеться 17 березня 2017 р. о 14⁰⁰ годині на засіданні спеціалізованої
вченої ради Д 35.052.14 у Національному університеті «Львівська політехніка» за
адресою: 79013, м. Львів, вул. С. Бандери, 28а, ауд. 807, V навч. корпус.

З дисертацією можна ознайомитись у бібліотеці Національного університету
«Львівська політехніка» (79013, м. Львів, вул. Професорська, 1).

Автореферат розіслано «14» лютого 2017 р.

Учений секретар
спеціалізованої вченої ради,
кандидат технічних наук, доцент



А.Є. Батюк

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. Зростаюча кількість ресурсів та послуг, доступних у глобальній мережі та можливостей, які пропонуються Web 2.0 штовхають кінцевих користувачів еволюціонувати від пасивних споживачів інформації в інформаційних виробників, які в змозі отримати доступ до існуючих інформаційного контенту та маніпулювати ним для того, щоб генерувати новий зміст. З точки зору людино-машинної взаємодії, це вимагає нових парадигм комунікацій, які повинні дозволити людям отримати доступ до інформаційних даних, перемістити їх в особисті інтерактивні робочі області, де вони можуть інтегрувати їх, а також, при потребі створювати новий інформаційний контент, використовуючи наявний. Одним з рішень даної проблеми є розробка систем динамічної інтеграції слабоструктурованих даних. Найперспективнішими в наш час є системи інтеграції даних, що працюють використовуючи технологію Mashup. Mashup - це технологія для проектування web-систем, яка дозволяє користувачам об'єднувати різного роду дані з декількох джерел в один інтегрований інструмент. Mashup системи відкривають нові і широкі можливості використання інформаційних ресурсів, і тому при створенні програмних продуктів все більше розробників переходять на використання Mashup технології інтеграції даних. Серед відомих корпорацій розробників Mashup систем є IBM, Google, Microsoft, Intel, Yahoo, тощо. Однак, існуючі методології та інструментальні засоби побудови програмних систем орієнтовані на добре структуровані задачі із достатньо формалізованими предметними областями і постійними локальними джерелами знань.

Розроблення технології динамічної інтеграції слабоструктурованих даних із врахуванням змісту даних, що знаходяться в різних web-системах та здатної враховувати особливості різних вхідних інформаційних систем актуально беручи до уваги такі фактори як: недостатність теоретичного обґрунтування методів семантичного опрацювання даних систем динамічної інтеграції даних, а також, необхідність уніфікації програмних засобів динамічного опрацювання інформаційних ресурсів інтегрованих систем Mashup. Практичний чинник опрацювання даних різноманітної структури і форматів подання, що знаходяться в різних web-системах при їх динамічній інтеграції пов'язаний з вирішенням задач покращення пошуку інформації в Інтернет в сьогоdnішню еру стрімкого зростання об'єму інформаційних даних різноманітного характеру та змісту, інтеграції гетерогенних інформаційних систем, що вміщують інформаційні ресурси різного роду і змісту та величезний попит на системи динамічної інтеграції даних на основі технології Mashup, що відкривають нові і широкі можливості використання інформаційних ресурсів.

Теоретичний чинник опрацювання даних різноманітної структури і форматів подання, що знаходяться в різних web-системах при їх динамічній інтеграції пов'язаний із ІТ семантичного динамічного опрацювання даних при їх інтеграції. Стандартні технології для вирішення однотипних завдань інтеграції почали розроблятися ще на початку 90-х років. Тоді був заснований консорціум OMG, для розробки та стандартизації технологій взаємодії між інформаційними системами. Дослідження в напрямку Mashup інтеграції даних активно почали розвиватися 8 років тому. В наукових роботах Фішера (Т. Fischer), Бакалова (F.

Bakalov), Наєрца (A. Nauertz), Алтінела (M. Altinel), Брауна (P. Brown), Клайна (S. Cline), Карфа (R. Kartha), Луї (E. Louie), Маркла (V. Markl), Мау (L. Mau), Сіммена (D. Simmen) та Сінга (A. Singh) наведено характеристику сучасних підходів створення Mashup систем та опис роботи деяких сучасних Mashup застосунків. Хендрік (Hendrik), Анйомшоа (A. Anjomshoaa) та Тйоа (A. Tjoa) описують, в своїх роботах, Mashup засоби для аналізу різного об'єму даних. Основні нюанси моделювання Mashup простору наводять такі науковці, як: Абїтбул (S. Abiteboul), Гріншпен (O. Greenshpan) та Мілоу (T. Milo). Предметно-орієнтовану мову для web-інтерфейсів та сервісів Mashup систем пропонують Максїмїльян (E. Maximilien), Вілкінсон (H. Wilkinson), Дїсей (N. Desai) і Тай (S. Tai). Згідно проведеним науковим аналізом встановлено, що даний сегмент ІТ є мало дослідженим. Кожний окремий проект реалізують практично з початку, фактично на основі своїх ідей та рішень. У літературі надзвичайно мало висвітлені суттєві теоретичні обґрунтування, дослідження, висновки, рекомендації, узагальнення для проектування систем динамічної інтеграції даних на основі технології Mashup та опрацювання даних в таких системах, враховуючи їх семантику. Виникла потреба в аналізі, узагальненні та обґрунтуванні існуючих підходів створення і роботи систем динамічної інтеграції даних на основі технології Mashup. Актуальною є задача створення комплексу технологічних засобів на основі теоретичного обґрунтування методів, моделей і принципів семантичного опрацювання даних різноманітної структури і форматів подання, що знаходяться в різних web-системах при їх динамічній інтеграції.

Зв'язок роботи з науковими програмами, планами, темами. Тема дисертаційної роботи відповідає науковому напряму кафедри інформаційних систем та мереж Національного університету «Львівська політехніка» на тему «Дослідження, розроблення і впровадження інтелектуальних розподілених інформаційних технологій та систем на основі ресурсів баз даних, сховищ даних, просторів даних та знань з метою прискорення процесів формування сучасного інформаційного суспільства». Результати досліджень, відображені у дисертаційній роботі, отримані у рамках виконання науково-дослідної роботи кафедри інформаційних систем та мереж Національного університету «Львівська політехніка» за темою «Розроблення інтелектуальних розподілених систем на основі онтологічного підходу з метою інтеграції інформаційних ресурсів» № держреєстрації 0115U004228 (автор вдосконалила структурну модель Mashup системи, розробила метод визначення структури і змісту отриманих вхідних даних та метод формування контенту об'єднаного динамічного набору даних, який має загальну структуру і єдиний зміст).

Мета і завдання дослідження. Мета дисертаційної роботи - підвищення якості результатів пошуку даних у web-системах шляхом розроблення інформаційної технології динамічної інтеграції слабоструктурованих даних у web-системах.

Для досягнення зазначеної мети необхідно вирішити такі основні **завдання дослідження**:

- провести аналіз сучасного стану опрацювання даних web-систем в сенсі застосування технології Mashup;

- розробити метод опису структури і змісту вхідних даних Mashup системи;
- розробити метод формування об'єданого динамічного набору вихідних даних, який має узгоджену структуру і єдиний зміст;
- удосконалити структурну модель системи динамічної інтеграції слабоструктурованих даних на основі технології Mashup;
- розробити програмні засоби динамічної інтеграції слабоструктурованих даних у web-системах.

Об'єкт дослідження – процеси динамічної інтеграції слабоструктурованих даних у web-системах.

Предмет дослідження – інформаційні технології динамічного об'єднання наборів даних на основі принципів Mashup.

Методи дослідження. Для вирішення проблеми аналізу та дослідження сучасного стану опрацювання даних web-систем в сенсі застосування технології Mashup було використано теорію систем та системний аналіз. Для розроблення методів та алгоритмів опрацювання даних інформаційних систем, що інтегруються було використано онтології та стандарт онтологічного моделювання IDEF5, теорію множин та математичну логіку. При побудові програмних засобів комплексного динамічного опрацювання даних різноманітної структури і форматів подання різних web-систем було використано методи об'єктно-орієнтованого програмування та методи сервісно-орієнтованої архітектури.

Наукова новизна одержаних результатів. У процесі теоретичних та експериментальних досліджень отримано такі нові наукові результати:

Вперше

- Розроблено метод опису структури і змісту вхідних даних Mashup системи шляхом застосування онтологій для класифікації інформаційних ресурсів, що дало змогу виконувати процеси інтеграції даних із врахуванням їх змісту.
- Розроблено метод формування об'єданого динамічного набору вихідних даних через застосування процедур лінгвістичного аналізу змісту запиту користувача, що забезпечило підвищення релевантності та узгодженості результатів динамічної інтеграції слабоструктурованих даних.

Удосконалено

- Структурну модель системи динамічної інтеграції слабоструктурованих даних на основі технології Mashup, що відрізняється від відомих доданням функціональних компонентів визначення та опису семантики вхідних та вихідних наборів даних, що дало змогу підвищити якість результатів застосування таких систем.

Отримала подальший розвиток Mashup технологія динамічної інтеграції даних через запровадження додаткових можливостей опрацювання семантики інформаційних ресурсів, що дало змогу підвищити якість результатів інформаційно-пошукових web-систем.

Практичне значення одержаних результатів. Розроблено рекомендації проектування структури Mashup системи динамічної інтеграції слабоструктурованих даних, відмінної від існуючих через наявність підсистем динамічного

семантичного опрацювання інформаційних ресурсів, що дає змогу покращити якість результатів роботи Mashup системи. Розроблено та впроваджено програмні засоби, які дають можливість підвищити якість даних, отриманих в результаті динамічної інтеграції: збільшення точності до 6 %, зменшення інформаційного шуму до 6%, покращення узгодженості результатів інформаційних даних до 10% у системі, що працює за технологією Mashup.

Одержані у роботі результати реалізовано в проектах: News Mashup from Social Networks - система моніторингу новин із соціальних мереж (<http://snmashup.in>), елементи якої використано при побудові проектів для: ТОВ «Торгова група «Тиса», ДП «Латориця» та ФОП Сидоряк Р.Б.; Shares and Discounts Mashup – система пошуку знижки на купівлю товару, отримання послуги тощо (<http://discountsmashup.in.ua>); системах аналізу та опрацювання даних, розроблених для КЗ ЛОР «Львівське обласне патологоанатомічне бюро» та стоматологічної клініки «Улюблений доктор», що підтверджено відповідними актами впровадження. Результати впроваджено також в навчальному процесі НУ «Львівська політехніка» під час викладання дисципліни «Об’єктно-орієнтоване програмування».

Особистий внесок здобувача. Усі наукові результати дисертаційної роботи отримані дисертантом особисто. У наукових працях, опублікованих у співавторстві, здобувачу належать: [1, 2, 10] – характеризувано проектування Mashup системи та формування контенту динамічного набору даних, який має загальну структуру і єдиний зміст; [4] – опис процесу та алгоритм побудови загальної структурної онтологічної інформаційної моделі даних систем, що інтегруються; [5] – концептуалізація та моделювання роботи Mashup системи, метод опису структури і змісту вхідних даних; [6, 14] – структурно-динамічна модель Mashup системи в певний дискретний момент часу; [7, 10, 11] – опис інтеграції слабоструктурованих даних на різних рівнях складності, принципи перетворення слабоструктурованих даних у структуровані; [14, 16] - досліджено використання онтологій для побудови Mashup систем; [3, 12, 13] – опис основних складових математичної моделі онтології інформаційної системи для семантичного пошуку і зберігання даних; [15] – характеристика трирівневої побудови Mashup системи; [8, 9, 17] - опис використання фасетного методу класифікації у системі підтримки прийняття рішень.

Апробація результатів дисертації. Основні результати роботи доповідалися на таких семінарах та конференціях: Міжнародна наукова конференція «Інтелектуальні системи прийняття рішень та проблеми обчислювального інтелекту» (Залізний Порт, 25-28 травня 2015); The 7, 8, 9-th International Scientific and Technical Conference «Computer Sciences and Information Technologies» (Lviv, November 20-24, 2012; Lviv, November 11-16, 2013; Lviv, November 18-22, 2014); The III International Scientific-Practical Conference «Information Control Systems and Technologies» (Odessa, September 23-25, 2014); II, III Міжнародна наукова конференція «Інформація, комунікація, суспільство 2014» (Львів-Славське, 16-19 травня 2013, 21-24 травня 2014); The 6-th International Conference of Young Scientists «Computer Science and Engineering» (Lviv, November 21-23, 2013); XI Відкрита наукова конференція ІМФН (Львів, 17-18 травня 2012, 13-14 червня

2013). Результати дисертаційних досліджень регулярно доповідалися на наукових семінарах кафедри «Інформаційні системи та мережі» НУ «Львівська політехніка» (2013-2015).

Публікації. Основні результати дисертації відображені у 17 наукових публікаціях, з них 6 статей у наукових фахових виданнях України, 1 публікація у науковому виданні України, яке входить до міжнародних наукометричних баз (Index Copernicus, Ulrich's Periodicals Directory, DRIVER, Bielefeld Academic Search Engine (BASE), ПИНЦ, ResearchBib, Directory of Open Access Journals (DOAJ), WorldCat, EBSCO), 2 публікації у наукових періодичних виданнях інших держав (включені до міжнародних наукометричних баз BazTech, Cabell's Directory, CNKI Scholar, ERIH PLUS, Index Copernicus, J-Gate, Google Scholar, ТЕМА Technik und Management) та 8 публікацій за матеріалами наукових конференцій.

Структура та обсяг дисертації. Дисертація, обсягом 164 сторінки, складається зі вступу, чотирьох розділів, висновків, списку використаної літератури та додатків. Основний текст роботи викладено на 137 сторінках, дисертація містить 29 рисунків та 13 таблиць. Бібліографічний список налічує 160 найменувань.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У **вступі** обґрунтовано актуальність теми, сформульовано мету та основні завдання дослідження, показано зв'язок із науковими програмами, планами, темами, сформульовано наукову новизну. Розглянуто практичну цінність, реалізацію та впровадження результатів роботи. Наведено дані про особистий внесок здобувача, апробацію роботи та публікації.

У **першому розділі** дисертаційної роботи розглянуто теоретико-методологічні засади динамічної інтеграції слабоструктурованих даних у web-системах. Дано характеристику сучасним методам динамічної інтеграції інформаційних ресурсів та сучасним технологіям реалізації інтеграції даних. Описано динамічну інтеграцію даних на основі технології Mashup та характеристику сучасних Mashup систем. Визначено задачі і проблеми динамічної інтеграції слабоструктурованих даних у web-системах.

На даний час, в контексті інтеграції даних, найбільш активно розвивається динамічна інтеграція даних з використанням технології Mashup. Все більше і більше web-систем можна зустріти на основі даної технології. Mashup системи – системи, які частково або повністю в якості джерел даних використовують інші web-системи, забезпечуючи користувачеві новий функціональний інструмент для роботи. Формально опис Mashup системи динамічної інтеграції даних подають наступним чином:

$$H = \langle DI, Q, DO, C, T \rangle, \quad (1)$$

де: $DI = \{di_i | i = \overline{1, N_{DI}}\}$ - набори вхідних даних; DO - глобальний динамічний набір вихідних даних; $C = \{c_i | i = \overline{1, N_C}\}$ - множина умов інтеграції; $T = \{t_i | i = \overline{1, N_T}\}$ - час транзакцій оновлення даних; $Q = \{q_i | i = \overline{1, N_Q}\}$ - множина запитів користувачів; $H = \{h_i | i = \overline{1, N_H}\}$ - набір вихідних характеристик роботи системи.

У процесі динамічної інтеграції даних із множини вхідних даних виділяють певну підмножину значень DI^* , яку описують певною схемою S . Із даних підмножин, шляхом об'єднання, накладання різнотипних даних та створення глобальної схеми, формується єдиний глобальний динамічний набір вихідних даних DO для подальшого представлення користувачеві. Джерелами даних Mashup систем можуть бути: різноманітні web-сайти, портали, форуми, соціальні мережі, тощо. Дані джерел інформаційних ресурсів Mashup систем подаються у слабкоструктурованому вигляді. При об'єднанні таких даних необхідно брати до уваги семантичний чинник подання інформації, а це буває досить складно, особливо у web-просторі.

Згідно проведеного огляду літературних джерел було виділено основні технології (frameworks) створення сучасних Mashup систем: технологія на основі парадигми програмування, технологія на основі скриптових мов, технологія на основі електронних таблиць, технологія на основі блочної парадигми, технологія на основі програмування через демонстрацію, технологія на основі автоматичного створення Mashups. Але існуючим технікам для повноцінного вирішення завдання динамічного відображення вхідних наборів даних у вихідний інтегрований набір даних зі збереженням змісту даних недостатньо як масштабованості, так і функціонального охоплення. Ці підходи в основному концентруються на простому переміщенні даних з однієї системи в іншу і застосуванні трансформації та агрегації до елементів даних на рівні структур і форматів без врахування змісту даних. Вирішити цю проблему допоможе застосування методів та технологічних засобів підвищення якості результатів динамічної інтеграції слабкоструктурованих даних у web-системах, шляхом розроблення інформаційної технології, яка ґрунтується на опрацюванні їх семантики. Це обґрунтовує мету, актуальність, доцільність та напрями дослідження.

У **другому розділі** дано характеристику принципам роботи web-систем динамічної інтеграції слабкоструктурованих даних, описано основні стани діяльності Mashup системи динамічної інтеграції слабкоструктурованих даних, деталізовано охарактеризовано процеси опрацювання даних в Mashup системі, розроблено метод опису структури і змісту вхідних даних, що дало змогу врахувати зміст даних у процесах їх інтеграції.

Нехай процес опрацювання даних з (1) описано оператором:

$$do_j(t_{p+1}) = Mashup(di_i, q_d, c_k, t_p). \quad (2)$$

Для того, щоб зрозуміти основну суть роботи Mashup системи і яким чином дана система працює з вхідними та вихідними даними було визначено основні стани діяльності Mashup систем. Найважливішими станами при роботі Mashup системи є пошук даних, витягнення даних та їх подання у вигляді сервісу. При роботі з інформаційними ресурсами на цих станах досить часто може виникнути ряд проблем, які можуть вплинути не найкращим чином на результати роботи системи. Від коректної роботи системи на даних трьох станах залежить безпосередньо кінцевий результат системи. Стани пошуку та витягнення даних належать до процесу роботи системи із вхідними даними, а стан подання даних у вигляді сервісу – процесу відображення вихідних даних. Звідси, отримуємо, що

процес відображення вхідних наборів у вихідний набір даних складається із: процесу визначення вхідних даних та процесу формування вихідних даних (2), який подано, як суперпозиція функцій

$$Mashup = \lambda \circ \beta, \quad (3)$$

де: λ - оператор визначення вхідних даних; β - оператор формування вихідних даних.

Mashup систему, враховуючи (3), подано як

$$H' = \langle DI, Q, DO, C, T, \lambda, \beta \rangle, \quad (4)$$

Оператор визначення вхідних даних λ є відображенням слабоструктурованих вхідних даних di_i у структурований вигляд із збереженим змістом даних.

$$\lambda : (di_i, t_p, c_s) \rightarrow (di_{i+1}, t_{p+1}), \quad (5)$$

де: $C_s = \{c_{is} \mid i = \overline{1, N_{c_s}}\}$, $C_s \subseteq C$ - множина умов визначення вхідних даних.

Нехай $di_i^* = di_{i+1}$ - структуровані дані, згідно дії λ на di_i . І їх подано за допомогою (6).

$$di_i^* = \left\{ \bigcup_{i=1}^{N_{di}} di_i \mid \forall di_i \in DI_{c_s}, \exists c_s \in C_{di_i}, \right. \\ \left. DI = DI_{c_s} \cup DI_{\overline{c_s}}, C_s = C_{di_i} \cup C_{\overline{di_i}}, s = \overline{1, N_c} \right\}. \quad (6)$$

Де множина умов визначення вхідних даних подана як

$$c_s = \left\{ \bigcup_{j=1}^K c_{is} \mid \forall c_{is} \in C_{di_i}, \exists di_i \in DI_{c_s}, c_{is} \notin C_{\overline{di_i}}, \right. \\ \left. C_s = C_{di_i} \cup C_{\overline{di_i}}, DI_{c_s} \subseteq DI, s = \overline{1, N_c}, i = \overline{1, M} \right\}. \quad (7)$$

Оператор формування вихідних даних β є відображенням структурованих вхідних даних di_i^* у набір вихідних даних із узгодженим змістом даних do_j .

$$\beta : (di_i^*, q_d, t_p, c_u) \rightarrow (do_j, t_{p+1}), \quad (8)$$

де: $C_u = \{c_{iu} \mid i = \overline{1, N_{c_u}}\}$, $C_u \subseteq C$ - множина умов формування вихідних даних; $q_d = \{q_{id} \mid i = \overline{1, N_{q_d}}\}$ - множина запитів користувачів.

Формування контенту динамічного набору даних подано як

$$do_j = \left\{ \bigcup_{i=1}^{N_{di^*}} di_i^*(q_d, t_r) \mid \forall di_i^* \in DI_{q_d}, DI_{q_d} = \beta(DI_{q_d}), \exists q_d \in Q, \exists c_u \in C_{di_i^*}, \right. \\ \left. DI = DI_{q_d} \cup DI_{\overline{q_d}}, C_u = C_{di_i^*} \cup C_{\overline{di_i^*}}, d = \overline{1, N_Q}, s = \overline{1, N_C} \right\}. \quad (9)$$

Умови формування вихідних даних подано як

$$c_u = \left\{ \bigcup_{j=1}^K c_{iu} \mid \forall c_{iu} \in C_{di_i^*}, \exists di_i^* \in DI_{c_u}, c_{iu} \notin C_{\overline{di_i^*}}, \right. \\ \left. C_u = C_{di_i^*} \cup C_{\overline{di_i^*}}, DI_{c_u} \subseteq DI, u = \overline{1, N_u}, i = \overline{1, L} \right\}. \quad (10)$$

Процес визначення вхідних даних забезпечує динамічне отримання структурованих даних із збереженим змістом даних для їх подальшого використання у Mashup системі.

$$S(di_i) \rightarrow di_i \rightarrow DI \rightarrow \lambda(di_i, t_p, c_s) \rightarrow di_i^* \rightarrow DI^* \rightarrow D(DI^*), \quad (11)$$

де: $S(di_i)$ - джерело даних; $D(DI^*)$ - база даних.

Для підвищення якісних показників результату процесу визначення вхідних даних $\lambda : DI \rightarrow DI^*$ було розроблено метод визначення структури і змісту вхідних даних, який складається із певних етапів - кроків для визначення структури і змісту вхідних інформаційних ресурсів, також, було описано завдання, які потрібно вирішити на кожному з цих кроків. Рішення завдань кожного кроку методу отримуються використовуючи такі семантично-орієнтовані технології, як онтології і дескриптова логіка.

Метод опису структури і змісту вхідних даних – комплекс заходів забезпечення підвищення якості процесу отримання структурованих даних з різних інформаційних джерел для їх подальшого використання у системі динамічної інтеграції слабоструктурованих даних. Метод подано, як послідовність кроків:

Крок 1. Визначення форми подання вхідних даних.

В дисертаційній роботі розглядаються певні наперед визначені системи для інтеграції (соц. мережа, web-сторінка, web-сайт, web-сервіс), тому вхідними даними виступають інформаційні ресурси цих систем. Формати інформаційних ресурсів, які розглядалися: web-формат, табличний формат, формат на основі розмітки та мультимедійний контент.

Крок 2. Встановлення границь та задоволення основних обмежень, що стосуються вхідного потоку web-інформації.

$$DI_1^* = \lambda_2(DI, \lambda_1(DI, C_{1S}, T), C_{2S}), \quad (12)$$

де: λ_2 - оператор встановлення границь та задоволення основних обмежень; C_{2S} - множина умов встановлення границь та задоволення основних обмежень; λ_1 - оператор визначення форми подання вхідних даних; C_{1S} - множина умов визначення форми подання вхідних даних.

Крок 3. Класифікація вхідної інформації відповідно до предметної області із виділенням і збереженням семантики даних шляхом формування глобальної метамоделі всіх систем, що інтегруються. Даний крок поділяється на етапи:

- структурна онтологія системи, що інтегрується;
- загальна структурна онтологія всіх систем, що інтегруються;
- глобальна метамоделі систем, що інтегруються.

$$DI_2^* = \lambda_3(DI, C_{3S}, T, DI_0^*, DI_1^*), \quad (13)$$

де: $\lambda_3 = \lambda_{31} \circ \lambda_{32} \circ \lambda_{33}$ - оператор формування глобальної метамоделі всіх систем, що інтегруються; C_{3S} - множина умов формування глобальної метамоделі всіх систем, що інтегруються.

Формально, отримана структурна онтологія системи, що інтегрується має вигляд:

$$O = \langle X, Att, f_{Att}(X), R \rangle, \quad (14)$$

де: $X = \{x_i \mid i = \overline{1, N_X}\}$ - множина класів (концептів), $x_i = \{name_i, type_i\}$;
 $Att = \{att_{ij}, i = \overline{1, N_{Att}}, j = \overline{1, M_{Att}}\}$ - множина атрибутів концептів; $f_{Att}(X)$ - функція, що

визначає для кожного концепту множину його атрибутів; $R = \{r_i \mid i = \overline{1, N_R}\}$ – множина відношень між концептами онтології.

Наступне завдання полягає в об'єднанні розрізнених метамodelей в одну єдину загальну онтологічну модель структури всіх інтегрованих систем.

$$O_S = \bigcup_{i=1}^n O_i, \quad (15)$$

За допомогою формули (15) зображено загальну модель структури всіх інтегрованих систем, однак її ще недостатньо для автоматичного логічного аналізу при витяганні семантичних метаданих інформаційних ресурсів. Для цього метою третього завдання є ще більше розширення моделі і перетворення її на глобальну метамodelь, що використовує інформацію про предметну область та світ поза нею. Опис глобальної метамodelі систем, що інтегруються міститься у формулі (16):

$$DI_2^* = \langle X, Att, R, f_{Att}(X), f_W(X), W \rangle, \quad (16)$$

де: $W = \{w_{ij} \mid i = \overline{1, N_W}, j = \overline{1, M_W}\}$ - словник, в якому визначаються терміни предметної області; $f_W(X)$ - функція інтерпретації концептів, зіставляє концепту набір термінів зі словника W .

Крок 4. Виділення у вхідному інформаційному ресурсі набору реквізитів (атрибутів), що відображають його основні характеристики і аспекти зазначеної предметної області.

Завдяки розробленій на другому кроці структурній інформаційній метамodelі можна на її основі виділити в обраній системі для інтегрування потрібні метадані інформаційних ресурсів, що зберігаються в системі. Тобто витягнути з баз даних семантичні метаописи інформаційних ресурсів, що зберігаються в них за допомогою глобальної метамodelі і механізмів автоматизованого логічного прийняття рішень.

Вага терміну y_j у метаописі описується коефіцієнтом k_W :

$$k_W(y_j) = k_{IN}(y_j) + k_F(y_j), \quad (17)$$

де: $k_{IN}(y_j)$ - коефіцієнт частоти кількості входжень деякого слова ($h_{IN}(y_j)$) до загальної кількості слів:

$$k_{IN}(y_j) = \frac{h_{IN}(y_j)}{\sum_{i=1}^{N_Y} y_i}, \quad (18)$$

$k_F(Y)$ - коефіцієнт форматування терміну визначається відповідно до форматування слова у тексті: без форматування, жирним, курсивом, підкресленням, хештегом. Якщо не має форматування – коефіцієнт дорівнює 0, якщо один вид форматування застосовано, то коефіцієнт дорівнює 1, якщо 2, то – 2, якщо 3, то - 3, якщо 4, то - 4.

В результаті отримано наступний вигляд глобальної метамodelі систем, що інтегруються:

$$DI_3^* = \langle X, Att, R, f_{Att}(X), f_W(X), W, Mir \rangle, \quad (19)$$

де: $Mir = \{Mir_i \mid i = \overline{1, N_{Mir}}\}$ - множина метаописів інформаційних ресурсів;

Крок 5. Формування опису кожного вхідного інформаційного ресурсу загальної структури, визначеної предметною областю.

$$DI_4^* = \langle URI, IR, TR, M, El, P_{RM}, P_R, C, FC \rangle, \quad (20)$$

де: URI - адреса Web-сервісу, представлена ідентифікатором URI даного сервісу; IR – ідентифікатор ресурсу; TR – тип ресурсу; M – подання ресурсу у метамоделі системи; $El = \{el_i \mid i = \overline{1, N_{El}}\}$ – множина елементів об'єктної схеми інформаційного ресурсу; $P_{RM} : El \rightarrow X$ - відображення, що ставить у відповідність елементу об'єктної схеми його концепт; $P_R : El \times El \rightarrow R$ - відображення, що ставить у відповідність зв'язкам між елементами об'єктної схеми відношення в онтології, і для будь-якого елемента $el \in El$ виконується умова: множина атрибутів елемента об'єктної схеми відповідає атрибутам його концепту, тобто $\{att : \langle att, d \rangle \in el\} = P_X(P_{El}(el))$; C (Calls) - кількість викликів web-сервісу; FC (Failed calls) - кількість невдалих викликів.

Третій розділ присвячено методу формування об'єднаного динамічного набору вихідних даних, який має узгоджену структуру і єдиний зміст, проведено аналіз вхідних інформаційних ресурсів для визначення їх спільних рис та виявлення зв'язків між ними, запропоновано процедуру визначення подібності схем даних систем, що інтегруються, запропоновано процедуру визначення подібності інформаційних ресурсів систем, що інтегруються, розроблено алгоритми створення загальної динамічної структури для множини вхідних інформаційних ресурсів із врахуванням їх змісту.

Для вирішення завдання другого кроку методу опису структури і змісту вхідних даних, що стосується класифікації вхідних даних відповідно до предметної області із виділенням і збереженням семантики даних шляхом формування глобальної метамоделі всіх систем, що інтегруються побудовано алгоритм побудови онтологічної моделі всіх систем, що інтегруються. Алгоритм побудови онтологічної моделі всіх систем, що інтегруються містить в собі 6 основних кроків, а саме:

1) Подання структури бази даних у вигляді RDF, тобто послідовне відображення схеми S в RDF формат.

$$X_i \rightarrow X(RDF)_i, Att_j \rightarrow Att(RDF)_j, \quad (21)$$

де: $X(RDF)_i, i = \overline{1, n}$ - концепти онтології, описані за допомогою RDF; $Att(RDF)_j, j = \overline{1, k}$ - властивості концептів в онтології.

2) Додавання семантичних властивостей та створення онтології.

Нехай отримано онтології: $O_1 = \langle X^1, Att^1, f_{Att}(X)^1, R^1 \rangle$ і $O_2 = \langle X^2, Att^2, f_{Att}(X)^2, R^2 \rangle$.

Було виділено такі варіанти можливих семантичних властивостей між елементами онтології:

- Властивість узагальнення: $s_1 : P(x_1) = x_2, x_2 = \{x_{2i}\}$, якщо $b_1 < k(x_1, x_{2i}) < b_2$, де $P(x_1)$ – відображення концепту x_1 ; b_2 – максимальне значення коефіцієнта семантичної подібності $k(x_1, x_{2i})$, при якому будується відображення концепту x_1 в онтологію O_2 ; b_1 – мінімальне значення подібності для встановлення відсутності еквівалентності концептів.

- Властивість еквівалентності: $s_2 : P(x_1) = x_2$, якщо $k(x_1, x_{2i}) \geq b_2$, де b_2 – максимальне значення коефіцієнта семантичної подібності $k(x_1, x_{2i})$, при якому будується відображення концепту x_1 в онтологію O_2 .

- Властивість часткової еквівалентності: $s_3 : P(x_1) = x_2$, якщо $b_1 < k(x_1, x_2) < b_2$, b_2 – максимальне значення коефіцієнта семантичної подібності $k(x_1, x_2)$, при якому будується відображення концепту x_1 в онтологію O_2 ; b_1 – мінімальне значення подібності для встановлення відсутності еквівалентності концептів.

- Властивість уточнення: $s_4 : P(x_1) = x_2, x_2 = \{x_{1i}\}$, якщо $b_1 < k(x_{1i}, x_2) < b_2$, де $P(x_1)$ – відображення концепту x_1 ; b_2 – максимальне значення коефіцієнта семантичної подібності $k(x_{1i}, x_2)$, при якому будується відображення концепту x_1 в онтологію O_2 ; b_1 – мінімальне значення подібності для встановлення відсутності еквівалентності концептів.

- Властивість відмінності: $s_5 : P(x_1) = 0, \exists x_1, \forall x_2 \in O_2, k(x_1, x_2) \leq b_1, b_1$ – мінімальне значення подібності для встановлення відсутності еквівалентності концептів.

3) Додавання онтологій верхніх рівнів та онтології предметної області. Реалізується даний крок за допомогою мови OWL, використовуючи команду owl:import. Завдяки правилу транзитивності в RDF, додаткові онтології розширюють предметні області і додають нові концепти і властивості.

4) Перевірка створеної онтології. Даний крок реалізується виконанням перевірки і аналізу витягнутої онтології на «зв'язність», тобто перевіряється чи не бракує ніде семантичних зв'язків. Якщо так, тоді перехід до п'ятого кроку, якщо ні – перехід до шостого кроку.

5) Редагування витягнутої онтології за допомогою редактора онтології додавання зв'язків між концептами. Далі повернення до кроку 4.

6) Зберігання отриманої загальної онтології структури всіх систем, що інтегруються у сховище метаданих у форматі RDF.

Процес формування вихідних даних забезпечує відповідно до запиту користувача динамічне відображення структурованих вхідних даних di_i^* у набір вихідних даних із узгодженим змістом даних do_j .

Для підвищення якісних показників результату формування вихідного набору даних $\beta : (DI^*, Q) \rightarrow DO$ було розроблено метод формування контенту об'єднаного динамічного набору даних, який має загальну структуру і єдиний зміст.

Метод формування об'єднаного динамічного набору вихідних даних, який має узгоджену структуру і єдиний зміст - комплекс заходів забезпечення

підвищення якості процесу відображення структурованих даних із врахуванням їх змісту, отриманих з різних інформаційних джерел для їх подальшого візуального подання у системі динамічної інтеграції слабоструктурованих даних. Метод подано, як послідовність кроків:

Крок 1. Виділення метаописів інформаційних ресурсів із метамоделі системи, що інтегрується.

Перший крок полягає у отриманні та деталізації метаописів інформаційних ресурсів певної визначеної кількості для кожного джерела. Не потрібно розглядати абсолютно всі метаописи, адже їх семантичні зв'язки вже зазначено у метамоделі системи. І знайшовши, при порівнянні із запитом користувача, у виділеній кількості метаописів відповідні запиту – буде знайдено, також, всі інші семантично пов'язані метаописи із даними.

Крок 2. Виділення пошукового образу запиту.

На даному кроці окремо виділяються терміни із пошукового запиту та семантично подібні терміни до термінів пошукового запиту. Останні отримуються за допомогою використання загальної онтології предметної області, створеної раніше. Наступним етапом є розставлення асоціативних зв'язків між термінами пошукового запиту та розставлення асоціативних зв'язків між семантично подібними термінами до термінів пошукового запиту.

Крок 3. Знаходження подібності запиту із метаописами.

Нехай X – запит користувача, який розглядається як рядок символів і Y – це метаопис IP, який порівнюється із запитом. Тоді коефіцієнт подібності рядків обчислюється за допомогою модифікованої формули обчислення лінгвістичного коефіцієнта Серенсена-Дайса, застосувавши N -грами замість символів:

$$k_p = \frac{2 \times f_{NG}(X \cap Y)}{f_{NG}(X) + f_{NG}(Y)}, \quad (22)$$

де: k_p – коефіцієнт подібності рядків, $0 \leq k_p \leq 1$; X і Y – рядки, що порівнюються; $f_{NG}(X)$ – функція, що визначає довжину ланцюжка N -грам рядка X .

При порівнянні знаходяться подібні дані запиту. Умовою знаходження результату є певне обмеження кількості всіх даних у загальній Mashup моделі результуючих даних, що забезпечує високу релевантність результуючих даних та зниження, таким чином, інформаційного шуму. Дані обмеження на кількість результуючих даних введені для кожного джерела інформаційних ресурсів можуть бути різними – в залежності від джерела. Відповідну обмеженням кількість інформаційних ресурсів із кожного джерела дістається ще за допомогою додання подібних даних із метамоделі до знайдених даних відповідно запиту користувача. Розглядається питання пошуку мінімального порогу b семантичної подібності, при якій елемент запиту користувача і елемент семантичних метаданих приймаються еквівалентними.

$$b_1 = \max(k(el_i, el_j) | \forall el_i \in Q_0, \forall el_j \in DI_{k5}^*) \times (p_1 / 100), \quad (23)$$

де: Q_0 – запит користувача; DI_{sk}^* – метаописи інформаційних ресурсів; p_1 – відсоток, при якому b_1 приймається порогом подібності для встановлення еквівалентності та коректного відображення el_i і el_j .

Показано, що b_1 – мінімальний поріг, при якому зменшення цього значення призводить до неможливості повного відображення елементів порівняння. Знаходиться граничне значення, при якому елементи приймаються частково еквівалентними.

$$b_2 = \max(k(el_i, el_j) | \forall el_i \in Q_0, \forall el_j \in DI_{k5}^*) \times (p_2 / 100), \quad (24)$$

де: p_2 – відсоток, при якому b_2 приймається порогом подібності для встановлення часткової еквівалентності елементів.

Показано, що b_2 – мінімальне значення в тому сенсі, що зменшення цього значення призводить до некоректного відображення елементів структур порівняння. Елементи вважаються різними, якщо мають значення міри семантичної подібності такої, що не перевершує поріг b_2 .

Крок 4. Узгодження результату.

Узгодження результату полягає у ранжуванні результуючих даних відповідно до зв'язків подібності. Тобто, чим вищий коефіцієнт подібності одних даних результату іншим, тим ближче ці дані будуть знаходитися один біля одного у результуючому Mashup. Ступінь впорядкованості елементів розглядається використовуючи величину, що відображає число перестановок пари сусідніх елементів до впорядкованого положення від положення ранжування по спаданню релевантності метаописів до запиту, як:

$$d_{\max} = kmq_2(kmq_1 + kmq_0) + kmq_1 kmq_0 + \dots + kmq_m(n - kmq_m) + kmq_{m-1}(n - kmq_m - kmq_{m-1}) \quad (25)$$

де: d_{\max} – ступінь впорядкованості; $kmq_i, i = \overline{1, m}$ – коефіцієнт подібності метаопису до запиту; n_1 – номер першого релевантного метаопису.

Коефіцієнт узгодження результату має вигляд:

$$k_u = 1 - \frac{d}{\sum_{i,j=0, i>j}^m kmq_i kmq_j}, \quad (26)$$

Результатом методу є виведення сформованого контенту об'єднаного динамічного набору даних.

В **четвертому розділі** на основі аналізу базових задач систем динамічної інтеграції даних, що працюють використовуючи технологію Mashup розроблено інструментальні засоби та технології для побудови таких систем. Розроблено функціональну схему систем динамічної інтеграції даних, що працюють використовуючи технологію Mashup з підсистемами пошуку, опрацювання, зберігання та подання інформації отриманої із різних інформаційних систем. Детально описано загальну архітектуру, завдання та принципи реалізації відповідних систем, а також обговорено функціональні елементи системи і подано схеми дії найсуттєвіших механізмів. Розроблено програмні засоби підвищення якості етапів отримання, опрацювання та подання інформаційних даних.

Підсистема опису структури та змісту вхідних даних забезпечує експорт метаданих структури (підключення до інформаційної системи, отримання структурної інформації, аналіз на подібність елементів структури, додавання семантичних зв'язків на основі аналізу, експорт онтології); розширення онтології структури (додавання онтологій верхнього рівня, додавання онтологій предметної області, додавання семантичних зв'язків, експорт розширеної онтології); отримання метаданих інформаційних ресурсів (імпорт розширеної онтології структури, підключення до інформаційної системи, витяг метаданих, аналіз на схожість, додавання семантичних зв'язків на основі аналізу, додавання семантичних зв'язків на основі онтології структури, експорт онтології в сховище).

Підсистема формування контенту Mashup забезпечує аналіз семантичних метаданих вхідних інформаційних ресурсів (імпорт отриманих семантичних метаописів ресурсів, їх деталізація, виділення зазначеної кількості, експорт даних у сховище); визначення подібності запиту користувача із семантичними метаописами інформаційних ресурсів (імпорт метаописів інформаційних ресурсів, виділення пошукового образу запиту, аналіз на подібність запиту користувача та метаописів, додання подібних метаописів до знайдених, експорт даних у сховище); узгодження результату (отримання даних зі сховища, які відображають схожі до запиту користувача метаописи, ранжування результуючих даних відповідно до зв'язків подібності, запуск сервера точки доступу, подання отриманих даних).

Таблиця 1 - Результати роботи системи News Mashup from Social Networks

Експеримент	Характеристика			
	Кількість знайдених результатів	Співвідношення знайдених джерел до всіх можливих, %	Точність, %	Доступність до повної інформації, %
1	140	100	87	100
2	114	100	89	100
3	110	67	81	100
4	90	33	83	100
5	115	67	90	100

Розроблені програмні засоби для підвищення якості процесів отримання даних, опрацювання отриманих даних та їх зберігання у вигляді сервісу у Mashup системі динамічної інтеграції даних використовуються у наступних проектах: News Mashup from Social Networks (табл. 1) - система моніторингу новин із соціальних мереж (<http://snmashup.in/>) та Shares and Discounts Mashup – система пошуку знижки на купівлю товару, отримання послуги тощо (<http://discountsmashup.in.ua/>). Подано програмні реалізації розроблених систем з підсистемами, що працюють на основі розробленої інформаційної технології динамічної інтеграції слабоструктурованих даних у web-системах.

Для тестування роботи систем з підсистемами, що працюють на основі розробленої інформаційної технології динамічної інтеграції слабоструктурованих даних у web-системах було здійснено 424 експерименти. Для аналізу якості пошуку і подання інформаційних даних (табл. 2), використано такі

характеристики: критерій оцінки точності (відношення кількості правильно знайдених об'єктів до суми всіх об'єктів), критерій оцінки повноти (відношення кількості не знайдених об'єктів, які відповідають запиту до їх суми із відповідними запиту об'єктами, які було знайдено), оцінку рівня інформаційного шуму (характеризує підмножину всієї множини результатів, що відповідає за неправильно знайдені об'єкти), узгодженість даних (семантична зв'язність об'єднаного набору даних).

Таблиця 2 - Порівняльний аналіз характеристик систем на основі інформаційної технології динамічної інтеграції слабоструктурованих даних із аналогами

Назва характеристики		IFTTT	Google Alerts	Shares and Discounts Mashup	News Mashup from Social Networks
Робота із вхідним та вихідним потоком інформації					
Зіставлення даних	Ручне	+	-	-	-
	Напівавтоматичне	-	-	+	+
	Автоматичне	-	+	+	+
Опрацювання семантики	Присутнє	-	-	+	+
	Відсутнє	+	+	-	-
Якісні показники опрацювання результату	Точність	0,85	0,78	0,91	0,89
	Повнота	0,67	0,72	0,64	0,68
	Втрати інформації	0,33	0,28	0,36	0,32
	Інформаційний шум	0,15	0,22	0,09	0,11
	Узгодженість результатів, %	43,9	68,6	73,7	84,5
Оновлення даних	Стратегія витягнення даних	+	+	+	+
	Стратегія вмщення даних	-	-	+	+
	Глобальний інтервал витягнення даних	+	+	+	+
	Локальний інтервал витягнення даних	-	-	+	+
	Інтервал оновлення	+	+	+	+
Вихідні дані	Людино-орієнтовані	+	+	+	+
	Машино-орієнтовані	-	-	-	-

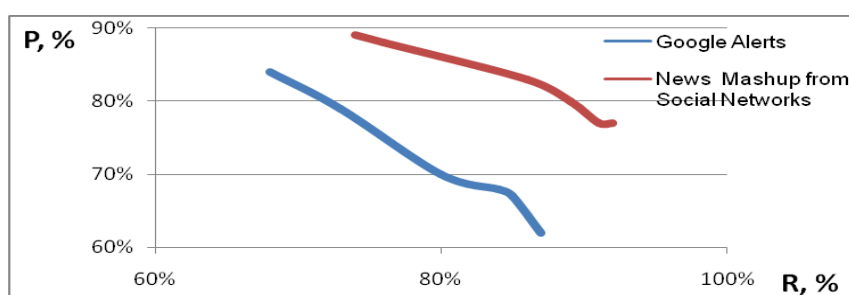


Рисунок 1 - Графік повноти-точності систем Google Alerts та News Mashup from Social Networks

Визначено якість пошуку і подання інформаційних даних для різних методів. Для порівняння проаналізовано отримання результату трьома іншими методами для одних і тих самих запитів (табл. 3). Таким чином, результати роботи системи з підсистемами, що працюють на основі розробленої інформаційної технології кращі, порівняно із результатами інших методів.

Таблиця 3 - Порівняння формування результатів роботи Mashup системи, використовуючи різні методи опрацювання даних

Технологія	Точність	Інформ. шум	Узгодженість
Розроблена ІТ	91%	9%	79%
Технологія на основі парадигми програмування	73%	27%	69%
Технологія на основі електронних таблиць	77%	23%	57%
Технологія на основі блокового подання	85%	15%	62%

Наявність підсистем визначення структури та змісту отриманої вхідної інформації та формування контенту об'єднаного динамічного набору даних, що має загальну структуру і єдиний зміст в Mashup системах сприяє збільшенню точності результуючих даних до 6%, зменшенню інформаційного шуму до 6%, а також покращенню узгодженості результатів даних до 10%.

ВИСНОВКИ

У дисертаційній роботі розв'язано актуальну наукову задачу розроблення інформаційної технології динамічної інтеграції слабоструктурованих даних у web-системах для підвищення якості пошуку даних у web-системах. Отримані в дисертації результати у сукупності є теоретичною та практичною основою для проектування Mashup систем динамічної інтеграції слабоструктурованих даних із врахуванням семантичного чинника опрацювання даних для підвищення точності результату, зниження рівня інформаційного шуму та покращення узгодженості результуючого набору даних. Отримано такі основні наукові і практичні результати:

1. Досліджено та проаналізовано принципи побудови та функціональні можливості Mashup систем динамічної інтеграції даних та здійснено концептуалізацію та моделювання роботи Mashup системи.

2. Вперше розроблено метод опису структури і змісту вхідних даних Mashup системи шляхом застосування онтологій для класифікації інформаційних ресурсів, що дало змогу виконувати процеси інтеграції даних із врахуванням їх змісту.

3. Вперше розроблено метод формування об'єднаного динамічного набору вихідних даних через застосування процедур лінгвістичного аналізу змісту запиту користувача, що забезпечило підвищення релевантності та узгодженості результатів динамічної інтеграції слабоструктурованих даних.

4. Отримала подальший розвиток Mashup технологія динамічної інтеграції даних через запровадження додаткових можливостей опрацювання семантики інформаційних ресурсів, що дало змогу підвищити якість результатів динамічної інтеграції слабоструктурованих даних.

5. Удосконалено структурну модель системи динамічної інтеграції слабоструктурованих даних на основі технології Mashup шляхом додання функціональних компонентів визначення та опису семантики вхідних та вихідних наборів даних, що дало змогу підвищити якість результатів застосування таких систем.

6. Розроблено та впроваджено програмні засоби динамічної інтеграції слабоструктурованих даних для підвищення характеристик якості пошуку і подання даних таких, як: збільшення точності до 6% та зниження інформаційного

шуму до 6%, а також покращення узгодженості результатів інформаційних даних до 10% для процесу отримання результуючої множини вихідних даних у Mashup системі динамічної інтеграції даних.

СПИСОК ПУБЛІКАЦІЙ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Кушнірецька І. І. Проектування системи динамічної інтеграції слабоструктурованих даних на основі технології Mash-Up / І.І. Кушнірецька, О.І. Кушнірецька, А.Ю. Берко // Вісник Національного університету "Львівська політехніка". – 2015. – № 832: Інформаційні системи та мережі. – С. 165-177.
2. Kushniretska I.I. Forming of the Semistructured Data Dynamic Integration Mash-Up System Content / I.I. Kushniretska // Applied Computer Science. - Volume 11, No. 3, Poland, 2015. - P. 34-44. (*Index Copernicus*).
3. Кушнірецька І.І. Семантичний пошук і зберігання даних науково-технічної інформаційної системи / О.І. Кушнірецька, І.І. Кушнірецька, А.Ю. Берко // Вісник Національного університету "Львівська політехніка". – 2015. – № 814: Інформаційні системи та мережі. – С. 310-319.
4. Kushniretska I.I. Designing of Structural Ontological Data Systems Model for Mash-UP Integration Process / I.I. Kushniretska, O.I. Kushniretska, A.Y. Berko // Applied Computer Science. - Volume 11, No. 1, Poland, 2015. - P. 39-50. (*Index Copernicus*).
5. Кушнірецька І.І. Визначення структури і змісту вхідних інформаційних ресурсів для роботи Мешап-системи / І.І. Кушнірецька, О.І. Кушнірецька, А.Ю. Берко // Технологический аудит и резервы производства — №6/3(20), 2014. С. 4-9 (*Index Copernicus*).
6. Кушнірецька І.І. Аналіз інформаційних ресурсів системи динамічної інтеграції слабоструктурованих даних web-середовищі / І.І. Кушнірецька, О.І. Кушнірецька, А.Ю. Берко // Вісник Національного університету «Львівська політехніка». – 2014. – № 805: Інформаційні системи та мережі. – С. 162-169.
7. Кушнірецька І.І. Застосування онтологій і метамоделей для динамічної інтеграції слабоструктурованих даних / І.І. Кушнірецька, О.І. Кушнірецька, А.Ю. Берко // Вісник Національного університету "Львівська політехніка". – 2014. – № 783: Інформаційні системи та мережі. – С. 128–137.
8. Кушнірецька І.І. Проектування системи підтримки прийняття рішень виборця у вигляді веб-сайта / І.І. Кушнірецька, О.М. Верес // Вісник Національного університету "Львівська політехніка". – 2012. – № 743 : Інформаційні системи та мережі. – С. 44–53.
9. Кушнірецька І.І. Система підтримки прийняття рішень для виборця з використанням фасетного методу класифікації / І.І. Кушнірецька, Л.В. Чирун // Вісник Національного університету «Львівська політехніка». – 2011. - № 699: Інформаційні системи та мережі. - С. 144-153.
10. Кушнірецька І.І. Отримання метаданих вхідного інформаційного ресурсу при його динамічній інтеграції / І.І. Кушнірецька, О.І. Кушнірецька, А.Ю. Берко // Матеріали XI Міжнародної наукової конференції «Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту» (ISDMCI'2015). – Залізний Порт, 2015.

11. Kushniretska, I.I. The ontological model of knowledge of scientific and technical information system / I.I. Kushniretska, O.I. Kushniretska, A.Y. Berko // Матеріали 9-ї Міжнародної науково-технічної конференції «Computer Sciences and Information Technologies» (CSIT'2014). — Львів: Видавництво НУ «ЛП», 2014. - Р. 47-48.

12. Кушнірецька І.І. Принципи семантичного пошуку науково-технічних інформаційних ресурсів / І.І. Кушнірецька, О.І. Кушнірецька, А.Ю. Берко // Матеріали III Міжнародної науково-практичної конференції «Інформаційні управляючі системи та технології» (ІУСТ-Одеса-2014). – Одеса: «ВМВ», 2014. – С. 253-255.

13. Кушнірецька І.І. Методика «чорної скриньки» та агентно-орієнтованого підходу для забезпечення Mash-Up інтеграції інформаційних ресурсів / І.І. Кушнірецька, А.Ю. Берко // Матеріали III Міжнародної наукової конференції «Інформація, комунікація, суспільство 2014» (ІКС-2014). — Львів: Видавництво НУ «Львівська політехніка», 2014. - С. 32-33.

14. Kushniretska, I.I. The principles of Integration the Semi-Structured Data in Web-Systems using Ontologies / I.I. Kushniretska, A.Y. Berko // Матеріали VIII Міжнародної науково-технічної конференції «Computer Sciences and Information Technologies» (CSIT'2013). — Львів: Видавництво НУ «Львівська політехніка», 2013. – Р. 50-51.

15. Kushniretska, I.I. Application of MashUp Technology for Dynamic Integration of Semi-Structured Data / I.I. Kushniretska, A.Y. Berko // Матеріали VI Міжнародної конференції молодих вчених CSE-2013. — Львів: Видавництво НУ «Львівська політехніка», 2013. — Р. 220–221.

16. Кушнірецька І.І. Використання онтологій для інтеграції слабоструктурованих даних у web-системах / І.І. Кушнірецька, А.Ю. Берко // Матеріали XI Відкритої наукової конференції ІМФН (PSC-IMFS-11) — Львів: Видавництво НУ «Львівська політехніка», 2013 С. 149-150.

17. Kushniretska I.I. Application of faceted classification in a decision support system for voter / I.I. Kushniretska, O.M. Veres // Матеріали VII Міжнародної науково-технічної конференції «Computer Sciences and Information Technologies» (CSIT'2012). — Львів: Видавництво НУ «Львівська політехніка», 2012 С. 181-182.

АНОТАЦІЇ

Кушнірецька І.І. Інформаційна технологія динамічної інтеграції слабоструктурованих даних у web-системах. – На правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 – інформаційні технології. Національний університет «Львівська політехніка» Міністерство освіти і науки України, Львів, 2016.

Дисертацію присвячено питанням розроблення інформаційної технології динамічної інтеграції слабоструктурованих даних у web-системах. Досліджено та проаналізовано принципи побудови та функціональні можливості Mashup систем динамічної інтеграції даних та здійснено концептуалізацію та моделювання роботи Mashup системи. Розроблено метод опису структури і змісту вхідних даних Mashup системи шляхом застосування онтологій для класифікації інформаційних ресурсів, що дало змогу виконувати процеси інтеграції даних із

врахуванням їх змісту. Розроблено метод формування об'єданого динамічного набору даних через застосування процедур лінгвістичного аналізу змісту запиту користувача, що забезпечило підвищення релевантності та узгодженості результатів динамічної інтеграції слабоструктурованих даних. Отримала подальший розвиток Mashup технологія динамічної інтеграції даних через запровадження додаткових можливостей опрацювання семантики інформаційних ресурсів, що дало змогу підвищити якість результатів інформаційно-пошукових web-систем. Удосконалено структурну модель системи динамічної інтеграції слабоструктурованих даних на основі технології Mashup шляхом додання функціональних компонентів визначення та опису семантики вхідних та вихідних наборів даних, що дало змогу підвищити якість результатів застосування таких систем. Розроблено та впроваджено програмні засоби динамічної інтеграції слабоструктурованих даних для підвищення характеристик якості пошуку і подання даних таких, як: збільшення точності до 6% та зниження інформаційного шуму до 6%, а також покращення узгодженості результатів інформаційних даних до 10% для процесу отримання результуючої множини вихідних даних у Mashup системі динамічної інтеграції даних.

Ключові слова: динамічна інтеграція, слабоструктуровані дані, Mashup технологія, Mashup система, метадані, якість даних.

Кушнірецька І.І. Информационная технология динамической интеграции слабоструктурированных данных в web-системах. - На правах рукописи.

Диссертация на соискание ученой степени кандидата технических наук по специальности 05.13.06 - информационные технологии. Национальный университет «Львовская политехника» Министерство образования и науки Украины, Львов, 2016.

Диссертация посвящена вопросам разработки информационной технологии динамической интеграции слабоструктурированных данных в web-системах. Исследованы и проанализированы принципы построения и функциональные возможности систем интеграции слабоструктурированных данных в web-системах, работающих используя технологию Mashup. Осуществлено концептуализацию и моделирования работы Mashup системы. Усовершенствован процесс семантического сбора и хранения данных систем при их динамической интеграции, учитывая особенности систем, работающих используя технологию Mashup динамической интеграции данных, путем разработки алгоритмов создания общей динамической структуры для множества входных информационных ресурсов с учетом их содержания. Впервые разработан метод описания структуры и содержания входных данных Mashup системы, который отличается от известных применением онтологий для классификации информационных ресурсов, что позволило учесть семантику данных в процессах их интеграции. Впервые разработан метод формирования контента объединенного динамического набора данных, который имеет общую структуру и единственный смысл путем применения процедур лингвистического анализа содержания запроса пользователя, что обеспечило повышение релевантности и согласованности результатов динамической интеграции слабоструктурированных данных.

Получила дальнейшее развитие Mashup технология динамической интеграции данных из-за введения дополнительных возможностей обработки семантики информационных ресурсов, что позволило повысить качество результатов результатов информационно-поисковых web-систем. Усовершенствована структурная модель системы динамической интеграции слабоструктурированных данных на основе технологии Mashup путем добавления функциональных компонентов определения и описания семантики входных и выходных наборов данных, что позволило что позволило повысить качество результатов применения таких систем. Разработаны и внедрены программные средства динамической интеграции слабоструктурированных данных для повышения характеристик эффективности поиска и представления данных таких, как: увеличение точности до 6% и снижение информационного шума до 6%, а также улучшение согласованности результатов информационных данных до 10% для процесса получения результирующего множества выходных данных в Mashup системе динамической интеграции данных.

Ключевые слова: динамическая интеграция, слабоструктурированные данные, Mashup технология, Mashup система, метаданные, качество данных.

Kushniretska I.I. Information technology of semi-structured data dynamic integration in the web-systems. - The manuscript.

Dissertation for a Candidate degree in Technical sciences by specialty 05.13.06 - Information Technologies. Lviv Polytechnic National University of Ministry of Education and Science of Ukraine, Lviv, 2016.

The dissertation discusses the development of information technology of semistructured data dynamic integration in the web-systems. The conceptualization and modeling of Mashup system work has been done. The process of semantic data collection and data storage of systems during their dynamic integration, given the characteristics of Mashup systems of dynamic data integration through the development of algorithms for creation of general dynamic structure for multiple incoming information resources with regard to their contents has been improved. The method of describing the structure and content of the received input information of Mashup system using ontologies for the classification of information resources allowing us to take into account the content of data during processes of its integration has been developed. The method of forming the content of a united dynamic data set that has the general structure through the application of linguistic analysis procedures of the user's request content that provided relevancy and consistency of the semistructured data dynamic integration results has been developed. The structural model of semistructured data dynamic integration Mashup system by adding functional components defining and describing the semantics of the input and output data sets that provided the quality improving of such systems application has been improved. The software for semistructured data dynamic integration to increase precision and reducing information noise and improve consistency of the information data results to 10% for the process of obtaining the resulting output data set in system that works using Mashup technology Mashup of dynamic data integration has been developed and implemented.

Keywords: dynamic integration, semistructured data, Mashup technology, Mashup system, metadata, data quality.