

## Математичне моделювання профілю Mashup системи

Жила В.Р.

Науковий керівник – к.т.н., доц. Муляр І.В.

Хмельницький національний університет

Важливим аспектом Web 2.0 є зміна пріоритетів та акцентів у використанні технологій та задоволенні потреб користувачів. Так, Web 1.0 орієнтувався на розвиток технологій комп'ютерної взаємодії, а Web 2.0 розвиває технології орієнтовані на користувачів. Нові інформаційні технології суттєво впливають на колективні способи спілкування, мислення та дій. До основних принципів Web 2.0 можна віднести невід'ємне право користувачів самостійно створювати контент, маніпулювати ним та управляти зв'язками між своїми та чужими матеріалами, отже мова йде про скоординовану активність окремих користувачів по формуванню та наповненню мережі контентом [1].

Найперспективнішими в наш час є системи інтеграції даних - них, що працюють використовуючи технологію Mashup. Mashup - це технологія для проектування вебсистем, 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 вебсистемах 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 [2]. Практичний чинник опрацювання даних різноманітної структури і форматів подання, що знаходяться в різних вебсистемах при їх динамічній інтеграції пов'язаний з вирішенням задач покращення пошуку інформації в Інтернет в сьогоднішню 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 [3].

Отже сервіс-орієнтована архітектура SOA абсолютно незалежна від мов програмування, платформ або протокольних специфікацій, за допомогою яких сервіси розробляються, а також від того, де і за допомогою чого вони розгорнуті. Практично архітектура SOA вимагає наявності не лише сервісів, але і засобів, за допомогою яких ці сервіси можуть бути виявлені і підключені незалежно від інфраструктури. SOA - це не продукт або специфікація, тому потрібне ретельне вибудовування цієї архітектури, що складається з безлічі компонентів, таких, як сервери додатків, єдине ПО, репозиторій і навіть спеціалізовані пакети централізованого управління SOA.

Формально опис Mashup системи динамічної інтеграції даних подають наступним чином:

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

де:  $DI = \{di_i \mid i = 1, N_{di}\}$  - набори вхідних даних;  $DO$  - глобальний динамічний набір вихідних даних;  $C = \{c_i \mid i = 1, N_c\}$  - множина умов інтеграції;  $T = \{t_i \mid i = 1, N_t\}$  - час транзакцій оновлення даних;  $Q = \{q_i \mid i = 1, N_q\}$  - множина запитів користувачів;  $H = \{h_i \mid i = 1, N_h\}$  - 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 [3]. Дані джерел інформаційних ресурсів Mashup систем подаються у слабоструктурованому вигляді). Слабоструктуровані дані (semistructured data) - дані, опис структури яких є частиною самих даних і має здатність змінюватися разом із їх змінами [4]. При об'єднанні таких даних необхідно брати до уваги семантичний чинник подання інформації, а це буває досить складно, особливо у web-просторі.

Композитний вебдодаток (mashup) - це вебдодаток, який використовує дані з більше ніж одного джерела для створення нового сервісу, що відображується одним графічним інтерфейсом [5]. Наприклад, комбінуючи картографічні дані Google Maps з відгуками людей про подорожі до якихось місць, можна створити унікальний веб-сервіс з функціональністю, що не передбачало жодне джерело.

Вихідні дані, орієнтовані на обробку - в основному орієнтовані на використання машинами. Це потрібно у випадку, коли необхідно використовувати дані, що підлягають подальшій обробці, наприклад, для вилучення знань. Слід зазначити, що ця категорія може, на деякій стадії, включати першу категорію, наприклад, вихідні дані у вигляді RSS можуть

бути як візуалізовані на HTML сторінці, так і використані в інших додатках для інших задач обробки даних.

Представлення вихідних даних залежить від головної мети системи. Фактично, більшість систем забезпечують багаті динамічні візуалізації колажів додатків. Але є й такі, які прагнуть до агрегації і маніпулювання даними, які можна використовувати з іншими додатками. Вихідні дані експортуються в RSS, Atom або XML інтерпретацію шляхом додавання інформації заголовка і вмісту в ньому конкретної інформації, потім інформація перетворюється в кортежі послідовностей в зазначений тип вихідної подачі інформації.

Аналізуючи діяльність Mashup систем можна виділити наступні стани діяльності (рис. 1):

Перший стан - реєстрація. Якщо реєстрація пройшла успішно, переходимо до другого стану, якщо ж ні - повертаємося знову до початку реєстрації.

Другий стан - авторизація, якщо все пройшло успішно, рухаємося далі, якщо ні - повертаємося до початку авторизації.

Третій стан - формування завдання. Якщо завдання сформовано відповідно до правил системи, рухаємося далі, якщо ні - повертаємося до початку третього стану.

Четвертий стан - формування джерел для Mashup.

П'ятий стан - пошук даних у виділених джерелах. Якщо результати пошуку задовільні - йдемо далі, якщо ні - повертаємося назад до пошуку.

Шостий стан - добування даних і перехід до наступного стану

Сьомий стан - зберігання даних у вигляді сервісу.

Восьмий стан - візуальне представлення готового Mashup.

Отже, було визначено 8 основних станів, які, зазвичай, наявні у будь-якій Mashup системі. Найважливішими станами при роботі Mashup системи є пошук даних (п'ятий стан), добування даних (шостий стан) та їх зберігання у вигляді сервісу (сьомий стан). 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 добування даних належать до процесу роботи системи із вхідними даними, а стан зберігання даних у вигляді сервісу - процесу відображення вихідних даних.

В основному джерелами Mashup застосунків є вебсистеми із даними, поданими у слабоструктурованому вигляді, адже сьогодні панівне становище в якості засобу представлення інформації в Інтернет займає мова гіпертекстової розмітки HTML, теги якої описують візуальне представлення документу, посилання, тощо, але не несуть інформації відносно семантичної структури документу.

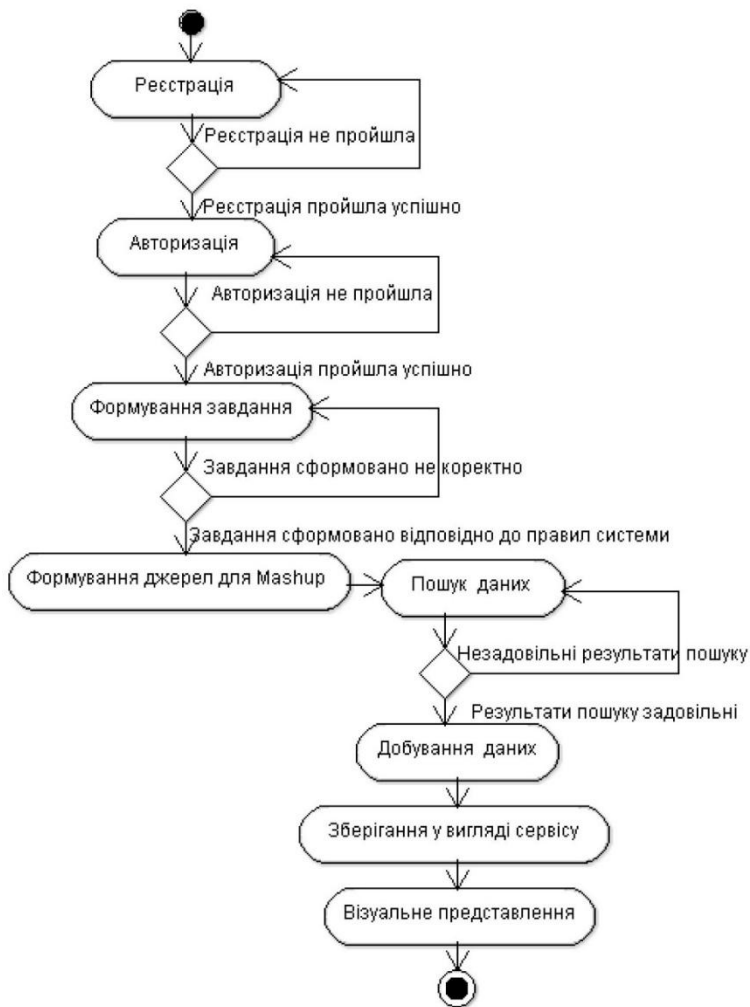


Рисунок 1 - UML-діаграма діяльності Mashup системи

З розвитком Інтернет та збільшенням обсягів збережених даних необхідність автоматизованого пошуку семантично релевантної інформації стає все більш актуальною. Mashup застосунки, здатні інтегрувати дані із декількох різних джерел є досить корисними для вирішення цього завдання. Але документи HTML мало придатні як засіб подання і отримання семантично узгодженої інформації для таких систем, так як описана засобами HTML інформація може бути охарактеризована більш як для машинного

читання, але не як машинного розуміння. Тому надзвичайно актуальною є проблема використання ефективних методів та ІТ забезпечення можливості подання семантичної структури документів в Інтернет, які дозволяли б опрацювати користувацький запит, враховуючи зміст інтегруючих даних.

В загальному основний принцип функціонування системи динамічної інтеграції слабоструктурованих даних полягає у здійсненні процесу відображення вхідних наборів даних у вихідний набір. Для опису процесу опрацювання даних Mashup системою було здійснено концептуалізацію та моделювання роботи системи динамічної інтеграції слабоструктурованих даних, використовуючи принципи STD та UML діаграми діяльності.

Показано, що процес опрацювання даних системою динамічної інтеграції слабоструктурованих даних потребує деталізації на два підпроцеси: визначення вхідних даних і формування вихідних даних. Для підвищення якості виділених підпроцесів запропоновано моделі опису структури і змісту вхідних даних та моделі формування об'єднаного динамічного набору вихідних даних.

Формалізовано процес опису структури і змісту вхідних даних, що дає змогу підвищити якість процесу отримання структурованих даних з різних інформаційних джерел для їх подальшого використання у системі динамічної інтеграції слабоструктурованих даних.

#### Перелік посилань

1. Abiteboul S. Modeling the mashup space / S. Abiteboul, O. Greenshpan, T. Milo // In *WIDM*, 2008. - P. 87-94.
2. Duda C. Ajaxsearch: crawling, indexing and searching web 2.0 applications / C. Duda, G. Frey, D. Kossmann, C. Zhou // *PVLDB*, 2008. - P. 1440-1443.
3. Муляр І.В. Аналіз підходів до структурної збірки web-додатків / І.В. Муляр, В.М Лоза, С.Б. Войнарович // Збірник наукових праць Військового інституту Київського національного університету імені Тараса Шевченка. – К.: ВІКНУ, 2017. – Вип. № 56. – С. 132- 138
4. Муляр І.В. Метод опрацювання слабоструктурованих даних у web-системах / І.В. Муляр, В.Р. Жила, Л.О. Ряба // Тези доповідей Всеукраїнської науково-практичної конференції молодих вчених, ад'юнктів, слухачів, курсантів і студентів «Молодіжна військова наука у Київському національному університеті імені Тараса Шевченка» 25квітня 2019 року. / за заг. редакцією І.В. Толока. –К. ВІКНУ, 2019. – с. 132.
5. 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.