

ВИКОРИСТАННЯ ВІКІ-ЗАСОБІВ ДЛЯ ПЕРСОНІФІКАЦІЇ ІНФОРМАЦІЙНОГО НАПОВНЕННЯ WORLD WIDE WEB

© Березко О.Л., 2008

Запропоновано використання вікі-засобів для персоніфікації інформаційного наповнення Вебу та розглянуто можливість застосування з цією метою вікі-шаблонів.

Using of wiki means for Web content personification is proposed and the possibility of the application of wiki templates for this purpose is considered.

Вступ

Сьогодні завдяки новим тенденціям, які набули поширення у World Wide Web і отримали спільну назву Веб 2.0 [2, 5], відбулося стрімке зростання обсягів інформаційного наповнення Мережі за рахунок вмісту, створеного “звичайними” користувачами у межах форумів, блогів, вікі-систем тощо. Паралельно з цим відбулась трансформація логічної структури інформаційного наповнення Вебу від множини документів (Веб-сторінок), з’єднаних між собою гіперпосиланнями, до множини інформаційних відображень користувачів (Веб-особистостей [4]), зв’язками між якими слугують особисті контакти, які документуються у Вебі в різних формах: приналежність до однієї спільноти, цитати, коментарі, спілкування на форумах, декларація стосунків у межах соціальних мереж тощо.

WWW все більше набуває ознак реальної глобальної спільноти людей, для ефективного і толерантного функціонування якої важливим є створення механізмів формування відкритої репутації, систем крос-сайтової ідентифікації та моніторингу активності її учасників.

Отже, актуальною є персоніфікація інформаційного наповнення Вебу – процес встановлення та оприлюднення відношень між інформаційними об’єктами та реальними особами, який полягає у:

- встановленні відповідностей між множиною псевдонімів, які використовуються для авторизації користувача на Веб-ресурсах різних типів та конкретною Веб-особистістю (наскрізна ідентифікація користувача на множині Веб-ресурсів);
- встановлення відношень типу “автор–твір” та “герой–твір” між інформаційними об’єктами та конкретними Веб-особистостями;
- достовірне оприлюднення встановлених зв’язків [6].

Ефективна персоніфікація неможлива без розроблення та впровадження відповідних методів та засобів. У статті розглянуто використання вікі-системи як підсистеми агрегації персоніфікованого інформаційного наповнення Вебу для подальшої обробки та оприлюднення.

Стан проблеми та способи її розв’язання

Попри переваги Вебу 2.0, величезна спільнота, яка отримала ефективний засіб масової інформації, стала джерелом багатьох проблем сучасного Вебу, зокрема:

- “інформаційного шуму” – величезної кількості неточної, неповної, поверхової та відверто неправдивої інформації, яку створюють у Вебі всі бажаючі (“культ аматора”, “антиелітність”) [2, 7];
- фактичної відсутності відповідальності (принаймні моральної) конкретного автора за свої дописи;
- відсутності ефективних засобів керування великими Веб-спільнотами.

Така орієнтованість на активного користувача Вебу та проблеми, які виникли на цьому ґрунті, вимагають нових підходів до структурування та анотування інформаційного наповнення Вебу залежно від його авторства. Сучасний Веб потребує механізмів ефективної ідентифікації, яка повинна діяти скрізь без винятку, а не лише в межах окремого Веб-сайту; відкритої системи формування і моніторингу он-лайн репутації користувача WWW – учасника глобальної спільноти тощо.

Тобто все актуальнішою стає персоніфікація інформаційного наповнення WWW, яка зокрема дає змогу:

- підвищити ефективність пошуку та моніторингу інформації про конкретних осіб;
- Веб-ресурсам отримати критичну масу користувачів;
- здійснювати ефективнішу он-лайн рекламу об'єктів авторського права певного автора доведення його ідей до широкого загалу тощо;
- підвищити рівень відповідальності автора за свої об'єкти авторського права;
- боротися з плагіатом та сприяти захисту авторських прав;
- підвищити популярність Веб-ресурсів з персоніфікованим інформаційним наповненням [6].

Сьогодні у Вебі існують засоби персоніфікації, спрямовані зокрема на:

- наскрізну ідентифікацію користувача на багатьох Веб-ресурсах (OpenID; Windows CardSpace; Gravatar тощо);
- створення відкритого соціального графу з можливістю автоматичного видобування даних (використання мікроформатів XFN та FOAF для декларації відношень та Google Social Graph API для видобування даних з Вебу) [12];
- створення семантичного інформаційного наповнення Вебу у різних форматах (XHTML, RSS, Atom тощо) з використанням мікроформату hCard (HTML vCard), який зокрема можна застосувати для опису контактної інформації осіб, що дозволяє подальше автоматичне видобування цієї інформації різноманітними засобами [13];
- пошук інформації про людей у Вебі (<http://www.pipl.com>, <http://www.zoominfo.com>, прес-портрети від Яндекса <http://news.yandex.ru/people/> тощо);
- створення каталогів відомих людей (наприклад, <http://www.nndb.com>).

Постановка задачі

Жоден з перелічених засобів не пропонує комплексного підходу до персоніфікації, який би передбачав залучення глобальної Веб-спільноти до акумуляції різнопланової інформації про Веб-особистостей у межах одного відкритого для користувачів та глобальних пошукових систем Веб-ресурсу (або системи кількох Веб-ресурсів), з метою формування відкритої репутації Веб-особистостей та крос-сайтової ідентифікації.

За таких умов актуальним є розроблення комплексних методів та засобів персоніфікації інформаційного наповнення Вебу.

Таким засобом є система персоніфікації [6], яка передбачає формування відкритого каталогу Веб-особистостей, де роль агрегаторів персоніфікованого інформаційного наповнення виконують Веб-сайти соціальних акцій [1].

Разом з тим, сьогодні у Вебі великої популярності набули Веб-сайти, які підтримують технологію вікі (wiki) [11, 16], яка дає змогу спільно редагувати вміст і організовувати моніторинг суспільної думки з різноманітних питань шляхом створення статей на відповідні теми у Вебі. На об'єктивність моніторингу впливає насамперед кількість активних редакторів конкретної вікі-спільноти.

Отже, актуальним є використання спеціалізованого вікі-ресурсу як альтернативного джерела персоніфікованих даних з їх наступним розміщенням у каталозі Веб-особистостей. На рис. 1 наведена узагальнена структура пропонованої системи – засобу персоніфікації інформаційного наповнення WWW. Альтернативним (або паралельним) варіантом застосування вікі-технології для акумуляції персоніфікованих даних є використання вже існуючих вікі-спільнот.

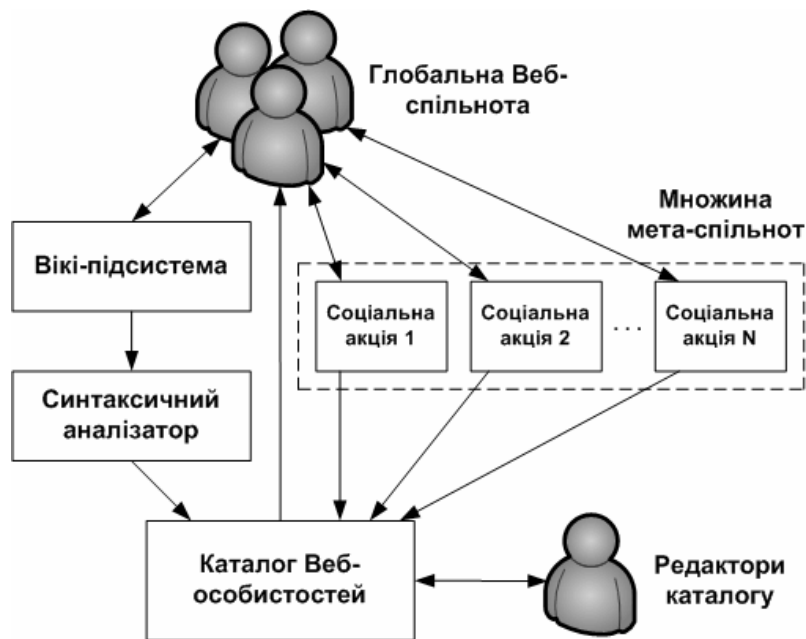


Рис. 1. Узагальнена структура системи

Результати дослідження:

1. Використання вікі-системи для акумуляції персоніфікованих даних

За Леуфом (В. Leuf) та Каннінґемом (W. Cunningham) вікі (wiki) – це сукупність зв’язаних між собою Веб-сторінок, яка може бути вільно розширена; гіпертекстова система для зберігання та модифікації інформації – база даних, де кожна сторінка може бути легко відредагована будь-яким користувачем за допомогою браузера, який підтримує HTML-форми [16].

Редагування інформаційного наповнення користувачами у вікі-системах може бути здійснене просто і швидко, часто навіть без реєстрації на вікі-ресурсі. Для редагування фрагмента вмісту користувачу достатньо перейти за відповідним посиланням (запрошення редагувати усю статтю, окремий абзац тощо), відредагувати текст запропонованими засобами та підтвердити зміни. Введений текст перетворюється на HTML-код засобами вікі.

Така легкість взаємодії користувачів перетворює вікі на ефективний засіб спільного написання тексту, автором якого можна вважати всю аудиторію користувачів ресурсу.

Іншою важливою рисою вікі є контроль за версіями статті в процесі її спільного редагування користувачами. Оскільки кожен користувач може вносити довільні зміни до довільної статті вікі, актуальною є проблема “поганих” користувачів, які з різних причин вносять неправильну, образливу та іншу некоректну інформацію у статті, знищують коректні дописи тощо [2].

Проте, як показує досвід великих вікі-спільнот (наприклад, wikipedia.org), за умов, коли стаття становить інтерес для достатньо великої кількості активних користувачів, в результаті виконання багатьох ітерацій редагування її вміст стає вичерпним, об’єктивним і неупередженим. Отже, першочерговим завданням, яке повинно бути вирішене для ефективного функціонування вікі-ресурсу (як і будь-якого соціального проекту Вебу 2.0), є формування активної аудиторії користувачів достатньо великого обсягу.

Сьогодні у Вебі є велика кількість популярних, авторитетних та багатих на інформаційне наповнення вікі-ресурсів, які використовують різні програмні системи залежно від потреб власників (наявність захищених сторінок, історії редагувань; підтримка конкретних систем управління базами даних тощо). Тематика вікі-сайтів також є різною: від туризму (наприклад, <http://wikitravel.org>) до спільного написання роману (<http://www.amillionpenguins.com>).

Створення спеціалізованого вікі-ресурсу, який виконуватиме функцію каталогу Веб-особистостей з можливістю вільного редагування вмісту всіма бажаючими (альтернативним варіантом є використання вже існуючого вікі-ресурсу) дасть змогу залучити глобальну Веб-спільноту до персоніфікації інформаційного наповнення WWW.

Згідно з інформаційною схемою каталогу Веб-особистостей [6], такий проект повинен передбачати такі типи статей:

- Веб-особистість: інформація про конкретну Веб-особистість (інформаційне відображення реальної особи) з даними про цю особу; опис її активності у Вебі, зокрема встановлення відповідності між нею та конкретними користувачами сайтів; коло інтересів та мови спілкування особи тощо.
- Користувач: інформація про конкретного користувача певного сайту, його контактні дані, мови спілкування, коло інтересів, персональна Веб-сторінка та канал новин у межах цього сайту тощо.
- Веб-сайт: інформація про сайт (як правило, це сайти спільнот різних типів), його історію, тематику, тип, мови, особливо важливі Веб-сторінки, канали новин тощо.

Для повного вікі-проекту доцільно застосувати програмну систему MediaWiki (<http://mediawiki.org>), яка є сьогодні найпопулярнішою системою такого класу. MediaWiki обрали і успішно використовують у авторитетних проектах різної тематики, зокрема wikipedia.org. Користувачам зручніше працювати з уже знайомим продуктом, що сприятиме зростанню активної аудиторії. Серед переваг цієї системи над іншими подібними програмними продуктами можна виділити ефективний механізм багатомовності, багату внутрішню мову розмітки документів, зручну систему роботи з різними версіями статей та наявність системи шаблонів [9, 11].

Оскільки статті у вікі-системах редагують усі бажаючі, немає можливості вимагати від редакторів дотримання складних правил розмітки. Тобто інформаційне наповнення сторінок вікі-сайтів важко піддається автоматичній інтерпретації програмними засобами і переважно сприймається ними як звичайний текст. З цієї ж причини існує проблема автоматичного розпізнавання зв'язків між різними інформаційними об'єктами у межах сайту [18].

У зв'язку з цим виникає потреба створення спеціалізованого синтаксичного аналізатора з метою видобування даних зі сторінок вікі-сайту (рис. 1). Для спрощення цієї задачі (надання сторінкам вікі семантичності) та уніфікації структури сторінок актуальним є створення та використання спеціальних шаблонів, які підтримує система MediaWiki.

2. Використання шаблонів для уніфікації структури вікі-сторінок та забезпечення можливості автоматичної інтерпретації їх вмісту

У системі MediaWiki інформаційне наповнення однієї сторінки може бути розміщене на інших сторінках за допомогою процесу transclusion [19]. Якщо сторінка є спеціально для цього призначена, її називають шаблоном [10, 15].

Шаблони набули великої популярності у вікі-системах і слугують стартовою точкою для створення статей на різну тематику, даючи авторам змогу використовувати існуючу типову структуру і концентруватися над створенням вмісту. Сьогодні вікі-шаблони є одним із реально діючих поєднань концепцій Вебу 2.0 та Семантичного Вебу, оскільки дають змогу створювати Веб-сторінки з семантичною розміткою зусиллями “звичайних” користувачів [2]. Серед найпопулярніших застосувань шаблонів можна виділити інформаційні та навігаційні блоки на сторінках <http://wikipedia.org> (рис. 2) [10].

Для ефективної агрегації персоніфікованих даних у межах пропонованого вікі-проекту необхідно забезпечити шаблонами такі типи статей:

Веб-особистість. Статті такого типу є статтями про реальних осіб та їх активність у Вебі.

В межах <http://wikipedia.org> існує шаблон Infobox Person [20], який забезпечує базову структуру для статей про реальних осіб. На рис. 2 зображена сторінка з цього сайту, яка містить таку статтю. Ця стаття використовує шаблон Infobox Person і є закритою до редагування звичайними користувачами (спільнота сайту вважає її вміст достатньо якісним).

Станом на жовтень 2008 року існує понад 130 варіацій шаблону Infobox Person, орієнтованих на створення статей про осіб, які займалися певним видом діяльності, посідали певну посаду, входили до певного офіційного списку тощо. Наприклад: актор, спортсмен, керівник держави тощо [21]. Серед цих шаблонів є шаблони, орієнтовані на Веб (“Інтернет-знаменитість” та “Користувач-шкідник Вікіпедії”), проте вони описують лише окремі випадки активності користувачів World Wide Web і не можуть бути використані для достатньо детального опису довільної Веб-особистості. Близькими за змістом шаблонами є “Письменник”, “Журналіст” тощо, проте вони не відображають специфіки Вебу.

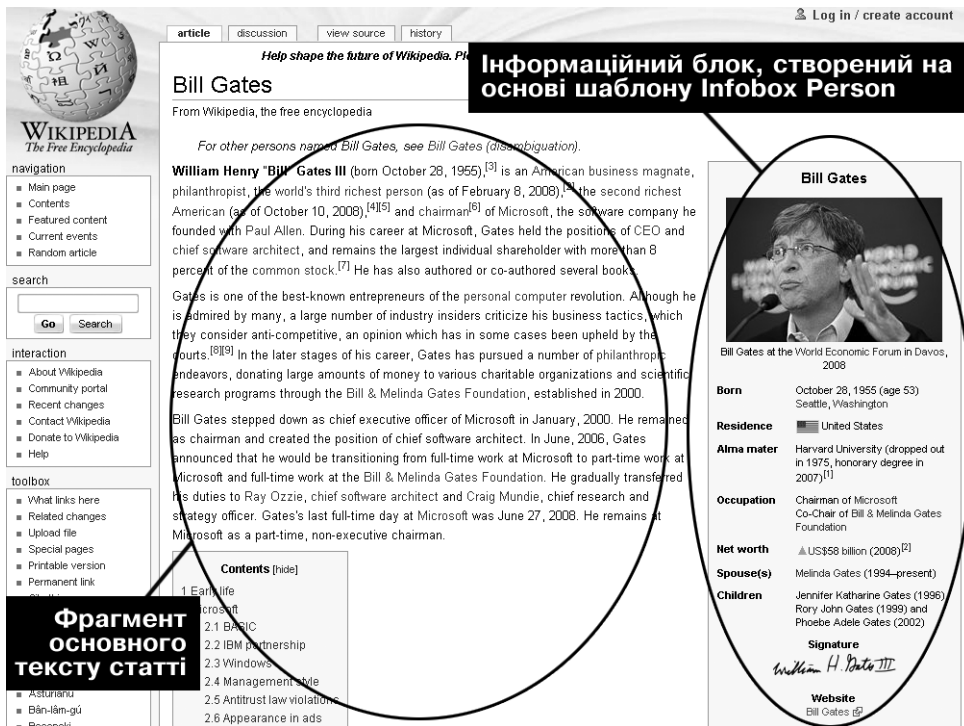


Рис. 2. Вікі-стаття про Біла Гейтса на сайті <http://wikipedia.org> (стаття належить до захищених від редагування, а отже, її структуру та інформаційне наповнення можна брати за приклад)

Важливою особливістю шаблону Infobox Person та його варіацій є підтримка мікроформату hCard [13]. Це означає, що HTML-код, згенерований за допомогою цього шаблону, має семантичну розмітку, яка може бути автоматично інтерпретована програмними засобами (синтаксичними аналізаторами).

Запропоновано шаблон “Веб-особистість” (Template:Infobox Web Personality), створений за принципом Infobox Person, який містить поля, необхідні для достатньо детального опису Веб-особистості спільнотою вікі-ресурсу. У таблиці наведено список параметрів, а на рис. 3 – відображення цього шаблону на відповідній сторінці сайту wikipedia.org, який був використаний для його тестування. Треба зазначити, що у статті наведені параметри шаблону у початковому стані. Спільнота Вікіпедії має можливість вільно редагувати створений шаблон.

Параметри шаблону “Веб-особистість”

Назва параметра	Опис параметра
1	2
name	Ім'я особи (як правило, збігається з назвою статті).
image	Назва зображення (фото особи-прототипа [4] Веб-особистості), наприклад, “pict.jpg”. З метою запобігання порушення авторських прав необхідно застерегти користувачів вікі-ресурсу від розміщення на сайті зображень у випадках, коли це не може бути потрактоване як доброзрядне використання інформаційного наповнення [22].
imagesize	Розмір зображення у пікселях. 200px завширшки за замовчуванням
alt	Атрибут “alt”, який дає змогу описати зображення для тих, хто з різних причин його не бачить.
caption	Заголовок зображення. У разі, якщо це фотографія, бажано вказати дату її створення.
birthname	Ім'я, дане особі-прототипу при народженні.
birthdate	Дата народження (можливі різні варіанти запису дати).
birthplace	Місце народження.
deathdate	Дата смерті (можливі різні варіанти запису дати).
deathplace	Місце смерті.
nationality	Національність особи-прототипу.
citizenship	Громадянство особи-прототипу.

1	2
education	Освіта особи-прототипу. Мається на увазі освітній напрямок. Цей параметр дає змогу точніше визначити тематику Веб-особистості.
alma_mater	Вищий навчальний заклад, студентом або випускником якого є особа-прототип. Цей параметр полегшує пошук інформаційного наповнення, пов'язаного з Веб-особистістю (у межах соціальних мереж, спільнот випускників тощо.)
occupation	Професія особи-прототипу (значення може бути відмінне від education).
nickname	Відомі псевдоніми Веб-особистості з посиланнями на відповідні сторінки у межах вікі-проекту.
subject	Перелік відомих тематичних категорій дописів Веб-особистості.
language	Перелік відомих мов дописів Веб-особистості.
notableworks	Найбільш значимі твори особи-прототипу, розміщені онлайн.
awards	Нагороди особи-прототипу, пов'язані з Вебом (наприклад, UA Webstar).
website_community	Перелік сайтів-спільнот, де є дописи Веб-особистості (форуми, колективні блоги, соціальні мережі різного типу тощо).
website_blogs	Перелік особистих блогів.
website_homepage	Перелік домашніх сторінок (статичних сайтів-візиток).
website_bookmarks	Перелік публічних сторінок в межах сервісів соціальних закладок (http://del.icio.us , http://stumbleupon.com , http://blinklist.com тощо).
website	Офіційний сайт особи-прототипу. У випадку наявності кількох таких сайтів, необхідно обрати серед них найважливіший.

Infobox Web Personality	
[[Image:{{{image}}} 200px]]	
{{{caption}}}	
Background information	
Born	{{{birthname}}} {{{birthdate}}} {{{birthplace}}}
Died	{{{deathdate}}} {{{deathplace}}}
Nationality	{{{nationality}}}
Citizenship	{{{citizenship}}}
Education	{{{education}}}
Alma mater	{{{alma_mater}}}
Occupation	{{{occupation}}}
Web activity	
Nicknames	{{{nickname}}}
Subjects	{{{subject}}}
Languages	{{{language}}}
Notable work(s)	{{{notableworks}}}
Notable award(s)	{{{awards}}}
Websites	
Community-driven	{{{website_community}}}
Personal blogs	{{{website_blogs}}}
Personal home page	{{{website_homepage}}}
Social bookmarks	{{{website_bookmarks}}}
[[{{{website}}} Official website]]	

Рис. 3. Відображення розробленого автором шаблону "Веб-особистість" у Вікіпедії – http://en.wikipedia.org/wiki/Template:Infobox_Web_Personality

Користувач. Статті такого типу мають цінність насамперед за умови відсутності інформації про відповідність множини користувачів певній реальній особі. Згідно з інформаційною схемою каталогу Веб-особистостей [6] шаблон “Користувач Веб-сайту” повинен містити такі базові поля:

- URL Веб-сайту (обов’язкове для заповнення поле);
- Псевдонім у межах сайту (обов’язкове для заповнення поле);
- Аватар – поточне зображення, яке позначає користувача у межах даного сайту, попередні зображення (як і старі значення інших параметрів) рекомендується розміщувати в основному тексті статті;
- Дата реєстрації (створення облікового запису);
- Дата знищення (блокування) облікового запису;
- E-mail, який відповідає обліковому запису користувача;
- URL блогу користувача в межах сайту;
- URL персональної сторінки користувача в межах сайту;
- Роль в межах сайту (модератор тощо);
- Авторитетність (карма) в межах сайту;
- Мови спілкування користувача в межах сайту;
- Місцезнаходження користувача;
- Теми, на які пише користувач в межах сайту;
- Веб-особистість, до якої належить користувач.

Веб-сайт. Такі статті повинні містити інформацію, необхідну для достатньо чіткого розуміння типу та тематики сайту, що є важливим для якомога повнішого опису Веб-особистостей, які мають до нього відношення.

У межах <http://wikipedia.org> існує шаблон Infobox Website [17], який дає змогу достатньо повно описати довільний Веб-сайт. Пропонується використати цей шаблон. Доцільним також є створення категорій [8] “Користувачі сайту” та “Веб-особистості, пов’язані з сайтом” для кожного з сайтів, сторінка про які буде створена.

Висновок

Запропоновано використання вікі-засобів для персоніфікації інформаційного наповнення Вебу, зокрема створення спеціалізованого вікі-ресурсу та використання вже існуючих вікі-спільнот для акумуляції персоніфікованих даних для подальшого їх розміщення у каталозі Веб-особистостей. Також запропоновано використання вікі-шаблонів для забезпечення необхідної уніфікованості структури інформаційного наповнення статей, надання їм семантичності та забезпечення можливості їх автоматичної інтерпретації.

1. Березко О.Л., Пелецишин А.М. Соціальні акції в середовищі Веб 2.0. // 12-й Міжнародний молодіжний форум “Радіоелектроніка і молодь в XXI ст.”: Зб. матеріалів форуму Ч.2. – Харків: ХНУРЕ, 2008. с.499. 2. Березко О., Пелецишин А. Особливості сучасних концепцій розвитку Вебу // Комп’ютерні науки та інженерія, матеріали 1-ї Міжнародної конференції молодих науковців CSE-2006. – Львів: Вид-во Нац. ун-ту “Львівська політехніка”, 2006. – С.9–12. 3. Березко О., Пелецишин А. WWW як соціальна мережа // Комп’ютерні науки та інженерія: Матеріали 2-ї Міжнародної конференції молодих науковців CSE-2007. Львів: Вид-во Нац. ун-ту “Львівська політехніка”, 2007. – С.30–32. 4. Пелецишин А.М., Березко О.Л. Формальні аспекти ідентифікації особистості у WWW. // Вісник Нац. ун-ту “Львівська політехніка”, 2005. – №546. – С.126–131. 5. Пелецишин А.М. Позичування сайтів у глобальному середовищі: Монографія. – Львів: Вид-во Нац. ун-ту “Львівська політехніка”, 2007. – 260 с. 6. Березко О. Каталог Веб-особистостей – засіб

персоніфікації інформаційного наповнення глобальної системи World Wide Web // *Комп'ютерні науки та інформаційні технології: Матеріали третьої Міжнародної конференції CSIT'2008*. – Львів: Вид-во ПП “Вежа і Ко”, 2008. – С.204–207. 7. Carr N. *The amorality of Web 2.0. Rough type, October 3, 2005*. – [Електронний ресурс]. – http://www.roughtype.com/archives/2005/10/the_amorality_o.php. 8. *Category: People infobox templates*. – [Електронний ресурс]. – http://en.wikipedia.org/wiki/Category:People_infobox_templates. 9. Daniel J. Barrett. *MediaWiki. O'Reilly Media, 2008*. 10. Di Iorio A., Vitali F., Zacchiroli S., *Wiki content templating. // Proceeding of the 17th international conference on World Wide Web, April 21-25, 2008, Beijing, China* <http://www2008.org/papers/pdf/p615-diiorioA.pdf>. 11. Ebersbach A., Glaser M., Heigl R., Warta A. *Wiki. Web collaboration. Second edition. Springer, 2008*. 12. *Google Social Graph API*. – [Електронний ресурс]. – <http://code.google.com/apis/socialgraph/>. 13. *hCard*. – [Електронний ресурс].- <http://microformats.org/wiki/hcard>. 14. *Help:Category*. – [Електронний ресурс]. – <http://meta.wikimedia.org/wiki/Help:Category>. 15. *Help:Template*. – [Електронний ресурс].- <http://meta.wikimedia.org/wiki/Help:Template>. 16. Leuf B., Cunningham W., *The Wiki way: quick collaboration on the Web. // Addison-Wesley Longman Publishing Co., Inc., Boston, MA, 2001*. 17. *Template:Infobox Website*.- [Електронний ресурс].- [http://en.wikipedia.org/wiki/Template: Infobox_Website](http://en.wikipedia.org/wiki/Template:Infobox_Website). 18. Völkel M., Krötzsch M., Vrandečić D., Haller H., Studer R., *Semantic Wikipedia. // Proceedings of the 15th international conference on World Wide Web, May 23-26, 2006, Edinburgh, Scotland* – <http://www2006.org/programme/files/pdf/4039.pdf>. 19. *Wikipedia:Transclusion*. – [Електронний ресурс]. – <http://en.wikipedia.org/wiki/Wikipedia:Transclusion>. 20. *Template:Infobox Person/doc*. – [Електронний ресурс]. – [http://en.wikipedia.org/wiki/Template: Infobox_Person/doc](http://en.wikipedia.org/wiki/Template:Infobox_Person/doc). 21. *Category: People infobox templates*. – [Електронний ресурс]. – http://en.wikipedia.org/wiki/Category:People_infobox_templates. 22. *Wikipedia:Non-free content*. – [Електронний ресурс]. – <http://en.wikipedia.org/wiki/Wikipedia:NONFREE>