

## КАТАЛОГ ВЕБ-ОСОБИСТОСТЕЙ

© Березко О.Л., 2009

**Розглянуто проблему персоніфікації інформаційного наповнення WWW та запропоновані засоби такої персоніфікації, зокрема каталог Веб-особистостей.**

**The problem of Web content personification is considered and the means of such personification are proposed, Web personalities directory particularly.**

**Вступ.** Протягом останніх років у Вебі можна виділити тенденції, які стали визначальними. Їх основою є нові підходи до розробки, експлуатації та організації взаємодії між Веб-ресурсами. Ці підходи отримали спільну назву “Веб 2.0” і позиціонуються багатьма дослідниками як новий якісний крок у розвитку Вебу [4].

Основою Вебу 2.0 є активність користувачів, а саме досягнення ними креативних та комунікативних цілей [2]. Веб-контент, створений “звичайними” користувачами (user generated content) як поодинці, так і колаборативно в межах Веб-спільнот, не лише став рівноправним за авторитетністю та популярністю з контентом, створеним професійними Веб-журналістами, а й часто витісняє останній (протистояння Вікіпедії та платних енциклопедій, особистих блогів та сайтів авторитетних видань тощо).

Такі зміни в джерелах інформаційного наповнення Вебу та правилах його створення стали причиною трансформації логічної структури мережі Інтернет від множини Веб-сторінок, з’єднаних між собою гіперпосиланнями, до множини Веб-особистостей, зв’язками між якими слугують особисті контакти, цитати, приналежність до однієї спільноти тощо (соціальний граф) [3], [10]. За таких умов актуальною є персоніфікація інформаційного наповнення глобальної системи World Wide Web, яка неможлива без розробки та впровадження відповідних методів та засобів.

**Огляд літературних джерел та стан проблеми.** Попри значну кількість позитивних сторін Вебу 2.0, величезна спільнота, яка отримала ефективний засіб масової інформації, стала джерелом деяких проблем сучасного Вебу, серед яких можна виділити:

- “інформаційний шум” – величезна кількість неточної, неповної, поверхової та відверто неправдивої інформації, яку створюють у Вебі всі бажаючі (“культ аматора”, “анти-елітність”) [2, 7];
- фактичну відсутність відповідальності конкретного автора за свої дописування;
- відсутність 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.

Очевидно, що Веб все більше набуває ознак реальної глобальної спільноти людей, у якій неможливо обійтися без механізмів ефективної ідентифікації, яка повинна діяти скрізь без винятку, а не лише в межах окремого Веб-сайта; відкритої системи формування і моніторингу он-лайн репутації користувача WWW – учасника глобальної спільноти тощо.

Як наслідок, все актуальнішою стає персоніфікація інформаційного наповнення WWW. Сьогодні у Вебі існує низка засобів персоніфікації, спрямованих 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) [10];

- пошук інформації про людей у Вебі (<http://www.pipl.com>, <http://www.zoominfo.com>, прес-портрети від Яндекс <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 для акумуляції різноманітної інформації про Веб-особистостей в межах одного відкритого для пошуківців Веб-ресурсу (або системи кількох Веб-ресурсів), з метою формування відкритої репутації Веб-особистостей.

За таких умов актуальною є розробка комплексних методів та засобів персоніфікації інформаційного наповнення Вебу.

### Результати дослідження:

**1. Персоніфікація інформаційного наповнення Вебу.** Веб-особистість – це множина даних, які стосуються конкретної особи і можуть стосуватись будь-якої категорії або будь-якої комбінації категорій даних, доступних в мережі Інтернет [4].

Якщо розглядати всі дані, доступні в мережі Інтернет, як деяку множину інформаційних об'єктів  $U$  (елементи Веб-контенту: тексти, зображення, відеофайли тощо), конкретна Веб-особистість  $Wp$  є підмножиною цієї множини:

$$Wp \subset U \quad (1)$$

Під *персоніфікацією інформаційного наповнення WWW* розуміємо 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 (це дослідження спрямоване на суб'єктів, а не об'єктів Вебу: ми не розглядаємо відношення типу “герой-твір”);
- достовірне оприлюднення встановлених зв'язків.

Персоніфікація інформаційного наповнення Вебу зокрема дозволяє:

– *підвищити ефективність пошуку та моніторингу інформації про конкретних осіб у Вебі*

Сьогодні у Вебі є велика кількість персоніфікованої інформації, зокрема доступної у соціальних мережах типу [facebook.com](http://facebook.com), [myspace.com](http://myspace.com), [bebo.com](http://bebo.com) тощо. Проте, інформація, яка розташована на конкретному сервісі, доступна лише користувачам цього сервісу і, відповідно, не може бути проіндексована глобальними пошуковими системами (належить до “прихованого Вебу” [13]).

Така ситуація ускладнює пошук та моніторинг інформації про конкретних осіб у Вебі, що негативно впливає на ефективність пошуку партнерів, ділових контактів, друзів тощо;

– *отримати Веб-ресурсам критичну масу користувачів*

Станом на липень 2008 року за приблизними оцінками у Вебі:

- понад 1,4 млрд. користувачів;
- понад 172 млн. діючих сайтів;
- понад 27 млрд. сторінок, проіндексованих глобальними пошуковими системами (сюди не входять сторінки “прихованого Вебу”) [11], [12], [15].

Такі дані свідчать про перенасиченість Вебу інформацією, що не дає змоги переважній більшості Веб-сайтів отримати свою аудиторію через жорстку конкуренцію. Особливо це стосується різноманітних Веб-сервісів, що передбачають реєстрацію, в основі яких лежать соціальні механізми Вебу 2.0, і які не можуть ефективно функціонувати без критичної маси активних користувачів.

Засоби персоніфікації дозволяють користувачам здійснювати моніторинг активності інших користувачів, які їх цікавлять (авторитетних партнерів, друзів тощо) в межах широкого кола Веб-

ресурсів, до яких вони отримують стимул приєднатися. Це дозволить багатьом справді цікавим і корисним Веб-сайтам отримати свою активну аудиторію.

**– здійснювати більш ефективну он-лайн рекламу об'єктів авторського права певного автора і самого автора, доведення його ідей до широкого загалу тощо**

Засоби Вебу 2.0 дозволяють всім бажаним активно рекламувати свої об'єкти авторського права, послуги, суспільну позицію тощо. Сьогодні одним з найефективніших методів цього є створення блогів (як самостійних, так і в межах відповідних Веб-сайтів) з професійними високоякісними дописуваннями, допомога іншим користувачам на авторитетних тематичних форумах тощо. Засоби персоніфікації дозволять користувачу акумулювати інформацію про себе та місцезоташування своїх дописувань для ефективнішого їх позиціонування у Вебі.

**– підвищити рівень відповідальності автора за свої об'єкти авторського права**

Сьогодні важко встановити авторство значної частини Веб-контенту. Особливо це стосується різноманітних засобів Вебу 2.0. Як наслідок, важко притягнути до відповідальності (принаймні моральної) користувачів за: “інформаційний шум”; Веб-контент, що зазіхає на честь та гідність інших осіб; спам та інший небажаний контент.

Персоніфікація Веб-контенту, зокрема відкрита репутація користувачів WWW, змусить авторів створювати якісніший контент, відмовитись від ідеологічного екстремізму тощо. Це, своєю чергою, сприятиме підвищенню якості інформаційного наповнення усього Вебу.

Іншим позитивним наслідком персоніфікації є підвищення керованості Веб-спільнот: ефективніша модерація дописів, коментарів, боротьба з небажаними дописами (флейм, флуд) тощо.

Яскравим прикладом переходу від анонімності до персоніфікації може слугувати Веб-система Knowl від компанії Google – аналог Вікіпедії, яка не передбачає анонімного створення контенту, на відміну від останньої [9].

**– боротися з плагіатом та сприяти здійсненню захисту авторських прав**

Персоніфікований Веб-контент передбачає можливість точного встановлення першоджерела інформації та її автора, що дає змогу ефективніше виявляти плагіат та неправомірне копіювання.

**– підвищити популярність Веб-ресурсів з персоніфікованим інформаційним наповненням**

Персоніфікація передбачає появу множини посилань, які встановлюють відповідність між частинами інформаційного наповнення та конкретними особами. Ці посилання можуть стати джерелом значного обсягу трафіку на відповідні ресурси. Окрім того, дані про конкретних осіб-авторів виконують функцію метаданих і сприяють появі нових шляхів до персоніфікованого контенту, зокрема через глобальні пошукові системи (наприклад, через запит, в якому фігурує прізвище автора).

## **2. Каталог Веб-особистостей**

Отже, ефективні комплексні засоби персоніфікації інформаційного наповнення WWW повинні передбачати можливість використання соціальних механізмів Вебу для акумулювання даних та зручний для користувачів і відкритий для глобальних пошукових систем спосіб їх оприлюднення.

На рис.1 наведена узагальнена структура пропонованої системи – засобу персоніфікації інформаційного наповнення WWW.

Джерелом даних у системі є глобальна Веб-спільнота, яка бере участь у соціальних акціях у Вебі (належить до мета-спільнот) [1] та здійснює колаборативне редагування вмісту спеціалізованого вікі-проекту [8]. Дані з цих Веб-ресурсів автоматично передаються у зведену базу даних каталогу Веб-особистостей, де їх обробляють та агрегують, зокрема за участю редакторів каталогу (користувачів з відповідними правами). Каталог Веб-особистостей є базовою компонентою системи. Дані в каталозі вільно доступні користувачам WWW та глобальним пошуковим системам.

У каталозі Веб-особистості класифікуються за такими основними ознаками, як: мова, розташування, тематична категорія та роль. Ці ознаки напряму залежать від основних ознак сайтів (іноді – окремих Веб-сторінок), де відповідні користувачі WWW досягають свої креативні та комунікативні цілі [5], [14].

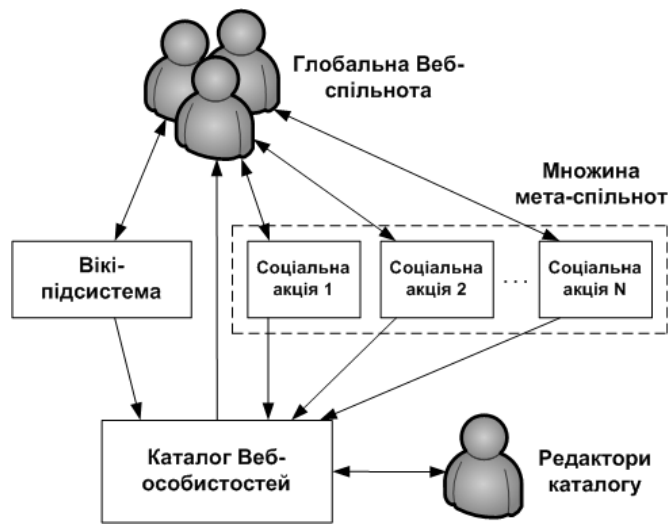


Рис. 1. Узагальнена структура системи

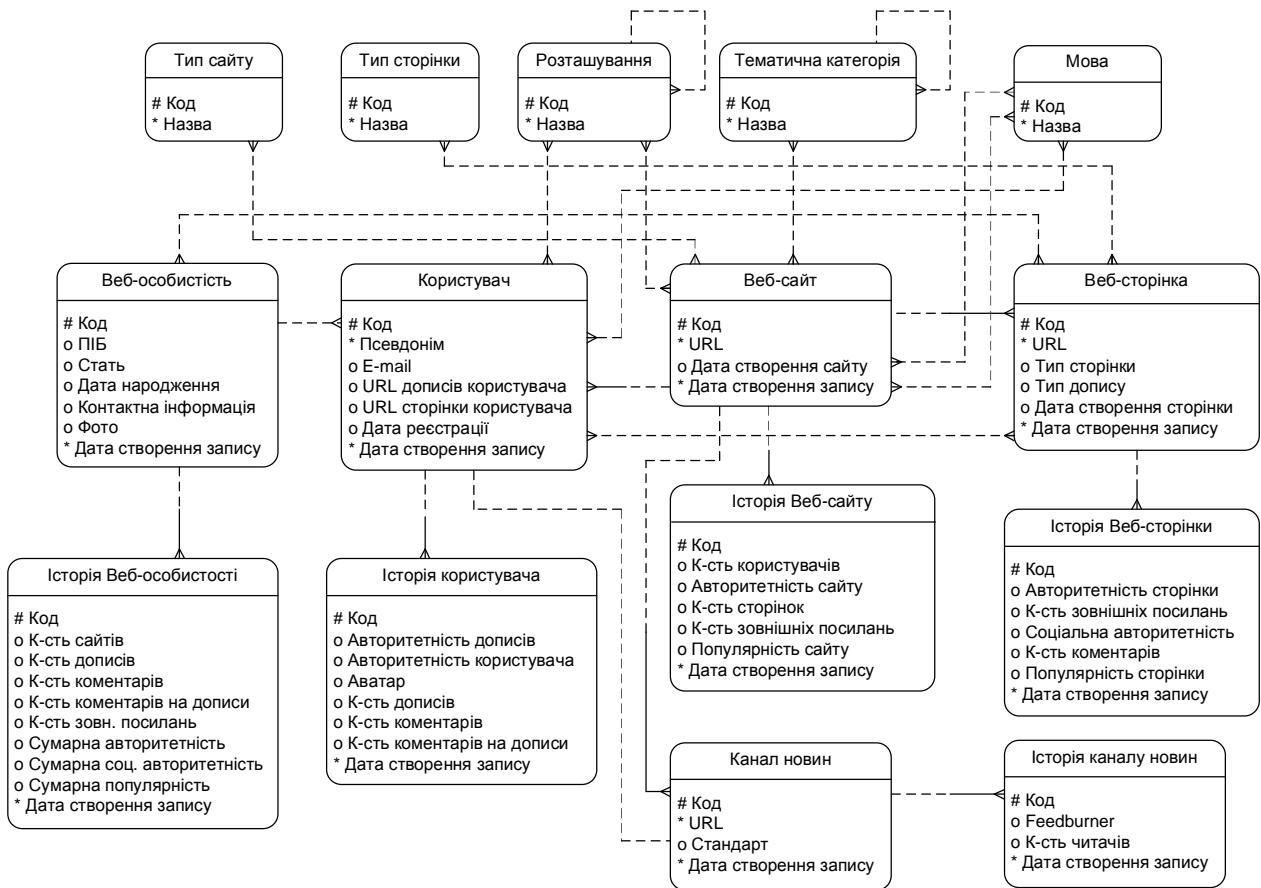


Рис. 2. Інформаційна схема каталогу Веб-особистостей

На рис.2 наведена інформаційна схема каталогу, виконана в нотації Баркера [6]. Коментарі до деяких відношень та зв'язків запропонованої схеми:

**Веб-особистість.** Це відношення містить дані про конкретну Веб-особистість (наприклад, Олександр Березко). Зв'язки з іншими відношеннями:

- **Користувач.** “Один до багатьох”, обидва зв'язки необов'язкові. До Веб-особистості можуть належати багато користувачів різних сайтів (Веб-особистість може мати багато псевдонімів).

- Веб-сторінка. “Багато до багатьох”, обидва зв’язки необов’язкові. Веб-особистість може бути автором дописувань на багатьох сторінках; сторінка може містити частини контенту авторства різних осіб (наприклад, дописування на форумі та коментарі до нього).

Користувач. Це відношення містить дані про певного зареєстрованого користувача конкретного сайта (наприклад, admin з сайта <http://pidkamin.ridne.net>). Зв’язки з іншими відношеннями:

- Веб-сторінка. Аналогічно до зв’язку “Веб-особистість – Веб-сторінка”. Іноді можливо встановити авторство лише на рівні користувача сайта, а не Веб-особистості.

**Висновки.** У статті розглянута проблема персоніфікації інформаційного наповнення глобальної системи World Wide Web, виділені потенційні корисні наслідки цього процесу. Запропоновано засоби персоніфікації Веб-контенту, зокрема каталог Веб-особистостей та наведено схему даних цього каталогу.

1. Березко О.Л., Пелецишин А.М. Соціальні акції в середовищі Веб 2.0. 12-ий 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 Міжн. конф. молодих науковців CSE-2006, Львів, Видавництво Національного університету “Львівська політехніка”, 2006. – С.9–12. 3. Березко О., Пелецишин А. WWW як соціальна мережа // Комп’ютерні науки та інженерія, матеріали 2-ї Міжн. конф. молодих науковців CSE-2007. Львів, Видавництво Національного університету “Львівська політехніка”, 2007. – С.30–32. 4. Пелецишин А.М., Березко О.Л. Формальні аспекти ідентифікації особистості у WWW. Комп’ютерні системи та мережі // Вісник Нац. ун-ту “Львівська політехніка” – Львів, Видавництво Нац. ун-ту “Львівська політехніка”. –2005. – №546. – С.126–131. 5. Пелецишин А.М. Позиціонування сайтів у глобальному середовищі: Монографія. – Львів: Видавництво Нац. ун-ту “Львівська політехніка”, 2007. – 260 с. 6. Barker R., "CASE Method - Entity Relationships Modelling", Oracle Corporation UK Limited, Addison-Wesley Publishing Company, 1990. 7. Carr N. The amorality of Web 2.0. Rough type, October 3, 2005 - [http://www.roughtype.com/archives/2005/10/the\\_amorality\\_o.php](http://www.roughtype.com/archives/2005/10/the_amorality_o.php) 8. Emigh W., Herring S.C. Collaborative Authoring on the Web: A Genre Analysis of Online Encyclopedias. Proceedings of the Thirty-Eighth Hawai’i International Conference on System Sciences (HICSS-38). Los Alamitos: IEEE Press, 2005 - <http://ella.slis.indiana.edu/~herring/wiki.pdf> 9. Google Knol. Introduction to Knoll. – <http://knol.google.com/k/knol/knol/Help> 10. Google Social Graph API – <http://code.google.com/apis/socialgraph/> 11. Internet Usage Statistics. Marketing Report. – Miniwatts Marketing Group. <http://internetworldstats.com/stats.htm> 12. June 2008 Web Server Survey. Technical Report. – Netcraft Ltd. [http://news.netcraft.com/archives/2008/06/22/june\\_2008\\_web\\_server\\_survey.html](http://news.netcraft.com/archives/2008/06/22/june_2008_web_server_survey.html) 13. Michael K. Bergman, Ann Arbor. White paper: The Deep Web: Surfacing Hidden Value. MI: Scholarly Publishing Office, University of Michigan, University Library vol. 7, no. 1, August, 2001. - <http://hdl.handle.net/2027/spo.3336451.0007.104> 14. Susan Dumais, Hao Chen, Hierarchical classification of Web content, Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval, p.256-263, July 24-28, 2000, Athens, Greece 15. The Size of the World Wide Web. – ILK Research Group. Tilburg university. <http://www.worldwidewebsite.com/>