

УДК 004.738.52

С.С. Щербак

Харківський національний університет радіоелектроніки, Харків

ТЕХНОЛОГІЯ МІКРОФОРМАТІВ ЯК СТРАТЕГІЯ ПІДВИЩЕННЯ ЯКОСТІ ІНФОРМАЦІЙНОГО ПОШУКУ

Розглянуто технологію мікроформатів в контексті впровадження семантичних розміток до веб-сторінок програмних прикладань та наводиться огляд існуючих мікроформатів, обробка яких підтримується сучасними пошуковими системами. Запропоновані моделі взаємодії користувачів з мікроформатами, які адаптовані до використання в Semantic Web, використовують засоби локального та розподіленого пошуку, та наводяться рекомендації щодо їх застосування.

Ключові слова: мікроформати, інформаційний пошук, онтології, пошукові механізми, семантика.

Вступ.

Роль мікроформатів у сучасному Веб

Мікроформати – це один з видів семантичної розмітки, які використовуються для додавання опису знань та даних, розташованих на веб-сторінках та інших типах електронних документів. З одного боку, семантичні розмітки, такі як RDF, OWL існують достатньо давно та мають широкі можливості для опису знань, але складність розміток та необхідність мати спеціалізовані знання для їх застосування привели до того, що веб-розробники не використовують їх. Така непопулярність не дозволяє застосовувати методи штучного інтелекту для організації семантичного пошуку, тому розробка засобів та технологій спрощення опису знань на веб-сторінках є актуальною. З іншого боку, в видавництвах багато років використовуються технології мікроформатів, подібних до vCard, які значним чином полегшують роботу з матеріалом, який планується опублікувати. Основою цих технологій є використання міток та їх груп, для визначення загальнозживаного представлення типової інформації, наприклад, про авторів, підручники, організації тощо. За кожною міткою, яка використовується, закріплене деяке загальнозживане значення, що може бути використане при машинній обробці.

Огляд існуючих мікроформатів

На сьогоднішній день розроблено досить багато різних мікроформатів, які можна використовувати при розробці веб-сторінок, та обробку яких вже підтримують такі пошукові машини як Яндекс та Гугл.

Найбільшу підтримку пошукових машин отримав мікроформат hCard. Цей мікроформат базується на vCard та використовується для опису людей, підприємств, організацій та публікації контактної інформації в структурованому вигляді. Крім того, треба зазначити, що цей мікроформат використовується в інших мікроформатах, таких як hCalendar та Geo.

hCalendar – це мікроформат для опису подій, часу їх проведення і таке інше.

hReview – мікроформат для розмітки відгуків наприклад, на події або матеріали.

hProduct – уніфікована розмітка товарів. Наприклад, за допомогою цього формату можна визначити ціну товару, його наявність, відгуки покупців.

hRecipe – уніфікована розмітка рецептів по приготуванню їжі.

Geo – мікроформат, який використовується для розмітки географічних координат на веб-сторінках.

XFN – мікроформат для розмітки соціальних взаємовідносин.

FOAF – ще одна розмітка для опису людей, груп та взаємовідносин між ними.

DOAP – FOAF-подібна розмітка для опису відкритих програмних проектів.

RDFa – це версія RDF, адаптована до використання на веб-сторінках.

Приклади мікроформатів

Розглянемо деякі приклади мікроформатів та відзначимо основні результати, які можна отримати при їх використанні.

Приклад 1. Деяка компанія, яка займається торгівлею меблів через Інтернет, має електронний магазин та хотіла б чітко вказати пошуковій машині, коли працює її офіс, де він розташований та з/до якого часу він працює.

Для визначення подібної інформації на веб-сторінці ми будемо використовувати hCard (рис. 1).

Перший рядок розмітки (див. рис. 1) дає вказівку програмному забезпеченню про використання мікроформату hCard для опису підприємства. На 2-4 рядках ми визначили назву підприємства та адресу її сайту в Інтернеті.

На 9-15 рядках визначимо телефон та факс офісу підприємства. Крім того, рядки 16-25 визначають адреси електронної пошти підприємства та час роботи офісу.

Приклад 2. Опис співробітника деякого підприємства (рис. 2), в якому представлена інформація про прізвище, ім'я співробітника та його мобільний телефон.

```

1 <div class="vcard">
2   <div>
3     Частное предприятие <a class="fn org url" href="http://щербак.net">Щербак Корпорейшн</a>
4   </div>
5   <div class="adr">
6     <span class="locality">г. Харьков</span>,
7     <span class="street-address">ул. Бассейная, д. 122</span> (район м. Центральный рынок )
8   </div>
9   <div class="tel">Телефон:
10    <abbr class="value" title="+380 (57) xxx-xx-xx">xxx-xx-xx</abbr>
11  </div>
12  <div class="tel">
13    <abbr class="type" title="fax">Факс:</abbr>
14    <abbr class="value" title="+380 (57) xxx-xx-xx">xxx-xx-xx</abbr>
15  </div>
16  <div>Електронная почта по вопросам трудоустройства :
17    <a class="email" href="mailto:hr@щербак.net">
18      hr@щербак.net
19    </a>
20  </div>
21  <div>Електронная почта по вопросам закупок и продаж:
22    <a class="email" href="mailto:sail@щербак.net">
23      sail@щербак.net
24    </a>
25  </div>
26  <div>
27    Время работы: <span class="workhours">пн-пт 9:00-18:00</span>
28  </div>
29  <span class="category">
30    <span class="value-title" title="Мебельные магазины"></span>
31  </span>
32 </div>

```

Рис. 1. Приклад використання мікроформату hCard для опису підприємства

```

1 <div id="myhcard" class="vcard">
2   <span class="fn n">
3     <span class="given-name">Сергей</span>
4     <span class="family-name">Щербак</span>
5     <span class="nickname">shcherbak</span>
6   </span>
7   <span class="tel">
8     <span class="type">mobile</span>:
9     <span class="value">+380 (67) xxx-xx-xx</span>
10  </span>
11 </div>

```

Рис. 2. Приклад використання мікроформату hCard для опису співробітника підприємства

Для правильної обробки мікроформатів необхідно в блоці HEAD веб сторінки додати тип профілю мікроформату, який використовується. Це можна зробити шляхом додавання до блоку HEAD, наприклад, для профілю hCard, наступного рядка:

```
<link rel="profile" href="http://microformats.org/profile/hcard">.
```

Далі, в залежності від конкретної задачі можуть використовуватися різні технології вилучення контексту мікроформатів. Найбільш ефективними є на сьогодні технології GRDDL (Gleaning Resource Descriptions from Dialects of Languages), XSLT (eXtensible Stylesheet Language Transformations), Javascript з DOM (Document Object Model). Розглянемо більш детально ці технології в контексті вирішення типових задач користувача.

Моделі взаємодії користувачів з мікроформатами

Мікроформати можуть бути використані не тільки пошуковими машинами для покращення інформаційного пошуку, але й для організації різного роду

сервісів. Інформація про підприємства складає основу для створення каталогу підприємств з можливостями організації інтелектуального пошуку; інформація про відгуки може слугувати основою для створення різного роду рейтингів; Гео – координати дають змогу застосовувати карти місцевості на своїх веб-сторінках з можливостями відтворення місцевості, де, наприклад, буде відбуватися деяка подія. Розглянемо рішення, які можуть бути застосовані для організації взаємодії користувачів з мікроформатами.

Рішення 1. Мікроформати, як елемент інформаційної системи можуть слугувати джерелами даних, які можуть бути оброблені процесором XSLT та завантажені до бази даних трійок RDF (RDF БД), як це показано на рис. 3.

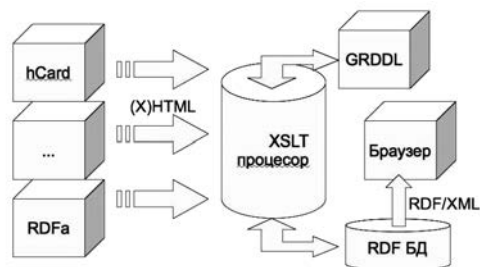


Рис. 3. Модель взаємодії користувача з мікроформатами на базі GRDDL

Обробка XSLT процесором відбувається на основі правил шаблонів GRDDL.

Дані, збережені до RDF БД, вилучаються за запитом користувача при необхідності побудови веб-сторінки за допомогою технології Sparql, що дозволяє завантажувати дані до браузера в форматі RDF/XML та відобразити їх засобами мов Javascript та (X)HTML.

Запропонована модель взаємодії користувача з мікроформатами відноситься до моделей з проміжним збереженням даних та може застосовуватися у випадках, коли необхідно забезпечити централізований доступ багатьох користувачів до певної кількості збережених мікроформатів.

Рішення 2. Мікроформати, як елемент інформаційної системи, можуть слугувати локальними джерелами структурованих даних конкретних веб-сторінок, які можуть бути використані для організації пошуку та слугувати основою для консультуючих експертних підсистем. Це може бути досягнуто шляхом організації взаємодії, як це представлено на рис. 4.



Рис. 4. Модель взаємодії користувача з мікроформатами на базі Javascript та DOM

Дані, які збережені в середині веб сторінки, представляють собою семантичну розмітку спрощеного типу, що дозволяє динамічно формувати персонального помічника-агента, який обізнаний про те, якого типу інформація на цій сторінці знаходиться та яка її структура. Крім того, спеціалізація агента задається набором попередньо визначених правил, які завантажуються динамічно з сервера у вигляді скриптів Javascript та може буди змінена в залежності від дій користувача.

Ці особливості можуть ефективно використовуватись на спеціалізованих тематичних веб-сайтах, корпоративних порталах або каталогах, де набір дій користувача може бути попередньо визначений та відповідні сценарії поведінки розроблені.

Запропонована модель взаємодії відноситься до моделей оперативної обробки даних без застосування серверних можливостей або з не значним використанням.

Висновки

Мікроформати – дуже проста в застосуванні технологія, яка дає змогу застосувати різноманітні технології штучного інтелекту для покращення якості інформаційного пошуку. Крім того, мікроформати створюють підґрунтя для спрощення процесів

розробки семантичних приладань на основі більш виразних мов, таких як RDF та OWL.

В статті розглянуто технологію мікроформатів в контексті впровадження семантичних розміток до веб-прикладань та зроблений огляд існуючих мікроформатів, обробка яких підтримується сучасними пошуковими системами Яндекс, Гугл та ін.

Запропоновані моделі взаємодії користувачів з мікроформатами, які використовують засоби локального та розподіленого пошуку, дозволяють сформувати з веб-сторінки спеціалізовану базу знань, яка може слугувати основою для побудови експертних систем та орієнтованих на користувача систем візуалізації інформації. Крім того, положення запропонованих моделей формують уніфікований підхід до організації веб-сторінок та оброблюючих їх сервісів/прикладань для підвищення ефективності пошуку та наводяться рекомендації щодо його застосування.

Список літератури

1. Щербак С.С. Сбор и преобразование разнородных данных из различных источников к единому формату в системах поддержки принятия решений / С.С. Щербак // *Материалы 12-й международной конференции по автоматическому управлению (30 мая – 3 июня 2005г.)*. – Х.: НТУ «ХПИ», 2005. – Т.3. – С. 140-141.
2. Волкова В.В. Технология извлечения семантически обогащенных данных и отношений как стратегия эффективного использования интеллектуальных систем / В.В. Волкова, С.С. Щербак // *Тези доповідей 3-ї міжнародної науково-практичної конференції «Математичне та програмне забезпечення інтелектуальних систем» (16-18 листопада 2005р.)*. – Дніпропетровськ: Вид. «ДНУ», 2005. – С. 32-33.
3. Рассел С. Искусственный интеллект: современный подход: Пер. с англ. / С. Рассел, П. Норвиг. – М.: Издательский дом «Вильямс», 2006. – 1408 с.
4. Козаловский М.Р. Перспективные технологии информационных систем / М.Р. Козаловский. – М.: ДМК Пресс, 2003. – 288 с.
5. Рябова Н.В. Развитие технологий Semantic Web: Обработка RDF - графов с помощью XSLT / Н.В. Рябова, С.С. Щербак // *Восточно-Европейский журнал передовых технологий*. – 2004. – №4(10). – С. 67-72.
6. Щербак С.С. Формальные модели представления знаний для семантических web-сервисов / С.С. Щербак // *Восточно-Европейский журнал передовых технологий*. – 2008. – №3/2(33). – С. 68-71.
7. Гришанова И.Ю. Разработка распределенных семантических приложений как стратегия развития современного Web / И.Ю. Гришанова, С.С. Щербак // *Системы управления, навигации та зв'язку*. – К.: ДП «ЦНДІ навігації та управління», 2010. – №2(14). – С.194-199.

Надійшла до редколегії 1.12.2010

Рецензент: д-р техн. наук, проф. Є.І. Кучеренко, Харківський національний університет радіоелектроніки, Харків.

ТЕХНОЛОГИЯ МИКРОФОРМАТОВ КАК СТРАТЕГИЯ ПОВЫШЕНИЯ КАЧЕСТВА ИНФОРМАЦИОННОГО ПОИСКА

С.С. Щербак

Рассмотрена технология микроформатов в контексте внедрения семантических разметок в веб-страницы программных приложений и приводится обзор существующих микроформатов, обработка которых поддерживается современными поисковыми системами. Предложенные модели взаимодействия пользователей с микроформатами, кото-

рые адаптированы к использованию в Semantic Web, используют средства локального и распределенного поиска, и приводятся рекомендации, относительно их применения.

Ключевые слова: *Semantic Web, информационный поиск, онтологии, поисковые механизмы, семантика.*

TECHNOLOGY OF MICROFORMATS AS A STRATEGY TO IMPROVE THE QUALITY INFORMATION SEARCH

S.S. Shcherbak

The technology of microformats in the context of the implementation of semantic markup in web pages and software applications have considered. Overview of microformats that are popular at the moment. The designed model for users to interact with microformats, which are adapted for use in the Semantic Web, use the means of local and distributed search, and provided recommendations.

Keywords: *Semantic Web, information retrieval, ontologies, search engine, semantic.*