

МОДЕЛИ ВЗАИМОДЕЙСТВИЯ ВЕБ-ПРИЛОЖЕНИЙ НА ОСНОВЕ МИКРОФОРМАТОВ

Щербак С. С.

Харьковский национальный технический университет радиозлектроники
г. Харьков, пр. Ленина, 14, Украина
e-mail: ontolog@gmail.com

Аннотация — Рассмотрена технология микроформатов в контексте применения семантических разметок в интеллектуальных веб-приложениях, сделан обзор существующих микроформатов, обработка которых поддерживается современными поисковыми системами. Предложены модели взаимодействия веб-приложений на основе микроформатов, которые адаптированы к использованию в Semantic Web, и приведены рекомендации по их применению.

I. Введение

Микроформаты – это один из видов семантической разметки, которые используются для добавления “машиночитаемых” описаний знаний и данных, расположенных на веб-страницах и других типах электронных документов. С одной стороны, семантические разметки, такие как OWL, существуют давно и имеют широкие возможности для представления знаний о различных предметных областях, но сложность подобных разметок и необходимость иметь специализированные знания для их применения привели к тому, что разработчики веб-приложений не используют их. В результате, такая не популярность современных семантических разметок не позволяет применять методы и средства искусственного интеллекта для организации семантического поиска в веб-приложениях. С другой стороны, технология микроформатов vCard хорошо зарекомендовала себя в издательских системах для представления типовой информации о людях, организациях, книгах, публикациях и т.п.

Основой vCard являются наборы меток для представления общепринятой информации о различных типовых объектах. За каждой меткой закрепляется общепринятое значение, которое может быть использовано при машинной обработке.

На сегодняшний день на базе vCard разработано множество специализированных микроформатов для описания различных типов объектов предметных областей, которые можно использовать на веб-страницах и автоматическую обработку которых поддерживают поисковые машины Яндекс и Google.

Наиболее популярным является базовый микроформат hCard, который есть адаптированной к использованию в Веб версии vCard, и применяется для описания людей, предприятий, организаций и публикации контактной информации.

Необходимо отметить, что этот микроформат используется как базовый компонент других микроформатов, таких как hCalendar и Geo.

К широко используемым микроформатам относят:

1. hCalendar – микроформат для описания событий, времени их проведения и т.п.
2. hReview – микроформат для разметки отзывов, например, на события или материалы.
3. hProduct – унифицированная разметка товаров. Например, с помощью этого микроформата можно представить информацию о цене товара, его наличии, отзывы.
4. Geo – это микроформат для представления гео координат на веб-страницах.

Рассмотрим пример микроформата hCard и определим основные результаты, которые можно получить при его использовании.

Пусть компания, которая занимается торговлей через Интернет, на веб-сайте хочет указать поисковой машине мобильный телефон ее менеджера/консультанта (рис. 1).

```
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 class="tel">
7 <span class="type">mobile</span>:
8 <span class="value">+380 (67) xxx-xx-xx</span>
9 </span>
10 </div>
```

Рис. 1. Пример микроформата hCard.

Fig. 1. hCard Microformat example

Для правильной обработки микроформата hCard необходимо в блок Head веб-страницы добавить тип профиля используемого микроформата. Для явного указания профиля микроформата hCard необходимо добавить следующую строку: <link rel="profile" href="http://microformats.org/profile/hcard">.

Далее, в зависимости от конкретной задачи могут использоваться различные технологии извлечения контента микроформатов. Наиболее эффективными на сегодня являются технологии GRDDL (Gleaning resource Description from Dialects of Language), XSLT (extensible Stylesheet Language Transformations) и Javascript с DOM (Document Object Model). Рассмотрим более детально эти технологии в контексте решения типовых задач пользователя.

II. Модели взаимодействия веб-приложений на основе микроформатов

Микроформаты могут быть использованы не только поисковыми машинами для улучшения качества информационного поиска, но и для организации различного типа сервисов. Информация о предприятиях может составлять базу для создания глобального каталога предприятий с возможностью организации интеллектуального поиска; информация про отзывы может стать основой для создания разного рода рейтингов; Geo – координаты дают возможности применения карт территории для определения, например, кратчайшего пути к месту события и т.д. и т.п.

Рассмотрим решения, которые могут быть применены для организации взаимодействия между веб-приложениями.

Решение 1. Микроформаты, как элементы информационной системы, могут служить источниками данных, которые могут быть обработаны процессором XSLT и загружены в базу триплетов (RDF БД), как это показано на рис. 2.

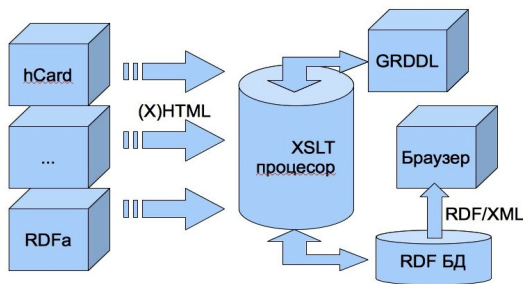


Рис. 2. Модель взаимодействия на базе GRDDL.

Fig. 2. The interoperability model based on GRDDL

Обработка XSLT процессором происходит на основе шаблонов GRDDL.

Данные, которые хранятся в базе триплетов извлекаются (по запросу пользователя) при необходимости построения веб-страницы с помощью Sparql, что позволяет загрузить данные в браузер в формате RDF/XML и визуализировать их средствами JavaScript и (x)HTML.

Эта модель взаимодействия на базе GRDDL относится к моделям с промежуточным хранением данных и может применяться в случаях, когда необходимо обеспечить централизованный доступ многих пользователей к некоторому количеству хранимых микроформатов.

Решение 2. Микроформаты, как элементы информационной системы, могут служить локальными источниками структурированных данных конкретных веб-страниц, которые могут быть использованы для организации поиска и служить основой для консультационных экспертных подсистем. Это может быть достигнуто следующим образом (рис. 3):

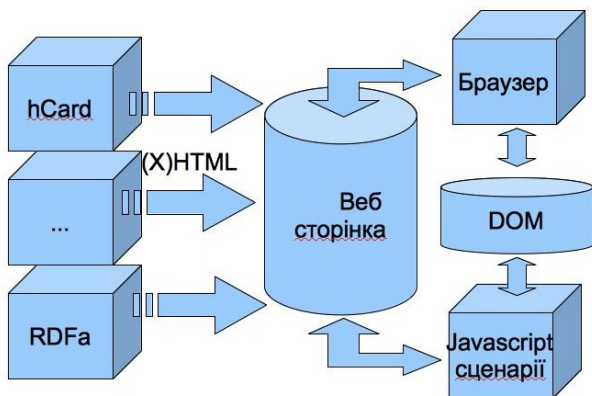


Рис. 3. Модель взаимодействия на базе DOM.

Fig. 3. The interoperability model based on DOM

Данные, которые хранятся на веб-странице, представляют собой простейшую семантическую разметку, что позволяет динамически формировать персонального агента-помощника, который знает о том, какого типа информация на этой странице расположена и какая у нее структура. Кроме того, специализация агента задается набором предварительно заданных правил, которые загружаются динамически с сервера в виде скриптов Javascript и может быть заменена в зависимости от действий пользователя.

Эти особенности могут быть эффективно использованы на специализированных тематических веб-сайтах, корпоративных порталах, каталогах, где набор действий пользователя может быть предварительно определен и соответствующие сценарии поведения разработаны.

Предложенная модель относится к моделям оперативной обработки данных без применения серверных возможностей или с незначительным их использованием.

III. Заключение

Микроформаты очень просты в применении, что позволяет применять различные технологии искусственного интеллекта для улучшения качества информационного поиска. Кроме того, микроформаты создают основу для упрощения процессов разработки семантических веб-приложений на основе более выразительных языков семантической разметки, таких как RDF и OWL.

В работе рассмотрена технология микроформатов в контексте применения семантических разметок в веб-приложениях и сделан обзор микроформатов, обработка которых поддерживается системами Google и Яндекс.

Предложенные модели взаимодействия позволяют сформировать на веб-странице базу знаний, которая может быть использована для построения экспертных систем и ориентированных на пользователя систем визуализации информации.

IV. Список литературы

- [1] Шчербак С. С. Формальные модели представления знаний для семантических web-сервисов // Восточно-Европейский журнал передовых технологий. - Харьков. - 2008. - №3/2(33). - С. 68-71.

INTEROPERABILITY WEB APPLICATION MODELS BASED ON MICROFORMATS

Shcherbak S. S.
Kharkov National University of Radioelectronics
14, Lenin Ave, Kharkov, Ukraine
e-mail: ontolog@gmail.com

Abstract — The overview of popular microformats is carried out. The technology of microformats with implementation of semantic markup web pages and software applications is considered. The interoperability web application models are designed.

I. Introduction

Microformats are widely used in development of Semantic Web Applications. The solution of this problem is in design of interoperability web application models. This paper concerns the features of the microformats with Semantic Web Applications.

II. Main Part

The interoperability web application model can be explained on the base of AI Paradigm. An architecture semantic web application is defined. The models are designed for users interacting with microformats which have been adapted for used in Semantic Web.

IV. Conclusion

Recommendations are proposed for local and distributed search.