

СЕМАНТИЧЕСКИ СТРУКТУРИРОВАННЫЕ САЙТЫ, ОРИЕНТИРОВАННЫЕ НА ПОДДЕРЖКУ ПОДГОТОВКИ И ПРОВЕДЕНИЯ НАУЧНО-УЧЕБНЫХ МЕРОПРИЯТИЙ

С.Г. Мошенко, Д.Г. Колб

*Белорусский государственный университет информатики и радиоэлектроники,
г. Минск, Республика Беларусь, moshenko@bsuir.by*

Abstract — This article deals with the main provisions of the development of web-sites that support the semantic search and navigation, and allow the transition from such web-sites to intelligent information-reference systems. The proposed approach is the presentation of information using semantic networks. The platform MediaWiki used to implement a semantically structured websites.

Организация различных научных и образовательных мероприятий является одной из важных задач дистанционного образования. В настоящее время формы проведения таких мероприятий разнообразны: видеофорумы, видеолекции или просто организация удаленного доступа к материалам семинаров по различным тематикам [1]. Ключевым моментом при организации таких мероприятий является систематизация и структуризация информации в рамках некоторого web-портала. Рассмотрим основные этапы разработки подобного web-сайта на примере одного из web-сайтов такого типа – web-сайте конференции.

В процессе анализа предметной области были выявлены следующие ключевые сущности:

- Персона;
- Доклад;
- Фотография;
- Мероприятие;
- Место размещения;
- Пункт питания;
- Организатор конференции;
- Конференция.

Кроме указанных сущностей были выделены специальные сущности, которые позволяют более полно описать основные. К таким сущностям можно отнесены: состав оргкомитета конференции, состав программного комитета конференции – для сущности конференция, географическое местоположение, адрес – для сущностей пункт питания и организатор конференции, персональные фото, коллективные фото – для сущности персона, мероприятие и ряд других.

Каждый экземпляр основной сущности описан на отдельной странице. Каждому такому экземпляру соответствует специальный набор семантических отношений, которые будут показывать связи данного экземпляра сущности с другими сущностями, которые присутствуют в рамках сайта. Для каждой вспомогательной сущности разработаны и специфицированы специальные отношения. Для того, чтобы более точно передать семантику используемых в рамках сайта отношений для каждого класса отношений разрабатывается специальная страница, которая его специфицирует.

В качестве формальной основы предлагаемого подхода будем использовать семантические сети с базовой теоретико-множественной интерпретацией. Основным способом кодирования информации для таких сетей является SC-код (Semantic Code)[2]. Интеллектуальные системы, построенные с использованием SC-кода, будем называть sc-системами.

Сущности web-страницы будем представлять как элементы базы знаний (БЗ). Такой подход позволяет рассматривать пользовательский интерфейс web-сайта как специализированную интеллектуальную систему, решающую задачу организации диалога человека и предметной интеллектуальной системы, и обеспечивающую решение основных задач предметной интеллектуальной системы. Основными классами задач такой специализированной интеллектуальной системы являются:

- просмотр текстов внешних языков, удобных и понятных для пользователя интеллектуальной системы;
- редактирование текстов внешних языков;
- трансляция текстов внешних языков в тексты SC-кода;
- трансляция текстов SC-кода в тексты внешних языков.

Для решения задач каждого класса используется отдельный класс компонент. Каждый компонент трактуется как специализированная интеллектуальная система, имеющая свою БЗ и машину обработки знаний (МОЗ). Пользовательский интерфейс web-сайта в целом является результатом интеграции всех его компонент.

В основу предметной интеллектуальной системы положено представление знаний предметной области с помощью SC-кода. Машину обработки знаний такой интеллектуальной системы составляет множество согласованных предметно-независимых операций [2].

Рассматривая в таком ракурсе web-сайт, мы приходим к заключению, что степень интеллектуальности web-сайта будет зависеть от степени реализации указанных классов компонент web-сайта.

Будем называть семантически структурированными гипертекстами гипертексты, информация в которых будет отображаться помощью SCn-кода (способа псевдо-естественного кодирования семантических сетей, представленных в SC-коде, Semantic Code natural). Разметка таких гипертекстов производится с помощью SCnML (SCn Markup Language) – языка разметки текстов SCn-кода [2]. Web-сайты, построенные с помощью указанных средств, будем называть семантически структурированными web-сайтами.

В семантической технологии проектирования баз знаний sc-систем SCn-код используется в качестве одного из основных способов представления знаний. Использование для записи текстов базы знаний языка близкого к естественному существенно повышает качество восприятия текстов баз знаний как инженерами по знаниям, так и пользователями баз знаний, поэтому SCn-код и может использоваться в качестве внешнего, понятного пользователю языка sc-системы.

SCn-код задается множеством всех sc.n-статей, каждая из которых описывает семантическую окрестность некоторого понятия предметной области. Каждая статья в свою очередь состоит из идентификатора sc-элемента, описываемого в этой sc.n-статье, и, возможно, одного или нескольких последующих sc.n-полей. При описании sc-элемента в sc.n-статье sc.n-поля описывают как, какими ролями и связками каких отношений, связан описываемый sc-элемент с другими sc-элементами. Ряд sc.n-полей может содержать мультимедиа или тексты логических утверждений. Мультимедиа может включать любые информационные конструкции, обозначаемые как внешние по отношению к SCn-коду, в том числе и sc.n-тексты.

Мультимедийные ресурсы, имеющиеся на сайте, включают в себя: фотографии форматов png и jpg, гипермедийные ссылки, документы в формате pdf, презентации, географические данные.

В рамках сайта все имеющиеся мультимедийные ресурсы можно классифицировать по способу предоставления информации на графические

(фотографии, географические данные) и текстовые (гипермедийные ссылки, текстовые документы). Также мультимедийные ресурсы классифицируются по интерактивности на интерактивные (гипермедийные ссылки, географические данные) и не интерактивные (фотографии, текстовые документы). Помимо этого, мультимедийные ресурсы различают по принадлежности к ключевым сущностям (фотография и доклад) и к вспомогательным (логотип, географические данные и др.)

Каждому sc.n-полю в языке SCnML соответствует отдельный тег. Такой подход позволяет однозначно определить соответствие между отображаемым sc.n-полем и тегом SCnML, с помощью которого это поле кодируется.

В рамках данной работы будем рассматривать семантическую структуризацию гипертекста на каждом из его слоев, в отличие от традиционных подходов, в которых семантическую структуризацию рассматривают на уровне скрытого слоя (невидимое пользователю web-сайта содержимое web-страницы). Это означает, что SCn-код позволяет семантически структурировать как отображаемое на экране содержимое web-страницы, так и скрытое (невидимое пользователю web-сайта) содержимое web-страницы. Такой подход позволяет существенно сократить время нахождения пользователем необходимой ему информации даже без использования специальных поисковых средств за счет простейшей гипертекстовой навигации, использующей явно указанные семантические связи между web-страницами.

В качестве опытной платформы для реализации семантически структурированных web-сайтов использована платформа MediaWiki (платформа, которую использует проект wikipedia). В основе подхода, используемого в проектах на основе MediaWiki, лежит понятие статьи. В традиционной трактовке статья MediaWiki определяется как семантическая окрестность некоторого понятия, записанная на естественном языке. Таким образом, для перехода от обычного гипертекста, используемого в статье Mediawiki, необходимо формально представить семантическую окрестность описанного на естественном языке указанного понятия с помощью средств SCn-кода.

Сравнивая типологию связей между статьями, создаваемыми на базе платформы Mediawiki и традиционного гипертекста можно заметить, что переход по ссылке дает возможность ответить на вопрос «Что это такое X?». Использование же предложенного подхода в совокупности с платформой Mediawiki позволяет существенно увеличить типологию семантических связей между статьями Mediawiki за счет явной формальной спецификации таких связей с помощью средств SCn-кода. Практически это приводит к тому, что переход по ссылке из семантически структурированного гипертекста позволяет ответить на вопрос “Что такое X, связанное с данным понятием отношением Y?”. Такие семантические связи при использовании предлагаемого подхода доступны пользователю web-сайта как через отображаемый слой гипертекста в виде псевдо-естественного (полуформального) текста, так и через скрытый слой в виде текстов SCnML.

Представленный в данной работе подход был опробован при разработке сайта конференций проекта OSTIS[3].

Литература

1. Интернет университет информационных технологий [Электронный ресурс]. – 2010. - Режим доступа: <http://www.intuit.ru>. – Дата доступа: 05.10.2011
2. Проект OSTIS [Электронный ресурс]. – 2010. - Режим доступа: <http://www.ostis.net>. – Дата доступа: 05.10.2011
3. Конференции OSTIS [Электронный ресурс]. – 2010. - Режим доступа: <http://conf.ostis.net>. – Дата доступа: 05.10.2011.