

Семантическая модель сайта конференции

Колб Д.Г.

Каф. ИИТ, ФИТиУ

БГУИР

г. Минск, Республика Беларусь

e-mail: kolb@bsuir.by

Аннотация — Приведены основные положения подхода к разработке web-сайтов, поддерживающих семантический поиск и навигацию, а также, позволяющих осуществить переход от таких web-сайтов к интеллектуальным информационно-справочным системам. В основе предлагаемого подхода лежит представление информации с помощью семантических сетей.

Ключевые слова: web-сайт; семантическая сеть; база знаний; интеллектуальная система.

I. ВВЕДЕНИЕ

В современных научных кругах различные конференции, международные форумы являются одним из средств обмена научным опытом, представления научной общественности результатов исследований. Поэтому важной практической и научной задачей является организация программной поддержки сайтов таких мероприятий. Очевидно, что для эффективного решения такой задачи, необходимо использовать методы и средства искусственного интеллекта, так как именно эти средства позволяют наиболее эффективно специфицировать и систематизировать знания в рамках, указанной предметной области, а также обеспечить сайт средствами семантической навигации и поиска.

II. СУЩНОСТИ МОДЕЛИ САЙТА КОНФЕРЕНЦИИ

В процессе анализа предметной области были выявлены следующие ключевые сущности:

- Персона – сущность содержит краткую информацию об участнике конференции: научная степень и звание, место работы, контактные данные, доклад и номер доклада, публикации;
- Доклад – сущность содержит краткую информацию о докладе: название доклада, в состав какой секции входит, авторы, иллюстрационные материалы, опубликованные материалы;
- Фотография – сущность содержит фотографию и её спецификацию: событие, когда она была сделана и кто на ней изображен;
- Мероприятие – сущность описывает событие конференции время его проведения основные этапы и фотографии мероприятия;
- Место размещения – сущность описывает информацию о месте размещения участников, приводятся телефон, адрес, адрес web-страницы в интернет;
- Пункт питания – сущность описывает информацию о пункте питания участников, приводятся телефон, адрес, адрес web-страницы в интернет;
- Организатор конференции – сущность описывает информацию об организаторе конференции, приводится список сотрудников, которые участвовали в конференции;
- Конференция – сущность содержит всю необходимую для участника конференции информацию: время проведения, формат проведения, состав оргкомитета и программного комитета, информационное письмо и правила оформления.

Кроме указанных сущностей были выделены специальные сущности, которые позволяют более полно описать основные. К таким сущностям можно отнести: *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.

III. СРЕДСТВА ПРЕДСТАВЛЕНИЯ МОДЕЛИ САЙТА КОНФЕРЕНЦИИ

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) [1]. 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 [1].

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 [1]. 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.

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) содержимое 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[2].

[1] Проект OSTIS [Электронный ресурс]. – 2010. - Режим доступа: <http://www.ostis.net>. – Дата доступа: 01.09.2011

[2] Конференции OSTIS [Электронный ресурс]. – 2010. - Режим доступа: <http://conf.ostis.net>. – Дата доступа: 01.09.2011.