

WEB-ОРИЕНТИРОВАННАЯ РЕАЛИЗАЦИЯ СЕМАНТИЧЕСКИХ МОДЕЛЕЙ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ ДЛЯ СИСТЕМ ДИСТАНЦИОННОГО ОБУЧЕНИЯ

Д.Г. Колб

Белорусский государственный университет информатики и радиоэлектроники, Минск, Беларусь, kolb@bsuir.by

Abstract. An approach to the construction of semantically-structured web-sites of different levels of intelligence for distance learning systems. The proposed approach is the use of uniform semantic networks with the basic set-theoretic interpretation.

Четкая и ясная структуризация и унификация подачи учебного материала является наиболее важной составляющей любой системы дистанционного обучения (СДО). Объемы информации, которые накоплены в сети Internet за последние годы, и неэффективность методов её представления и обработки указывают на необходимость использования новых средств представления и обработки информации. Наиболее перспективными и активно развивающимся технологиями для решения указанной проблемы являются технологии, которые позволяют представлять знания СДО в виде семантических сетей. Данный формализм активно используются как формальная база для различных прикладных интеллектуальных систем с помощью методов и средств искусственного интеллекта.

Традиционным подходом для организации СДО является организация её в виде web-портала или web-сайта. При такой организации СДО можно выделить следующие уровни, на которых можно осуществлять интеллектуализацию: *уровень внешнего представления, уровень языков разметки, уровень серверных скриптов, уровень хранилищ данных.*

Уровень внешнего представления – это часть web-ресурса, которая видима пользователю. В настоящее время существуют масса подходов к структуризации информации на этом уровне web-ресурса. Наиболее известными из них являются: структуризация в виде обычного текста; структуризация в виде линейного спискового представления; структуризация в виде иерархического спискового представления; структуризация в виде табличного представления; структуризация за счет выделения элементов различной значимости шрифтом, фоном или оформлением и структуризация на основе индуктивного подхода к организации web-ресурса.

Уровни языков разметки, серверных скриптов, хранилищ данных пользователю не видимы, однако именно эти уровни определяют функции поиска необходимой пользователю информации, как с помощью внешних поисковых систем, так и с помощью встроенных в СДО средств поиска.

Обозначим направления работы по интеллектуализации СДО, организованных в виде web-портала:

- Семантическая структуризация информации на уровне внешнего представления;
- Семантическая структуризация информации на уровне языков разметки;
- Реализация операций семантического поиска и навигации по информационному пространству web-ресурса;
- Использование хранилищ знаний вместо баз данных.

Рассмотрим один из возможных подходов к решению указанных в предыдущем абзаце задач. В качестве формальной основы предлагаемого подхода будем использовать семантические сети с базовой теоретико-множественной интерпретацией.

Основным способом кодирования информации для таких сетей является SC (Semantic Code)-код [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 web-сайт как специализированную интеллектуальную систему, решающую задачу организации диалога человека и предметной интеллектуальной системы, и обеспечивающую решение основных задач предметной интеллектуальной системы.

Рассматривая в таком ракурсе web-сайт, мы приходим к заключению, что степень интеллектуальности web-сайта будет зависеть от степени внедрения на каждом из выделенных уровней web-сайта семантических технологий.

Основу уровня внешнего представления при таком подходе составляют **семантически структурированные гипертексты** – гипертексты, информация в которых будет отображаться помощью SCn-кода (способа псевдо-естественного кодирования семантических сетей, представленных в SC-коде, Semantic Code natural)[2]. 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 (объекта предметной области, закодированного с помощью SC-кода), 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.

Основу уровня языков разметки задает SCnML(SCn Markup Language) – модель языковых средств разметки семантически структурированных гипертекстов. SCnML – определяет общие правила разметки текстов языка SCn. SCnML–содержит следующие классы тегов для разметки sc.n-статей: тег описываемого объекта, теги связей, теги компонентов связей, теги поисковых запросов. Любая прикладная реализации SCnML обязана реализовать указанные классы тегов.

Уровень серверных скриптов задается набором операций трансляции и интерпретации текстов одной из реализаций SCnML-разметки. Можно выделить два класса операций трансляции: класс операций, которые обеспечивают трансляцию разметки в одно из внешних представлений, базовым из которых является SCn-представление, и класс операций, которые обеспечивают трансляцию в некоторое внутреннее представление, закодированное с помощью SC-кода. Операции интерпретации обеспечивают обработку тегов поисковых запросов на множестве SCnML текстов.

Уровень хранилищ данных, в рамках предлагаемого подхода, можно обеспечить тремя способами. Первый способ заключается в организации на основе традиционных СУБД хранилищ знаний. При таком способе СУБД могут быть использованы для хранения связей и компонентов связей в рамках таблиц базы данных – такое высокоуровневое хранение позволяет обеспечить эффективных поиск для простых информационно поисковых запросов, что может быть использовано в простейших справочных системах, какими обычно являются информационные web-сайты. Второй способ заключается в использовании в качестве хранилища знаний сетевые СУБД. Использование сетевых СУБД позволяет хранить знания в более близком к используемым в данном подходе семантическим сетям представлении. Второй способ позволяет обеспечить выполнение более широкий класс поисковых запросов, чем при

использовании первого способа. И последний способ заключается в организации специализированного хранилища, ориентированного на хранение знаний представленных в SC-коде, в полном объеме, поддерживающем номенклатуру элементов.

Разработка web-сайтов, при использовании предложенного подхода проходит в следующем порядке: разработка формального описания предметной области, по которой разрабатывается web-сайт (sc-модели); выделение сущностей предметной области, которые будут соответствовать страницам web-сайта и представление их в виде SCn-текстов; выбор необходимой для решения задачи реализации SCnML и разработка операций обработки SCnML-текстов; выбор и использование одного из возможных хранилищ знаний; тестирование и отладка web-сайта.

Отметим, что приведенные этапы предусматривают интеллектуализацию web-сайта на всех выделенных в данной работе уровнях. Однако для промышленной разработки не всегда необходима реализация всех этих уровней. Предлагаемый подход к разработке web-сайта позволяет реализовать необходимое для решения задачи количество уровней, используя предложенные средства. Остальные же уровни можно реализовывать традиционным способом. Такая гибкость достигается за счет того, что в основе web-сайта лежит использования единой формальной основы – sc-модели и каждый уровень, который мы выделили, лишь обеспечивает интерпретацию данной модели с помощью разных языковых и программных средств.

Отметим ряд достоинств предлагаемого подхода к разработке СДО на базе web-сайтов:

- подход является полностью независим от программных средств реализации за счет с использования единой формальной основы sc-модели предметной области web-сайта;
- использование семантических моделей дает возможность реализации в рамках таких web-сайтов средств семантического поиска;
- при использовании предложенного подхода web-сайты будут совместимы на уровне SCn и SCnML представлений, что позволяет, используя средства интеграции баз знаний, построенных с использованием SC-кода объединять web-сайты в web-порталы знаний, которые будут построены по единым унифицированным принципам;
- при использовании предложенного подхода web-сайты можно рассматривать как исходники баз знаний, построенных с использованием технологий проекта OSTIS.

Предложенный подход разработан в рамках проекта OSTIS[2] и опробован на следующих сайтах: сайт конференций OSTIS (<http://conf.ostis.net>), информационно справочная система по геометрии (<http://ostisgeometry.sourceforge.net/>), информационно справочная система по русскому языку (<http://ostisrussianlan.sourceforge.net/old/index.php>) и ряде других web-сайтов.

Работа выполнена при поддержке гранта БРФФИ Ф10М-085 и гранта БРФФИ-РФФИ Ф10Р-175.

Литература

1. Представление и обработка знаний в графодинамических ассоциативных машинах /В. В. Голенков, [и др]; – Мн. : БГУИР, 2001.
2. Проект OSTIS [Электронный ресурс]. – 2010. - Режим доступа: <http://www.ostis.net>. – Дата доступа: 27.11.2010