

УДК 004.896

А.В. Григорьев, Е.В. Павловский
Донецкий национальный технический университет
grigorie@r5.dgtu.donetsk.ua, pavlovskiy.evgeny@mail.ru

Представление сайта на языке HTML как модели структурного уровня САПР

В предлагаемой статье рассматривается специфика представления сайта, выполненного на языке HTML, как типичной модели структурного уровня абстракции САПР. Приводится пример выявления структурной модели в типичном сайте. Показывается, что данная модель структуры соответствует возможностям концептуальной модели предметной области инструментальной оболочки для автоматизации построения САПР, что позволяет сформулировать задачу построения такого рода оболочки.

САПР, онтология, средний эксперт, база знаний, И-ИЛИ дерево, HTML

Введение

Ранее авторами рассматривался подход представления веб-приложений с точки зрения онтологического подхода как модель некоторого объекта предметной области, призванного обеспечить представление знаний для базы знаний (БЗ) [1]. В данном случае база знаний относится к САПР веб-приложений.

Приведенная ранее онтология [1] соответствует стандартному подходу представления веб-страницы. Однако в таком виде она не обеспечивает возможность построения базы знаний о методиках проектирования веб-приложений в силу своей упрощенной и примитивной структуры.

Чтобы возможным было обеспечить построение такой базы знаний, необходимо обеспечить представление, например, страницы сайта с точки зрения стандартных методов представления объектов проектирования в САПР технических объектов. Такие модели делятся по уровню абстракции, например: структурные модели, функционально-логические модели, количественная микро и макро модели [2].

Представить модель объекта, как структурную модель, означает ввести отношения: принадлежности к множеству для компонентов модели: класс – подкласс, часть – целое, экземпляр – класс и т.д.

Общая постановка проблемы

В настоящее время разработка сайтов становится все более актуальной задачей, как для различных компаний и фирм, так и для

частных лиц, поскольку создание сайта направлено на решение многих важных задач.

Актуальной на сегодняшний день является задача автоматического проектирования сайтов, так как растет количество пользователей интернета и потребность создания сайтов с минимумом временных затрат. При этом проектировщик не всегда является профессионалом в области веб-программирования.

Нашей общей задачей является создание инструментальной оболочки, способной автоматизировать создание базы знаний о методиках проектирования веб-приложений на требуемых языках программирования (HTML). Данная база знаний и является требуемым САПРом. Такой САПР должен обеспечить синтез требуемого веб-приложения, исходя из параметров, задаваемых конечным пользователем.

Постановка задач исследования.

Сейчас тенденции в программировании ведут к созданию генераторов программ, решающих ту или иную проблему в виде БЗ. Имеется два подхода к построению генератора программ: 1) генерация предметно независимого языка с кодированием в проблемно зависимый язык; 2) генерация на проблемно зависимом языке [1].

Целью предлагаемой работы являются:
- выявление структурной модели, имеющейся в типичном сайте, организованном с помощью средств языка HTML;

- представление данной модели в рамках возможностей концептуальной модели предметной области инструментальной оболочки для построения интеллектуальных САПР.

Для раскрытия данной темы необходимо осветить следующий ряд вопросов:

- дать характеристику языку разметки HTML, как языку строгой иерархии;

- описать роль онтологий, как средств наиболее полного представления страниц HTML;

- рассмотреть типичную логическую структуру страницы сайта;

- построить структурную модель HTML страницы в рамках средств концептуальной модели предметной области инструментальной оболочки для построения интеллектуальных САПР.

Конечной задачей является создание инструментальной оболочки, способной автоматизировать создание базы знаний о методиках проектирования вэб-приложений на требуемых языках программирования.

Решение задач и результаты исследований.

1. Язык разметки HTML, как язык строгой иерархии

Сам по себе язык разметки HTML строго структурирован. Язык состоит из множества тегов. Теги могут быть вложены друг в друга и могут иметь атрибуты [3]. Обычно страница сайта разбивается на блоки. Это может быть лента новостей, форма авторизации, переключение языка, гостевая книга и т.д. Каждый блок имеет свои атрибуты и позиционирование относительно других блоков сайта. Распространенной является табличная верстка. Страница сайта разбивается на блоки с помощью таблицы, т.е. каждый отдельный блок – это отдельной ячейке таблицы. В свою очередь, блоки могут содержать неограниченное количество подблоков. Таким образом, каждый блок имеет входы и выходы. Его входами являются любые существенные воздействия среды на синтезируемый объект, выходами – связи объекта со средой. Язык HTML однозначно позиционирует блоки относительно друг другу. Вложенность тегов друг друга – это мощное структурная особенность языка. На сегодняшний день является актуальным представление информации в виде онтологии. Причем каждый трактует это понятие по-своему. Есть базовое понятие и базовое формальное описание. Для конкретных задач можно формальное описание дополнять какими-либо параметрами.

2. Онтология, как наиболее полное представление страницы HTML

Базовое понятие онтологии [4] соответствует стандартному подходу представления вэб-страницы. Однако в таком виде она не обеспечивает возможность построения базы знаний о методиках проектирования вэб-приложений в силу своей упрощенной и примитивной структуры.

Чтобы возможным было обеспечить построение такой базы знаний, необходимо обеспечить представление, например, страницы сайта с точки зрения стандартных методов представления объектов проектирования в САПР технических объектов. Такие модели делятся по уровню абстракции [5], например: структурные модели, функционально-логические модели, количественная микро и макро модели.

Представить модель объекта, как структурную модель означает ввести отношения: принадлежности к множеству для компонентов модели, класс – подкласс, часть – целое, экземпляр – класс и т.д.

Проанализировав существующие структуры данных, сделан вывод, что онтология – это оптимальный вариант для представления и связи компонентов HTML страницы в единую систему.

С точки зрения формального описания онтологии, для представления страницы HTML в виде онтологии, минимально необходимым является знание о составе объекта предметной области (в нашем случае это блоки) и связей между ними (в нашем случае вложенность блоков).

В статье рассматривается специфика представления сайта, выполненного на языке HTML (см. рис.1), как типичной модели структурного уровня абстракции САПР. Приводится пример выявления структурной модели в типичном сайте. Показывается, что данная модель структуры соответствует возможностям концептуальной модели предметной области инструментальной оболочки для автоматизации построения САПР, что позволяет сформулировать задачу построения такого рода оболочки. Это своего рода редактор онтологий типа диаграмм IDEF0 или IDEF3 в технологии SADT. Причем, мы должны строить онтологии как множественные [6], т.е. учитывающие все альтернативные пути декомпозиции.

Мы не описываем множественной модели, а показываем, что путем декомпозиции блоков можно описать один сайт.

Каждый блок (прямоугольник) – это ячейка главной таблицы (id="table_main"). С помощью этой таблицы сайт разбивается на блоки. Эти блоки и есть ячейки главной таблицы. Другими словами, ячейки главной таблицы – это блоки, которые являются вершиной иерархии блочной структуры сайта.

В свою очередь эти блоки могут иметь подблоки, которые принимают от внешнего блока его атрибуты, а также имеют свои. На многих сайтах при переходе на другие страницы меняется только центральный блок. Остальные блоки, такие как верхний или боковой, могут отсутствовать.

Будем строить структурную модель для данного типичного сайта, исходя из общей концептуальной модели предметной области мета-эвристической оболочки, т.е. специальной оболочки для создания интеллектуальных САПР [9].

5. Семантика концептуальной модели

Рассмотрим семантику концептуальной модели. Выделим следующие сущности:

- «собственное» свойство – идентификатор точки есть некоторая фазовая переменная данной предметной области, задающая потенциал точки для некоторой субстанции;

- совокупность «собственных» свойств точки определяется уровнем представления модели и предметной областью;

- «чужое» свойство точки – задание факта наличия связи-отношения, задающего меру влияния потенциала («собственного» свойства) другой точки на потенциал данной точки;

- отношение зависимости потенциалов точки-хозяина и потенциала влияющей точки задается функцией, определенной над «чужим» и «собственным» свойством точки-хозяина;

- единственное «кольцевое» отношение среди всех свойств точки имеет только «собственное» свойство, все прочие свойства связывают только свойства различных точек;

- отсутствие в наборе свойств некоторой точки1 «чужого» свойства точки2 говорит о невозможности влияния потенциала точки2 на потенциал точки1 (однаправленная связь);

- функция задает поток изменения фазовой переменной во времени, исходя из существования данной «простой» связи внутри «пространственной» связи;

- для отработки функции-связи необходимо задание всей истории поведения данной связи во времени с целью косвенного задания ее параметров;

- переход на новый уровень предполагает не только декомпозицию избранного набора

блоков (точек) и свойств на более мелкие, но и увеличение числа этих точек, т.е. изменение (расширение) вышележащих уровней декомпозиции [11].

Основой языка формального описания модели является семиотическая модель [12] декомпозиции отношений между сущностями, к которым относятся блоки, свойства и т.д. Неформально проиллюстрируем систему уровней. Отношения в модели трактуются как связи между блоками по свойствам. Заострив свое внимание на структурной модели САПР, выполним подобное представление для типичной страницы сайта. На рис. 1 представлен исходный текст главной страницы сайта.

6. Структурная модель объекта проектирования

В САПР различаются следующие уровни моделей: [13]

- структурный;

- системный;

- функционально-логический, с дискретными значениями параметров;

- количественный, с непрерывными сосредоточенными параметрами (макромодел);

- количественный, с непрерывными распределенными параметрами (микромодел).

Покажем, что с точки зрения САПР, в задаче автоматизации создания требуемого Вэб-приложения речь идет о структурной модели объекта проектирования. Таким образом, в типичной структуре сайта нужно выявить в соответствии со структурной моделью в САПР [14]:

- 1) Отношения причины-следствия в структурной модели, это фактически отношения передачи информации.

- 2) Отношения принадлежности к множествам – это декомпозиция разных блоков одинаковым путем как блоков некоторого одного типа.

- 3) Структурные отношения между подблоками.

- 4) Отношения иерархической подчиненности.

Проделаем подробный анализ, сделав акцент на отношения причины-следствия и отношений иерархической подчиненности.

Многие элементы страницы (теги) имеют атрибуты. Например, ``. Тег `img` – является изображением, адрес которого находится в теге `src`. Также элемент содержит атрибуты ширины и высоты изображения, а также идентификатор `id`. Элемент `` является подблоком главного

блока (таблица id="table_main"). Атрибуты находятся на внутренней границе блока.

Как известно, заголовок HTML страницы находится в теге <title>. Из этого следует, что текст заголовка – это подблок блока title. Вход для блока с текстом заголовка страницы – является выходом из блока title. Текст заголовка – это подблок блока title (тега <title>). Вышеизложенные сущности представлены на рис.3. Все, что мы видим на странице сайта, находится в теге <body>, т.е. все элементы являются подчиненными блоку body.

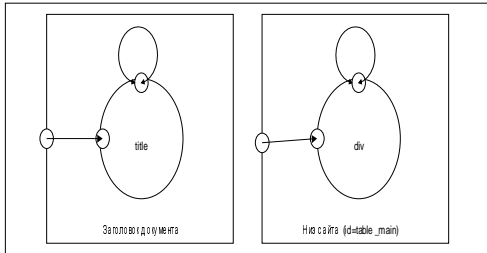


Рисунок 3 – Блок заголовка страницы и блок с автором сайта

Изображение вверху сайта находится в верхнем блоке главной таблицы. К нему на вход подаются атрибуты внешнего блока (см. рис. 4). Как видно из листинга страницы (см. рис. 1), блок с изображением занимает три колонки таблицы. Этот атрибут подается на внутреннюю границу блока изображения.

Текст приветствия находится в центральной ячейке главной таблицы. Текст находится в теге <p>, который находится во внешнем блоке главной таблице. Ему на вход передаются атрибута внешнего блока.

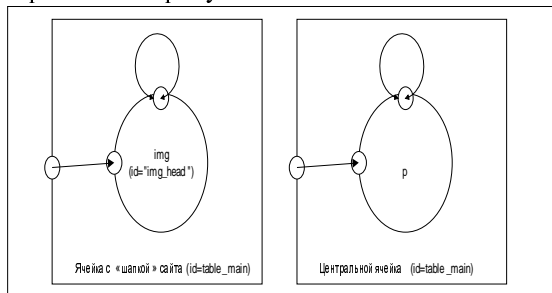


Рисунок 4 – Блок «шапки сайта» и центральный текст главной страницы

Форма авторизации находится во вложенной таблице id="table_auth". Блок table, у которого id="table_auth", является подблоком блока <form> с id="form_auth", который, в свою очередь, является подблоком блока главной таблицы id="table_main".

Для полей ввода логина и пароля входами является выход блока пользователь, т.е. информация которую вводит пользователь. Выходы блоков логина и пароля является входами для базы данных.

Поля для ввода – это ассоциативный массив вида [[name1,value], [name2,value], ..., [nameN,value]]. Name – это имя элемента для ввода, value – это значение, введенное пользователем. Их содержимое отправляется на сервер, после чего сопоставляется с содержимым базы данных.

Как видно из рис. 5, элементы html страницы могут быть вложены друг в друга. Форма авторизации <form id="form_auth"></form> вложена в ячейку таблицы <table id="table_auth"></table>. Последняя в свою очередь содержится в ячейке главной таблицы <table id="table_main"></table>.

Приходим к выводу, что подблок должен иметь вход только от одного внешнего блока, который ближе к нему. Ведь атрибуты дальнего внешнего блока, будут переданы на вход подблоку. Этого достаточно для построения четкой иерархии, которую предусматривает формат HTML.

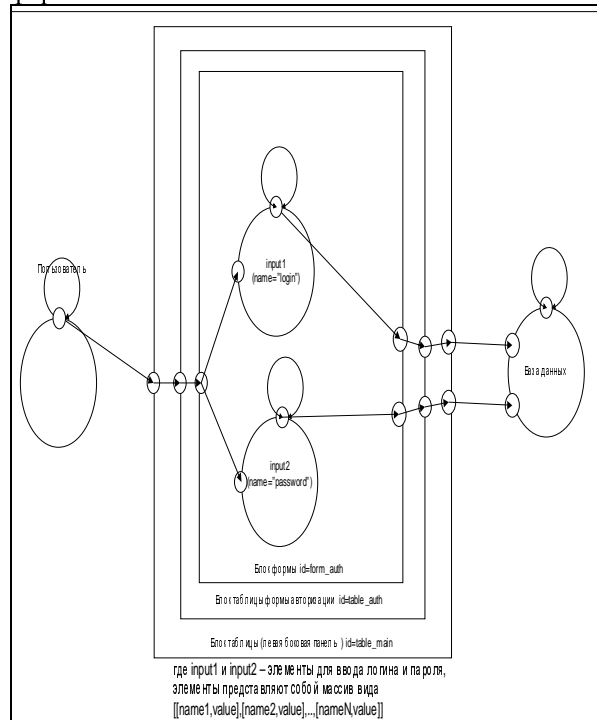


Рисунок 5 – Блок формы авторизации

Из рис. 1 видно, что на странице сайта присутствует лента новостей. Обычно эта информация извлекается из базы данных и в удобочитаемом виде располагается на странице (обычно в таблице) [15].

Новость состоит из даты и самого текста новости. Данные из базы данных возвращаются в массиве. Каждый элемент – это дата и текст новости. Упрощенно – это массива вида

[[div1,div2],[div3,div4]...[divN-1,divN]], где каждый элемент – это одна новость.

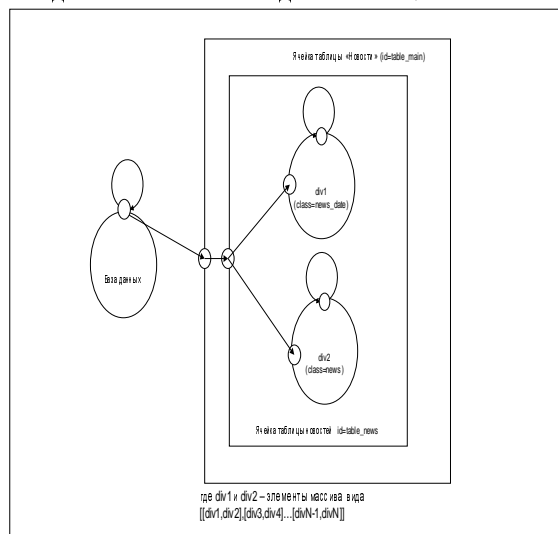


Рисунок 6 – Блок ленты новостей

Из рис.6 видно, что данные из базы данных подаются на границу внешнего блока главной таблицы `id="table_name"`, затем внешний блок передает данные и атрибуты на подблок `id="table_news"`.

Вывод

Подобное представление структуры html файла позволяет создать систему модулей знаний о методах построения блоков различных типов, имеющих в html-файле. Подобное представление соответствует заданию структуры объекта на универсальном языке формальных спецификаций. При наличии транслятора «язык формальных спецификаций - > проблемно-ориентированный язык (HTML)» можно обеспечить синтез желаемого сайта пользователем, эксплуатирующем готовую базу знаний.

Таким образом, разработка инструментального комплекса, способного обеспечить создание в его среде баз знаний для автоматизации синтеза сайтов как интеллектуального САПР, является перспективной задачей.

Стоит отметить, что задача имеет ряд специфических особенностей, что требует разработки соответствующего комплекса средств концептуальной модели, способного учесть специфику задачи.

Список литературы

1. Григорьев А.В. Анализ методов построения онтологий для построения экспертных систем по синтезу моделей сложных систем в САПР / А.В. Григорьев, Е.В. Павловский // Научные труды Донецкого государственного университета: Серия: Вычислительная техника и автоматизация. – 2011. – № 21. – С. 87-94.
2. Григорьев А.В. Анализ эффективности и перспектив развития методов построения двухсторонних трансляторов в задаче создания интеллектуальных надстроек над проблемно-ориентированными САПР / А.В. Григорьев, О.В. Морозова // Научные труды Донецкого государственного университета: Серия: Вычислительная техника и автоматизация. – 2011. – № 21. – С. 116-129.
3. Матросов А. Сергеев. HTML 4.0: А. М. Чаунин / А. Матросов. — М., БХВ-Петербург, 2008. – 672 с.
4. Онтология [Электронный ресурс]. Энциклопедия Википедия. Режим доступа: <http://ru.science.wikia.com/wiki/Онтология>.
5. Кроль Т.Я. Многоуровневое объектное описание интегрированной среды автоматизированного проектирования / Т.Я. Кроль, П.Н. Чистяков, А.С. Капитонихин // «Вестник ИГЭУ». – 2005. – № 4.
6. Ломов П.А. Интеграция онтологий с использованием тезауруса для осуществления семантического поиска / П.А. Ломов, М.Г. Шишаев – Режим доступа: http://crider.rork.ru/download/papers/3_integration_in_tezaurus.pdf.
7. Григорьев А.В. Пути создания интеллектуальных САПР при различных уровнях квалификации экспертов / А.В. Григорьев // Научно-теоретический журнал «Искусственный интеллект». – 2005. – №3. – С. 758–763.
8. Григорьев А.В. Методы поиска новых решений в специализированной инструментальной оболочке для создания интеллектуальных САПР / А.В. Григорьев // Десятая национальная конференция по искусственному интеллекту с международным участием КИИ-2006 (25-28 сентября 2006г., Обнинск): Труды конференции; в 3-т. -- М.: Физматлит, 2006. – Т.3. – С. 1031-1046.
9. Григорьев А.В. Адаптивная система ограничений на сложность при синтезе новых решений в интеллектуальных САПР / А.В. Григорьев // Искусственный интеллект. – 2001. – № 2. – С. 152-167.
10. Гаврилова Т.А. Использование онтологий в системах управления знаниями / Т.А. Гаврилова // Труды международного конгресса «Искусственный интеллект в XXI веке», Дивноморское, Россия. – М.: Физматлит, 2001 – С. 21-33.

11. Григорьев А.В. Организация пространственного и временного логического вывода в концептуальной модели интеллектуальных САПР / А.В. Григорьев // Научные труды Донецкого государственного университета : Серия: Информатика, кибернетика и вычислительная техника. – 2008. – № 9(132). – С.296-311.

12. Григорьев А.В. Семиотическая модель базы знаний САПР / А.В. Григорьев // Научные труды Донецкого государственного университета. Серия: «Проблемы моделирования и автоматизации проектирования динамических систем». – 1999. – № 10. – С. 30-37.

13. Норенков И.П. Введение в автоматизированное проектирование технических устройств и систем / И.П. Норенков. – М.: Высш. шк., 1986. – 304 с.

14. Григорьев А.В. Вербальная модель предметной области для интеллектуальных САПР / А.В. Григорьев // Серия: Вычислительная техника и автоматизация. – 2000. – № 20. – С. 171–180.

15. Мейер Д. Теория реляционных баз данных / Д. Мейер : пер. с англ. – М.: Мир, 1987 – 608 с.

Надійшла до редколегії 10.10.2011

О.В. ГРИГОР'ЄВ, Є.В. ПАВЛОВСЬКИЙ
Донецький національний технічний університет

O.V. GRIGORIEV, E.V. PAVLOVSKY
Donetsk national technical university

Представлення сайту мовою HTML як моделі структурного рівня САПР

Submission HTML language site as a model of structural abstraction level CAD

У пропонованій статті розглядається специфіка представлення сайту, виконаного на мові HTML, як типової моделі структурного рівня абстракції САПР. Наводиться приклад виявлення структурної моделі в типовому сайті. Показується, що дана модель структури відповідає можливостям концептуальної моделі предметної області інструментальної оболонки для автоматизації побудови САПР, що дозволяє сформулювати завдання побудови такого роду оболонки.

In this paper we consider the specifics of the submission site, performed in the language of HTML, as a typical model of structural abstraction level CAD. An example of identifying the structural model of a typical site. It is shown that this model corresponds to the structure features of the conceptual domain model tool for automating the construction of the shell of CAD that allows us to formulate the problem of constructing such a shell.

САПР, онтологія, середній експерт, база знань, І-АБО дерево, HTML

CAD, ontology, the average expert knowledge base, the AND-OR tree, HTML