



## МЕТОДОЛОГИЯ СОЗДАНИЯ СИСТЕМЫ УПРАВЛЕНИЯ КОНТЕНТОМ НА БАЗЕ INTERNET-ТЕХНОЛОГИЙ

С.В. Зыков (ООО "Нефтегазовая компания "ИТЕРА")

Представлена методология создания системы управления контентом на базе Internet-технологий, использующая семейство моделей данных для предметной области и среды вычислений, а также инструментальные средства для проектирования индустриальных информационных систем. Рассмотрены этапы создания модели данных предметной области и вычислительной среды.

### Введение

Современные производственные процессы характеризуются все возрастающей значимостью информационно-технологической составляющей. Автоматизированные системы, управляющие производственными процессами и бизнес-проектами, оперируют значительными объемами информации, зачастую разнородной, а также *слабо структурированной*, т.е. включающей аудио-, видео- и фотоматериалы, отсканированные чертежи и документы других типов. С учетом быстрого роста объемов такой информации, создание систем для ее обработки весьма значимо и непосредственно связано с эффективностью производства и конкурентоспособностью бизнеса. В связи с этим актуальной является задача создания методологии проектирования и реализации информационных систем управления контентом. Методология использует как семейство моделей данных для предметной области и среды вычислений, так и инструментальные средства для проектирования индустриальных информационных систем на основе Internet-технологий.

Предлагаемая методология проектирования основана на строгом математическом фундаменте, что позволяет обеспечить более эффективную интеграцию разнородных данных, в том числе и из "унаследованных" прикладных систем. Для динамического моделирования объектов данных используются теории, основанные на понятии функции — лямбда-исчисление [1] и комбинаторная логика [2]. Семантика моделей описывается на основе теории вычислений Д.Скотта, оперирующей правильно построенными языковыми конструкциями [3] и визуализируется при помощи *фреймов* — графов, представляющих элементы знаний о предметной области с точки зрения их доступа и обработки. Технологию управления контентом — данными и метаданными различной степени структурированности — формализуют абстрактные машины с состояниями [4], формирующие значения объектов в среде вычислений.

Методология создания индустриального Internet-ПО включает стадии моделирования предметной об-

ласти и реализации информационной системы управления контентом.

### Общая схема проектирования моделей данных для предметной области и среды вычислений

Каждый этап проектирования моделей данных для предметной области и среды вычислений содержит несколько уровней, включающих универсум рассуждений, объекты данных, а также среду и язык для манипулирования ими [5].

На первом этапе проектирования мир рассуждений представляет собой предметную область в терминах естественного языка (например, управление сайтом), где объектами являются понятия (например, сотрудник, размещающий контент на сайте), а средством манипулирования ими — классическая логика. Описание предметной области представляется на естественном языке (например, в форме ситуации: "Менеджер публикует контент на сайте"). Для описания динамических объектов и их взаимосвязей используется *логика высших порядков*, в которой кроме значений "истина" и "ложь", присутствует еще и значение "неопределенность".

На втором этапе проектирования происходит *концептуализация* — переход к модели данных предметной области со сменой универсума рассуждений на концептуальную

модель [6]. При этом понятия предметной области переходят в *концепты* (т.е. сущности модели, например, "Контент-менеджер" — в MANAGER), соотношения на естественном языке (например, "Менеджер публикует контент на сайте") заменяются ролевыми отношениями (Publish(Manager,Content,Website)), а языком описания становится логика предикатов — функций со значениями "истина" или "ложь". Динамические предметные области описываются логикой высших порядков с использованием семантических *доменов* — множеств, расширенных неопределенным элементом и рекурсивными функциями [3]. Домены формируют ограничения для переменных, моделирующих объекты предметной области. При этом объектная модель базируется на логике высших по-

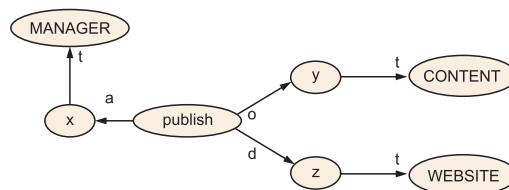


Рис. 1. Ситуативная модель публикации контента в виде фрейма

рядков с тремя упомянутыми выше значениями истинности.

Для визуализации *динамики* (т.е. смены состояний) объектов данных используется язык фреймов. Метаданные включают логические и причинные связи между объектами данных, в том числе типы, а также ограничения целостности, например, в форме диапазонов значений. Уровень *абстракции* (т.е. степени детализации представления предметной области) адекватно отображается фреймовой нотацией. Синтез алгебры фреймов с лямбда-исчислением и теорией категорий позволяет наглядно моделировать все стадии процесса проектирования ПО и управления контентом (рис. 1-3).

В силу громоздкости перечисленных нотаций для "ручного" построения моделей сложных предметных областей разработано инструментальное средство ConceptModeller, которое используется на третьем этапе методологии – первом этапе CASE-проектировании [7] (рис. 2). Универсум рассуждений сменяется на иерархию фреймов. Средство ConceptModeller осуществляет визуализацию проектирования модели предметной области и поддерживает иерархию фреймов, различные их типы (для функций, событий и др.), а также операции, аналогичные принятым в алгебре фреймов. Репозиторий ConceptModeller осуществляет централизованное хранение данных и метаданных – идентификаторов, ограничений целостности и др.

На четвертом этапе (промежуточной стадии CASE-проектирования) визуальное представление модели предметной области транслируется средством ConceptModeller в спецификации стандарта UML (рис.3). При этом универсумом рассуждений является UML-модель в виде семейства взаимосвязанных диаграмм различных типов (сценариев, классов и др.). Строгое математическое основание обеспечивает трансляцию элементов фреймов в UML-диаграммы с сохранением иерархических отношений и ограничений целостности модели предметной области.

#### Модель данных предметной области

Моделирование предметной области производится по схеме двукратного *свертывания*, т.е. установления отношений между классами объектов данных объединенной предметной области, которые в общем случае моделируются с помощью доменов [9].

Описание класса объектов данных представляет собой семейство упорядоченных пар вида *<атрибут,*

*тип>* (например, для размера фотоизображения по вертикали, *<PhotoImageVSize, int>*). Отметим, что в состав атрибутов входят как данные, так и метаданные (размерности объектов, ограничения целостности и др.). К метаданным класса относится и признак принадлежности его объектов к шаблонам ИС управления контентом (например, объект-фотоизображение связан с шаблоном биографии руководителя компании).

При конкретизации класса шаблоном ИС управления контентом, атрибуты метаданных конкретизируются объектами метаданных согласно ограничениям целостности шаблона на число элементов, их типы и т. д. (в частности, формируются размеры и расположение фотоизображения на HTML-странице с биографией руководителя компании). Второе соотнесение конкретизирует шаблон значениями контента, при этом формируется представление статической HTML-страницы (например, биографии Председателя Совета директоров ООО "НГК "ИТЕРА" И.В.Макарова).

Классы объектов строятся с учетом *критериев отбора* в виде предикатов. Двукратное применение соотнесений переводит эти классы сначала в объекты, а затем в значения. Например, класс "Фотоизображение" переводится в объект "Фото руководителя" шаблона HTML-страницы "Биография", а затем – в конкретный графический файл Макаrov.jpg с оцифрованным фотопортретом И.В.Макарова. При этом двунаправленные соотнесения (от классов к значениям и обратно) обеспечивают адекватный *реинжиниринг*, т.е. восстановление модели предметной области по CASE-схеме данных ИС.

Переменные домены строятся как семейства типизированных объектов, полученных из предметной области с применением критериев отбора. Основной принцип моделирования базируется на взаимодействии классов, объектов и значений. Суть его заключается в использовании для описания классов таких критериев отбора, которые однозначно определяют объекты (скажем, позволяют отличить фотопортрет И.В.Макарова от прочих фотоизображений и графических объектов иных типов). Для идентификации объектов данных используется функция вычисления значения, переводящая классы языка моделирования в классы предметной области. Таким образом, формируется соответствие между объектами предметной области и языком описания объектов данных на уровне модели.

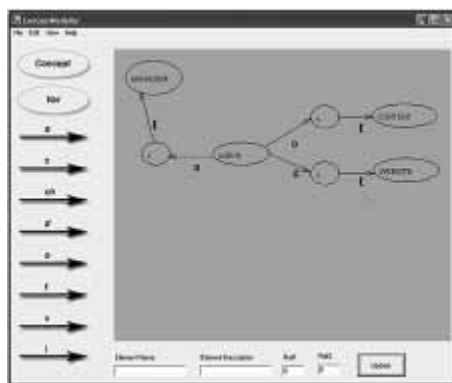


Рис. 2. Представление ситуативного фрейма в инструментальном средстве ConceptModeller

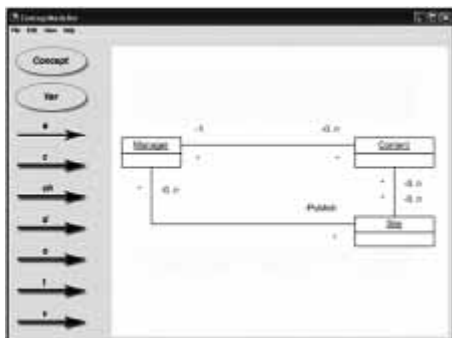


Рис. 3. Представление UML-диаграммы в инструментальном средстве ConceptModeller

*Большинство писателей-фантастов не любят Internet-технологии, так как в их описаниях будущего Internet-а просто нет*

И. Свиблов

#### Модель данных вычислительной среды

В качестве модели управления объектами корпоративного хранилища данных и метаданных создана абстрактная машина для управления контентом на основе состояний (таких, как "новая HTML-страница", "опубликованная HTML-страница" и т.д.). Динамика абстрактной машины формализуется явным перечислением всевозможных смен состояний. С точки зрения модели при отображении шаблонов в содержимое HTML-страниц, формируемых абстрактной машиной, производится связывание переменных, характеризующих элементы шаблона (например, элемент "Фото руководителя" шаблона "Биография"), с их значениями ("http://www.itera.ru/contentimg/Makarov.jpg").

Семантика абстрактной машины для управления контентом основана на теории семантических доменов Д.Скотта [3]. Типы атомарных шаблонов получают из *стандартных* (наиболее часто используемых) доменов (таких, как "Заголовок", "Текст", "Фото"), а типы более сложных шаблонов строятся посредством *конструкторов* — операций получения новых доменов на основе имеющихся. Формальная семантика абстрактной машины для управления контентом включает стандартные и *конечные* (содержащие явно перечислимые элементы) домены, а также их конструкторы: функциональное пространство, декартово произведение, последовательность и дизъюнктивную сумму.

Язык абстрактной машины содержит выражения: константы, идентификаторы, *присваивание* (означивание шаблона, например "Биография", контентом — файлом с фотопортретом И.В.Макарова) и другие операции, а также множество команд, включая сравнение и последовательное исполнение. Синтаксис языка абстрактной машины определяется описанием синтаксических доменов идентификаторов, команд и выражений.

Домен состояния абстрактной машины определяется состоянием памяти с учетом значений на входе (контент) и выходе (HTML-страница) ИС управления контентом, где память понимается как отображение из домена идентификаторов (например, "Фотопортрет И.В.Макарова") в домен значений (например, http://www.itera.ru/contentimg/Makarov.jpg). Домен значений представляет собой дизъюнктивную сумму доменов, содержащих типы контента, существующие в языке абстрактной машины.

Семантические предложения описывают значения *денотатов* — правильно построенных конструкций — языка управления контентом. Денотатом идентифика-

тора является идентификатор, связанный со значением в форме упорядоченной тройки вида *<значение в памяти, идентификатор, состояние>*, например, *<"Фотопортрет И.В.Макарова", "http://www.itera.ru/contentimg/Makarov.jpg", "Опубликовано">*. Причем изменения состояния абстрактной машины не происходит.

Вычисление значения выражения в среде абстрактной машины приводит к связыванию переменной со значением контента. При этом, как и в случае с командой, происходит изменение состояния абстрактной машины. Выполнение присваивания приводит к смене состояния с подстановкой в память значения контента (т.е. содержимого HTML-страницы сайта, например, файла "Makarov.jpg") вместо идентификатора шаблона (например "Биография").

Исключительные ситуации (ошибки) в работе абстрактной машины управления контентом, например, в случае несоответствия типов идентификатора и значения (скажем, "Фотоизображение" и "Текст"), моделируются с помощью значения "неопределенность".

Таким образом, инструментальные средства, используемые в процессе моделирования предметной области, позволяют оперативно дополнять, развивать и модифицировать семейство моделей в процессе эксплуатации системы управления контентом. Семейство моделей данных для предметной области и вычислительной среды являются входными данными для информационной системы управления контентом, проектирование и реализация которой также является частью предложенной методологии. В общем случае информационная система управления контентом на базе Internet-технологий является инвариантной к используемой предметной области.

#### Список литературы

1. *Барендрегт Х.* Лямбда-исчисление, его синтаксис и семантика. М.: Мир. 1985.
2. *Curry H.B., Feys R.* Combinatory logic, Vol. 1.- North Holland, Amsterdam. 1958.
3. *Scott D.S.* Lectures on a mathematical theory of computations. Oxford University Computing Laboratory Technical Monograph. PRG-19. 1981.
4. *Cousineau G., Curien P.-L., Mauny M.* The categorical abstract machine // Science of Computer Programming 8(2). 1987.
5. *Zykov S.V.* Integrated Methodology for Internet-Based Enterprise Information Systems Development // Proceedings of the 1st International Conference on Web Information Systems and Technologies (WEBIST2005).- Setubal: INSTICC Press. 2005.
6. *Wolfengagen V.E.* Event Driven Objects // Proceedings of the 1st Workshop on Computer Science and Information Technologies (CSIT'99). М.:МЭФИ. 1999.
7. *Zykov S.V.* ConceptModeller: a Problem-Oriented Visual SDK for Globally Distributed Enterprise Systems // Proceedings of the 7th International Workshop on Computer Science and Information Technologies (CSIT'2005), Vol.1. Ufa: USATU. 2005.

*Зыков Сергей Викторович — канд. техн. наук, ведущий специалист ООО "НГК "ИТЕРА". Контактный телефон(495) 411-85-00. E-mail: szykov@ngk.itera.ru*