МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

им. М. В. Ломоносова

ФАКУЛЬТЕТ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И КИБЕРНЕТИКИ

КАФЕДРА АВТОМАТИЗАЦИИ СИСТЕМ ВЫЧИСЛИТЕЛЬНЫХ КОМПЛЕКСОВ

**Семантическое отображение канонической информационной модели СИНТЕЗ в формальный язык спецификаций AMN**

*Дипломная работа*

*студента А. Ю. Сергеева*

**Руководители:**

*профессор, д. ф.-м. н. Л. А. Калиниченко*

*профессор, д. т. н. В. А. Сухомлин*

Москва 2010

Оглавление

[Аннотация 3](#_Toc261537203)

[Введение 4](#_Toc261537204)

[Постановка задачи 8](#_Toc261537205)

[Язык СИНТЕЗ 10](#_Toc261537206)

[Язык AMN 12](#_Toc261537207)

[Трансформация моделей и язык ATL 14](#_Toc261537208)

[Список литературы 16](#_Toc261537209)

## Аннотация

В данной дипломной работе рассматривается отображение

## Введение

В настоящее время в различных областях деятельности людей и науки в частности наблюдается экспоненциальный рост накапливаемых, экспериментальных данных, описанных с помощью всевозможных информационных моделей. Таким образом нынешний период развития информационных технологий сопровождается созданием множества различных информационных моделей. Наряду с созданием моделей для распределенных систем (таких как архитектуры OMG, архитектуры Семантического Веба, архитектуры цифровых библиотек, архитектуры информационных грид систем), можно выделить разработку в области конкретных информационных моделей - моделей данных (таких как SQL, UML, XML), моделей потоков работ (MQSeries, SAP/R3, Eastman), семантических моделей (включая онтологические модели и модели метаданных). В основе почти всех из них находятся разнообразные понятия и парадигмы, зачастую несовместимые между собой. Также можно выделить другую тенденцию, заключающуюся в накапливании основанных на данных моделях информационных ресурсах, количество которых непрерывно растет. На этом фоне возникает необходимость в использовании и интеграции компонентов и сервисов, представленных в различных моделях и приложениях, также как и их повторное использование для создания новых информационных систем. Можно заметить противоречивость такого рода тенденций, так как чем больше появляется всевозможных моделей наряду с уже использующимися, тем труднее становится задача их интеграции между собой. Данная проблема вынуждает разработчиков новых систем создавать новые ресурсы, вместо грамотного повторного использования уже существующих.

В течение последних лет в лаборатории композиционных методов проектирования ИС ИПИ РАН проводились интенсивные исследования по разработке методов решения задач над неоднородными информационными ресурсами, таких как композиционное проектирование ИС [2], регистрация ресурсов в предметных посредниках [3], интеграция моделей ресурсов [4].

Представленные выше методы нашли свое отражение в ряде проектов лаборатории. В частности, была разработана многоуровневая инфраструктура, включающая в себя уровень информационных ресурсов, промежуточный слой, обеспечивающий интероперабельность ресурсов благодаря технической унификации их интерфейсов и введению дистанционных механизмов обращения к ресурсам (адаптеры, осуществляющие преобразование запросов, выраженных в канонической информационной модели посредников, в их представление в информационной модели ресурса, включаются в состав промежуточного слоя), уровень *предметных посредников*, каждый из которых создает спецификацию предметной области для решения некоторого класса задач, используя каноническую информационную модель для представления семантики предметной области и унифицированного отображения разнообразных видов информационных моделей ресурсов (моделей данных, сервисных моделей, онтологических моделей, процессных моделей), уровень задач (приложений), формулируемых в терминах одного или нескольких посредников. Следует заметить, что посредники, в свою очередь, могут рассматриваться как информационные ресурсы для использования в посредниках более высокого уровня. В такой инфраструктуре предметные посредники играют ключевую роль для решения семантических проблем, перечисленных выше. В частности, при интеграции неоднородных ресурсов в посреднике нужно уметь семантически отождествлять объекты, представленные в различных информационных моделях, и семантически правильно отображать схемы интегрируемых

ресурсов в схему посредника. Поскольку в общем случае ресурсы неоднородны (представлены в различных моделях), при интеграции неоднородных ресурсов для однородного представления их семантики требуется приведение различных информационных моделей к унифицированному виду в рамках некоторой унифицирующей информационной модели, которая называется *канонической*. Для унификации разнородных спецификаций прежде всего требуется умение сопоставлять спецификации различных ресурсов друг с другом так, чтобы можно было отвечать на вопрос, можно ли при реализации посредника использовать спецификацию существующего ресурса вместо фрагмента спецификации посредника. Для этого достаточно доказать, что рассматриваемые спецификации находятся в отношении уточнения. Говорят, что спецификация A уточняет спецификацию D, если A можно использовать вместо D так, что пользователь D не будет замечать этой замены. Средства доказательства факта уточнения спецификации некоторого компонента спецификацией другого компонента (реализуемые на основе теоретико-модельных нотаций и соответствующего инструментария [6,7]) составляют фундамент применяемых методов конструирования унифицирующих (канонических) моделей представления информации в посредниках. Каноническая информационная модель служит в качестве общего языка для адекватного выражения семантики разнородных моделей представления информации, используемых в разнообразных информационных ресурсах, релевантных посреднику. Методы отображения информационных моделей и синтеза расширяемых канонических информационных моделей для слоя предметных посредников подробно рассмотрены в [7]. Там же предложена архитектура Унификатора информационных моделей. Задачей Унификатора моделей является унификация множества информационных моделей(называемых *исходными*), использующихся в различных информационных ресурсах некоторой информационной системы. Унификация исходной модели R есть приведение ее к канонической информационной модели C, т. е. создание такого расширения E канонической модели (которое может быть и пустым) и такого отображения M исходной модели в расширенную каноническую, что исходная модель *уточняет* расширенную каноническую модель. Уточнение моделей означает, что для любой допустимой спецификации r модели R ее образ M(r) при отображении M уточняется спецификацией r. Синтез канонической модели заключается в построении указанных расширений ядра канонической модели для всех исходных моделей и объединении таких расширений. В качестве ядра канонической информационной модели применяется язык СИНТЕЗ [1], который ориентирован на спецификацию предметных посредников и синтез канонических моделей. Спецификация предметного посредника для класса задач включает определения понятий предметной области, выражаемых соответствующими онтологическими спецификациями, спецификации классов объектов предметной области, спецификации типов экземпляров названных классов и их методов, определяющих их поведение, спецификации процессов решения задач данного класса как совмещенных во времени последовательностей действий, реализуемых методами классов, сервисами и другими процессами. Предполагается, что такие спецификации преобразуются в спецификации канонической модели, имеющей формальную семантику.

Регистрация релевантных посреднику ресурсов рассматривается как задача композиционного проектирования систем [2, 3]. Регистрация ресурсов есть процесс целенаправленной трансформации спецификаций, включающий декомпозицию спецификаций посредника на непротиворечивые фрагменты, поиск среди спецификаций релевантных ресурсов подходящих типов данных — кандидатов для уточнения ими спецификаций типов посредника, построение выражений, определяющих классы ресурсов в виде композиции классов посредника. Для подобного манипулирования спецификациями разработано специальное *исчисление спецификаций* [4].

Принципиальным моментом в этой схеме является реализация доказательства уточнения фрагментов спецификаций посредника спецификациями ресурсов в процессе отображения таких спецификаций.

## Постановка задачи

Таким образом, можно заметить, что необходимым требованием к разработанным методам и средствам является обеспечение возможности автоматизированного доказательства факта уточнения спецификации требований спецификацией существующих компонентов. Для таких манипуляций спецификациями требуется выражение их семантики в некотором формальном языке, позволяющем осуществлять доказательство непротиворечивости уточнения спецификаций. В качестве такого языка выбрана Нотация Абстрактных Машин (Abstract Machine Notation, AMN [6]), которая позволяет осуществлять необходимые рассуждения относительно спецификаций. Для AMN разработаны специальные инструментальные средства, составляющие так называемую В-технологию [7]. Для того, чтобы можно было манипулировать спецификациями информационных моделей в рамках В-технологии (в частности, доказывать корректность отношения уточнения между спецификациями, выраженными в различных моделях), необходимо корректно отобразить такие модели в язык AMN с сохранением семантики. Метод такого отображения создан [16] и существует возможность применять его на примере ядра канонической информационной модели (языка СИНТЕЗ).

В настоящее время в Лаборатории композиционных методов проектирования информационных систем разрабатываются прототипы инструментальные средства для регистрации ресурсов в предметных посредниках, а также средства унификации информационных моделей. В качестве объединяющей платформы для этих средств выбран Eclipse [17]. Задачей данной дипломной работы является программная реализация семантического отображения канонической информационной модели СИНТЕЗ в формальный язык AMN, с использованием языка трансформации моделей ATL (являющегося частью Eclipse Modeling Project). Компонент, реализующий отображение, рассматривается как неотъемлемая часть архитектуры упомянутых средств интеграции неоднородных ресурсов.

В нижеследующих разделах рассматриваются основные черты языков, используемых в дипломной работе: СИНТЕЗ, AMN, ATL.

## Язык СИНТЕЗ

В данной работе в качестве ядра канонической модели был выбран язык СИНТЕЗ [1], разработанный в Лаборатории композиционных методов проектирования информационных систем Института проблем информатики РАН. Он ориентирован на семантическую интероперабельность и композиционное проектирование информационных систем в широком диапазоне существующих неоднородных информационных компонентов. Метаязыком, при помощи которого в ядре канонической информационной модели представляются средства определения информационных источников, является язык фреймов. Фрейм - это структурированная символическая модель некоторого понятия. Фрейм можно рассматривать как множество атрибутов, называемых слотами, каждый из которых используется для представления свойства понятия. Каждый слот может иметь множество значений. Единицей определения канонической модели является модуль. Каждый модуль задает обобщенное представление информационных источников, либо является модулем спецификации предметной области или концептуального проекта информационной системы. Каноническая модель содержит унифицированную систему типов, включающую универсальный конструктор типов (Абстрактный Тип Данных, АТД), а также представительный набор встроенных типов. АТД по умолчанию является объектным, хотя он также может быть объявлен и как необъектный. Описание абстрактного типа данных инкапсулирует спецификации атрибутов, методов и инвариантов типа. Методы и инварианты описываются встроенным типом функции. Типы языка организованы в иерархию обобщения, базирующуюся на отношении тип-подтип. Значение подтипа может использоваться всюду, где ожидается значение супертипа; подтип наследует элементы спецификации супертипа; допускается множественное наследование спецификаций подтипом. Отношение тип-подтип основывается на понятии уточнения. Спецификация типа функции включает описание параметров функции и предикативную спецификацию функции. Для задания предикативных спецификаций функций в канонической модели используется типизированный вариант логики предикатов первого порядка. Предикативная спецификация позволяет задавать смешанные и постусловия, определяющие действия, реализуемые функцией. Для обращения к информационным источникам в языке предусмотрены формулы специального вида - правила. В языке правил не разрешается использование конструкций, связанных с изменением состояния системы, однако допускается использование широкого спектра операций композиции источников, характерных для языков запросов: например, соединение, пересечение, объединение.

Однородные совокупности объектов предметной области (информационные источники) представляются в канонической модели коллекциями. Экземплярами коллекций являются значения АТД. В случае если экземпляр типа объектный, коллекция называется классом. Явно разграничивая объявление типа и объявление коллекции, язык СИНТЕЗ подчеркивает роль коллекции как представителя множества объектов и роль типа как спецификации структуры и поведения объектов, связанных с коллекцией.

## Язык AMN

Формальная семантика необходима для проведения доказательных рассуждений о моделях информационных ресурсов. Имея формальную семантику можно строго формулировать утверждения о непротиворечивости и уточнении спецификаций, выраженных в модели. Появляется возможность доказательства корректности отображений одних моделей информационных ресурсов в другие модели. Для таких манипуляций информационными моделями и выраженными в них спецификациями требуется выражение их семантики в некотором формальном языке, позволяющем осуществлять доказательство непротиворечивости и уточнения спецификаций. Таким языком может являться теоретико-модельный язык Нотация Абстрактных Машин (Abstract Machine Notation, AMN [6]), основанный на теории множеств и логике предикатов первого порядка, который позволяет осуществлять необходимые рассуждения относительно спецификаций. Язык AMN обеспечивает манипулирование теоретико-множественными спецификациями в логике первого порядка и дает возможность доказать правильность уточнения спецификаций. Для AMN разработаны специальные инструментальные средства, составляющие в совокупности так называемую В-технологию[7]. Язык AMN как теоретико-множественная нотация позволяет рассматривать состояния и операции системы интегрировано как спецификацию пространства состояний и поведения (определенного операциями на состояниях) абстрактных машин. Спецификация состояния абстрактной машины вводится переменными состояния вместе с инвариантами - ограничениями, которые должны всегда удовлетворяться. Операции определяются на основе расширения формализма охраняемых команд Дейкстры. Ключевым понятием AMN является уточнение, позволяющее соотносить спецификации систем различных уровней абстракции. Уточняющая спецификация может быть гораздо более детальной, чем уточняемая спецификация. Конструируется уточняющая спецификация на основе алгоритмического уточнения и уточнения данных. Уточнение формализуется в AMN путем формулировки ряда теорем специального вида, так называемых *proof obligations*. Такие теоремы формулируются автоматически при помощи инструментальных средств поддержки В-технологии на основании склеивающих инвариантов, соотносящих состояния уточняемой и уточняющей системы.

## Трансформация моделей и язык ATL

Для трансформации моделей созданы специальные языки, которые развиваются в контексте *Движимой моделями архитектуры* (Model-Driven Architecture, MDA [8]) - подхода, поддерживаемого стандартом OMG MOF (Meta-Object Facility) [12]. Базовыми составляющими MDA являются модели. Они рассматриваются как первичные сущности, и наиболее важными операциями манипулирования моделями становятся преобразования моделей, отображения моделей из одной в другую.

Модель определяется в соответствии с семантикой некоторой метамодели, при этом говорят, что модель конформна метамодели. Стандартом MOF определена четырехуровневая архитектура моделей: модели уровня М0 описывают объекты реального мира, модели уровня М1 называются обычно *схемами*, модели уровня М2 представляют собой собственно *информационные модели* (например, в архитектурах систем управления данными представляющие собой совокупность языка определения данных и языка манипулирования данными), модели уровня М3 – метаметамодели, предназначенные для описания моделей уровня М2.

Примером языка трансформации моделей может являться QVT (Query-View-Transformation) [13].

В качестве языка трансформации моделей в данной работе рассматривается язык ATL (ATLAS Transformation Language) [9], разработанный научной группой ATLAS INRIA & LINA. ATL является ответом на RFP (Request for Proposal) языка QVT. Система типов и операций над типами языка ATL очень близка (но не тождественна) системе типов языка OCL [10]. Для языка ATL на базе платформы Eclipse реализована интегрированная среда разработки, называемая ATL Development Tools (ADT). В качестве модели уровня М3 рассматривается модель *Ecore* [11].

Трансформация представляет собой модуль языка ATL. Модуль состоит из заголовка (включающего имя модуля и имена переменных, соответствующих исходной и целевой моделям) и множества правил, определяющих способ построения элементов целевой модели на основе элементов исходной модели. Так называемые *сопоставляющие правила* (matched rules), составляющие ядро трансформации, позволяют описывать:

* какой элемент исходной модели должен быть взят;
* количество и тип порождаемых элементов целевой модели;
* способ, при помощи которого элементы целевой модели инициализируются на основании элементов исходной модели.

Для построения трансформации языка СИНТЕЗ в язык AMN, абстрактные синтаксисы обоих языков представляется в виде моделей уровня М2, конформных модели *Ecore*. Трансформация позволит осуществлять преобразование произвольных моделей уровня М1, конформных модели СИНТЕЗ уровня М2, (т.е. спецификаций языка СИНТЕЗ) в соответствующие модели уровня М1, конформные модели AMN уровня М2, (т.е. спецификации языка AMN).

## Построение модели языка AMN

Одной из основных задач данной работы является построение модели языка AMN. Согласно условиям трансформации моделей данная модель должна быть конформна модели *Ecore* и быть уровня M2 согласно стандарту MOF.

Основной синтаксической конструкцией языка являются компоненты. Основным элементом этого класса, использующимся при построении трансформации, является абстрактная машина. Она включает в себя секции определения данных, их свойств, а также операций. Порядок секций может быть произвольным. Ниже дается краткое описание секций абстрактной машины.

Таблица 1 – Секции абстрактной машины

|  |  |
| --- | --- |
| CONSTRAINTS | Определение типа и свойств формальных скалярных параметров |
| SEES | Список экземпляров видимых машин |
| INCLUDES | Список экземпляров включенных машин |
| PROMOTES | Список используемых операций включенных абстрактных машин |
| EXTENDS | Список экземпляров расширяемых машин |
| USES | Список экземпляров используемых машин |
| SETS | Абстрактные множества, либо определение перечислимых множеств |
| CONCRETE\_CONSTANTS | Список конкретных констант |
| ABSTRACT\_CONSTANTS | Список абстрактных констант |
| PROPERTIES | Тип и свойства машинных констант |
| CONCRETE\_VARIABLES | Список конкретных переменных |

Продолжение таблицы 1

|  |  |
| --- | --- |
| ABSTRACT\_VARIABLES | Список абстрактных переменных |
| INVARIANT | Тип и свойства переменных |
| ASSERTIONS | Определение свойств, выводимых из инварианта |
| INITIALISATION | Инициализация переменных |
| OPERATIONS | Список и определение операций |

Секции IMPORTS, SEES, INCLUDES, EXTENDS, или USES являются секциями видимости. С их помощью строятся связи между различными компонентами системы.

Составляющими элементами секций PROPERTIES, INVARIANT, ASSERTIONS являются предикаты. В языке AMN предикат является формулой, которая может быть доказана либо опровергнута. Предикаты используются для выражения свойств данных или для задания условий применения подстановок.

Все многообразие предикатов языка может быть разбито на два класса:

* атомарный предикат
* составной предикат

Атомарный предикат состоит из выражений, объединенных каким-либо условием.

Составной предикат состоит из предикатов, объединенных отрицанием, конъюнкцией, дизъюнкцией, импликацией либо эквивалентностью.

Секция INITIALISATION представляет собой подстановку, инициализирующую переменные абстрактной машины.

Подстановки представляют собой математические нотации, преобразующие предикаты. Пусть *S* – подстановка, а *P* – предикат. Тогда запись *S* [ *P* ] представляет предикат, полученный после трансформации *P* подстановкой *S*. Подстановки позволяют моделировать динамический аспект В-технологии: позволяют установить, как свойства данных преобразуются операциями компонента. Механизм трансформации предиката подстановкой позволяет генерировать теоремы специального вида - *proof obligations –* относящиеся к инициализации и операциям. Например, для того, чтобы абстрактная машина была семантически корректна, необходимо доказать, что каждая операция машины сохраняет инвариант.

Выражение является основным конструкторским элементом языка. Оно является формулой, которая определяет элемент данных. В языке AMN выражения разбиты на семейства, объединенные по типу данных использующихся в выражении элементов. Например, простейшие выражения, арифметические выражения, выражения над множествами и т.д.

При моделировании было решено абстрагироваться от стандартной классификации и объединить выражения по более общему признаку. Было выделено три больших класса выражений, объединенных общими свойствами:

* унарный оператор
* бинарный оператор
* функциональное выражение

Также был выделен класс простейших константных выражений: численных, булевых, строковых, множеств.

## Список литературы

1. *Kalinichenko L.A., Stupnikov S.A., Martynov D.O.* SYNTHESIS: a Language for Canonical Information Modeling and Mediator Definition for Problem Solving in Heterogeneous Information Resource Environments. Moscow: IPI RAN, 2007. - 171 p.

2. *Briukhov D.O., Kalinichenko L.A.* Component-based information systems development tool supporting the SYNTHESIS design method // Advances in Databases and Information Systems: Proc. of the Second East European Conference. – Berlin-Heidelberg: Springer-Verlag, 1998. – P. 305-327.

3. *Briukhov D.O., Kalinichenko L.A., Skvortsov N.A.* Information sources registration at a subject mediator as compositional development // Advances in Databases and Information Systems: Proc. of the 5th East European Conference. – Berlin-Heidelberg: Springer-Verlag, 2001. – P. 70-83.

4. *Kalinichenko L.A.* Method for Data Models Integration in the Common Paradigm // Advances in Databases and Information Systems: Proc. of the First East-European Conference. – St. Petersburg: Nevsky Dialekt, 1997. – P. 275-284.

5. *Kalinichenko L.A., Martynov D.O., Stupnikov S.A.* Query rewriting using views in a typed mediator environment // Advances in Databases and Information Systems: Proceedings of the East European Conference. – Berlin-Heidelberg: Springer-Verlag, 2004. – P. 37-53.

6. *Abrial J.-R.* The B-Book: Assigning Programs to Meanings // Cambridge: Cambridge University Press, 1996.

7. *Abrial J.-R.* B-Technology// Technical overview. – B-Core (UK) Ltd., 1993.

8. OMG Model Driven Architecture // http://www.omg.org/mda/

9. ATL Project // <http://www.eclipse.org/m2m/atl/>

10. OMG/OCL Object Constraint Language (OCL) 2.0. OMG Final Adopted Specification. ptc/03-10-14, 2003.

11. *Budinsky, F., Steinberg, D., Ellersick, R., Grose, T.* Eclipse Modeling Framework, Chapter 5 ”Ecore Modeling Concepts”. - Addison Wesley Professional, 2004.

12. Meta Object Facility (MOF) 2.0 Core Specification // <http://www.omg.org/cgi-bin/apps/doc?ptc/03-10-04.pdf>, 2003.

13. Meta Object Facility (MOF) 2.0 Query/View/Transformation Specification // http://www.omg.org/cgi-bin/doc?ptc/2007-07-07, 2007.

14. *Kalinichenko L.A., Briukhov D.O., Martynov D.O., Skvortsov N.A., Stupnikov S.A.* Mediation Framework for Enterprise Information System Infrastructures. Proc. of the 9th International Conference on Enterprise Information Systems ICEIS 2007. -- Funchal, 2007. -- Volume Databases and Information Systems Integration. -- P. 246--251.

15. *Захаров В.Н., Калиниченко Л. А., Соколов И. А., Ступников С. А.* Конструирование канонических информационных моделей для интегрированных информационных систем // Информатика и её применения, 2007. Т. 1. Вып. 2. С. 15–38.

16. *Stupnikov S.A.* Mapping of canonical model core specifications in Abstract Machine Notation // Formal Methods and Models for Compositional Infrastuctures of Distributed Information Systems: The Systems and Means of Informatics, Special Issue. -- Moscow: IPI RAN, 2005. -- P. 69--95

17. Eclipse // http://www.eclipse.org/