

Григорьев М.В.
Тюменский государственный университет
Кафедра информационных систем
Доцент, к.т.н.
mgrigoriev@utmn.ru

Гаркуша Н.А.
Тюменский государственный университет
Кафедра иностранных языков
и межкультурной
профессиональной коммуникации
естественнонаучных направлений
Доцент, к.п.н.
Sochy2001@yandex.ru

**ПОДХОДЫ К АНАЛИЗУ ОПИСАНИЙ И СПЕЦИФИКАЦИЙ
ПРЕДМЕТНЫХ ОБЛАСТЕЙ В ДЕЯТЕЛЬНОСТИ АРХИТЕКТОРА БАЗ
ДАННЫХ**

**SOME APPROACHES TO ANALYSIS OF APPLICATION DOMAIN
DESCRIPTION IN DATABASE ARCHITECT'S ACTIVITY**

В деятельности архитектора баз данных процесс анализа для идентификации объектов и классов объектов признан одним из самых сложных областей объектно-ориентированной разработки.

Подходы к анализу текстов

Существует ряд различных подходов, предлагаемых исследователями к распознаванию объектов, в том числе:

- использование грамматического анализа для описания системы на естественном языке (Abbott, 1983; Robinson, 1992 и т.д.);

- использование значимых элементов предметной области (Shlaer и Mellor, 1988, Coad и Yourdon, 1990, Wirfs-Brock и др., 1990);
- использование поведенческого подхода (Rubin и Goldberg, 1992);
- использование анализа на основе сценариев (Beck и Cunningham, 1989).

Основные принципы текстового анализа были разработаны Расселом Дж. Абботом в статье «Разработка программы посредством неформального описания на английском языке» (“Program Design by Informal English Descriptions”) [5].

По мнению Г. Буча, Аббот предложил «чрезвычайно простой метод», основанный на текстовом описании.

Аббот представил идею о том, как выделить типы данных, переменные, операторы и управляющие структуры из текста на естественном языке для разработки программ на языке программирования Ада. В своей статье он пишет: «Мы определяем типы данных, объекты, операторы и управляющие структуры, глядя на английские слова и фразы в неформальной стратегии» [5,884].

Согласно методу Аббота необходимо описать задачу или ее часть на простом английском языке, а потом подчеркнуть существительные и глаголы. Существительные – кандидаты на роль классов, глаголы могут стать именами операций [5, 128].

1. Имя существительное нарицательное в неформальной стратегии предполагает тип данных (data type).

2. Имя существительное собственное или прямая ссылка означает объект (object).

3. Глагол, определение, предикат, или описательные выражения предполагают оператор (operator).

4. Управляющие структуры выделяются без труда по-английски.

Глаголы и существительные могут быть дополнительно подразделены.

В таблице 1 представлено общее описание подхода Аббота к текстовому анализу.

Текстовый анализ
(Рассел Дж. Аббот)

Part of Text Часть текста	Model Component Компонент модели	Example Пример
Proper Noun Имя существительное собственное	Instance, Object Экземпляр класса объекта, Объект	J. Smith, Euro Дж. Смит, Евро
Common Noun Имя существительное нарицательное	Class, Type, Role Класс, Тип, Роль	toy, currency, seller игрушка, валюта, продавец
Doing Verb Глагол действия	Operation Действие	buy покупать
Being Verb Глагол be	Classification Классификация	is an
Having Verb Глагол have	Composition Композиция	has an
Stative Verb Глагол состояния	Invariance-Condition Инвариантность-состояние	are owned
Modal Verb Модальный глагол	Data Semantics, Pre Condition, Post Condition or Invariance Condition Семантика данных, предусловие, постусловие, условие инвариантности	must be
Adjective Имя прилагательное	Attribute Value or Class Значение атрибута или класса	unsuitable
Adjective Phrase Конструкция, имеющая свойства прилагательного	Association, Operation Ассоциация, операция	The customer with children, the customer who bought the kite
Transitive Verb Переходный глагол	Operation действие	enter
Intransitive Verb Непереходный глагол	Exception or Event Исключение или событие	depend

Подход Аббота полезен, так как он прост и заставляет разработчика заниматься словарем предметной области [5,128].

Нотации моделирования данных

Важнейшим этапом разработки прикладной системы является построение концептуальных моделей, как можно более полно описывающих особенности предметной области, характер решаемых задач, информационные потребности и ресурсы, технологические ограничения и т.д. В результате должны быть построены модели двух типов - информационная, отражающая существующие информационные структуры и взаимосвязи между ними, и функциональная, описывающая технологию и способы обработки информации, используемые в данной области.

В качестве стандартного средства информационного используется в том или ином виде аппарат моделей «сущность-связь» или ER-моделей (Entity Relationship Model). Этот формализм позволяет представлять информационные потребности в виде, наглядном и удобном для восприятия, что делает их хорошим средством коммуникации между проектировщиками и пользователями в процессе уточнения постановки задач [2].

Теоретической основой этого подхода является известная модель «сущность – связь», введенная Ченом в 1976 году [4] и получившая широкое развитие и распространение в качестве средств концептуального проектирования баз данных.

Проведем обзор некоторых нотаций моделирования данных, используемых при построении диаграмм «сущность-связь».

Нотация Питера Чена

Модель «сущность-связь» была предложена в 1976 году Питером Пин-Шен Ченом (англ. Peter Pin-Shen Chen) [4] – американским профессором компьютерных наук в университете штата Луизиана. Фактически Чен не изобретал модель, он взял идеи из более ранних работ таких практиков, как А. Браун [6] и других. Однако Питер Чен сделал больше, чем кто-бы то ни было до него для формализации и популяризации ER-модели, а так же для её внедрения в научную литературу.

В нотации Питера Чена сущность представляет собой множество экземпляров реальных или абстрактных объектов (людей, событий, состояний, идей, предметов и т.п.), обладающих общими атрибутами или характеристиками. Любой объект системы может быть представлен только одной сущностью, которая должна быть уникально идентифицирована. При этом имя сущности должно отражать тип или класс объекта, а не его конкретный экземпляр.

Отношение в самом общем виде представляет собой связь между двумя и более сущностями. Именованное отношение осуществляется с помощью грамматического оборота глагола (имеет, определяет, может владеть и т.п.).





Другими словами, сущности представляют собой базовые типы информации, хранимой в базе данных, а отношения показывают, как эти типы данных взаимосвязаны друг с другом. Введение подобных отношений преследует две основополагающие цели:







- обеспечение хранения информации в единственном месте (даже если она используется в различных комбинациях);
- использование этой информации различными приложениями.

Основные элементы, используемые при построении информационно-логической модели по нотации Чена, представлены в таблице 2.

Таблица 2

Описание элементов нотации Питера Чена

Элемент диаграммы	Обозначает
	независимая сущность
	зависимая сущность
	родительская сущность в иерархической связи
	Связь

	идентифицирующая связь
	Атрибут
	первичный ключ
	внешний ключ (понятие внешнего ключа вводится в реляционной модели данных)
	многозначный атрибут
	получаемый (наследуемый) атрибут в иерархических связях

Связь соединяется с ассоциируемыми сущностями линиями. Возле каждой сущности на линии, соединяющей ее со связью, цифрами указывается класс принадлежности (Рис. 1).

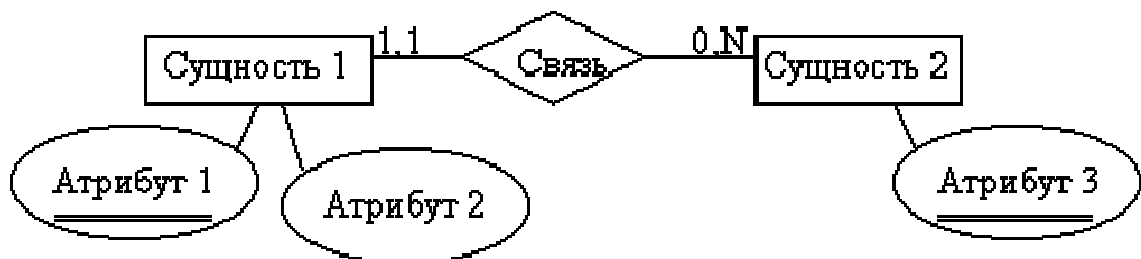


Рисунок 1 – ER-модель по нотации Питера Чена

Нотация Мартина

Основные элементы, используемые при построении информационно-логической модели по нотации Мартина [3], представлены в таблице 3.

Таблица 3


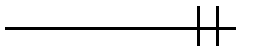
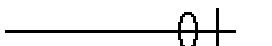



Описание элементов по нотации Мартина

Элемент диаграммы	Обозначает
	независимая сущность
	зависимая сущность
	родительская сущность в иерархической связи

Список атрибутов приводится внутри прямоугольника, обозначающего сущность. Ключевые атрибуты подчеркиваются. Связи изображаются линиями, соединяющими сущности, вид линии в месте соединения с сущностью определяет кардинальность связи (таблица 4).

Таблица 4

Описание связей в нотации Мартина

Обозначение	Кардинальность
	нет
	1,1
	0,1
	M,N
	0,N
	1,N

Имя связи указывается на линии ее обозначающей (Рисунок 2).

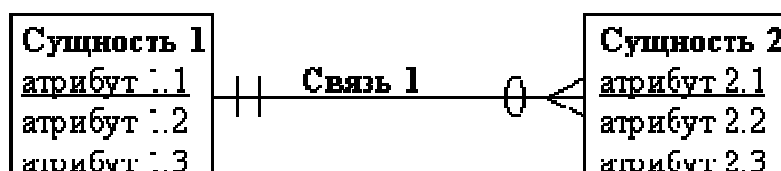


Рисунок 2 – ER-модель по нотации Мартина






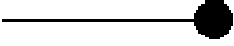
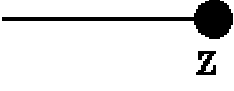
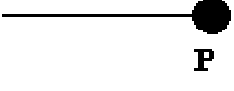
Нотация IDEF1X

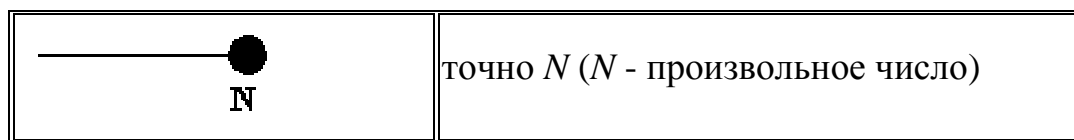
Согласно нотации IDEF1X, список атрибутов приводится внутри прямоугольника, обозначающего сущность. Атрибуты, составляющие ключ сущности, группируются в верхней части прямоугольника и отделяются горизонтальной чертой [1].

Обозначения сущностей, связей и кардинальных связей по нотации IDEF1X представлены в таблицу 5.

Таблица 5

Описание сущностей, связей по нотации IDEF1X

Элемент диаграммы	Обозначает
Обозначения сущностей:	
	независимая сущность
	зависимая сущность
Обозначения связей:	
	идентифицирующая связь
	неидентифицирующая связь
Обозначение кардинальности связей:	
	1,1
	0,М
	0,1
	1,М



ER-модель по нотации IDEF1X представлена на рисунке 3.

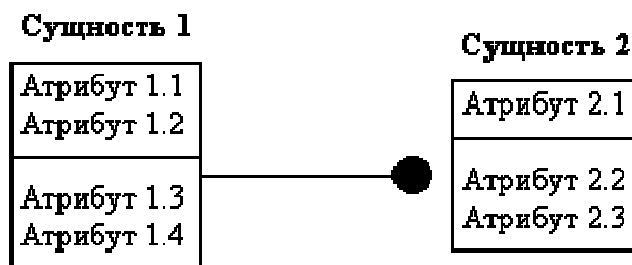


Рисунок 3 – ER-модель по нотации IDEF1X

Следует отметить, что в IDEF1X вводится понятие «отношение категоризации». Отношение полной категоризации (сущности-категории составляют полное множество потомков родительской сущности) обозначается следующим образом (Рисунок 4).

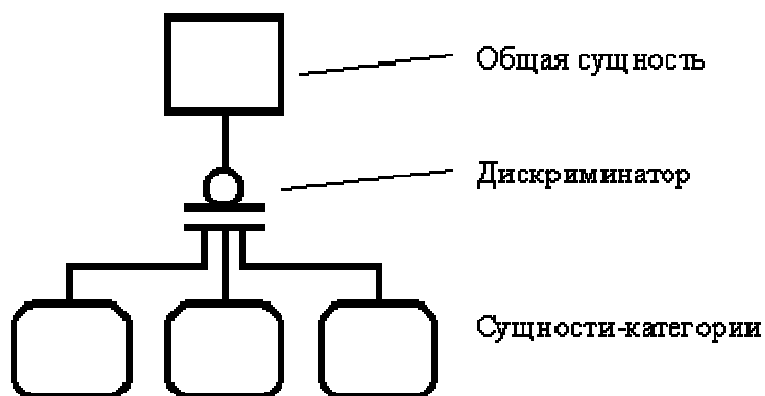


Рисунок 4 – Отношение полной категоризации по нотации IDEF1X

Также может существовать отношение неполной категоризации (сущности-категории составляют неполное множество потомков общей сущности) (Рисунок 5).

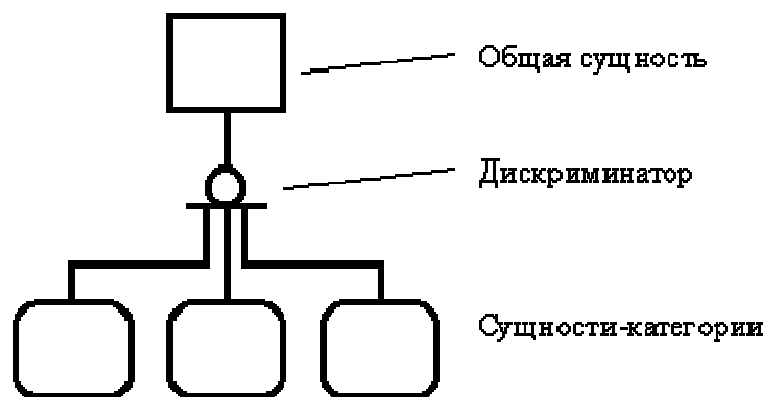


Рисунок 5 – Отношение неполной категоризации по нотации IDEF1X

Нотация Баркера

Дальнейшее развитие ER-подход получил в работах Баркера, предложившего оригинальную нотацию, которая позволила на верхнем уровне интегрировать предложенные Ченом средства описания моделей.

В нотации Баркера используется только один тип диаграмм – ER. Сущность представляется прямоугольником любого размера, содержащим внутри себя имя сущности, список имен атрибутов (возможно, неполный) и указатели ключевых атрибутов (знак "#" перед именем атрибута).

Все связи являются бинарными и представляются линиями с двумя концами (соединяющими сущности), для которых должно быть определено имя, степень множественности и степень обязательности. Для множественной связи линия присоединяется к прямоугольнику сущности в трех точках, а для одиночной связи - в одной точке. При обязательной связи рисуется непрерывная линия до середины связи, при необязательной - пунктирная линия.

Читается связь отдельно для каждого конца, показывая, как сущность 1 связывается с сущностью 2, и наоборот.

Связи обозначаются линиями с именами, место соединения связи и сущности определяет кардинальность связи (таблица 6).

Описание связей по нотации Баркера

Обозначение	Кардинальность
-----	0,1
—————	1,1
----->	0,N
—————>	1,N

ER-модель по нотации Баркера представлена на рисунке 6. Для обозначения отношения категоризации вводится элемент «дуга»

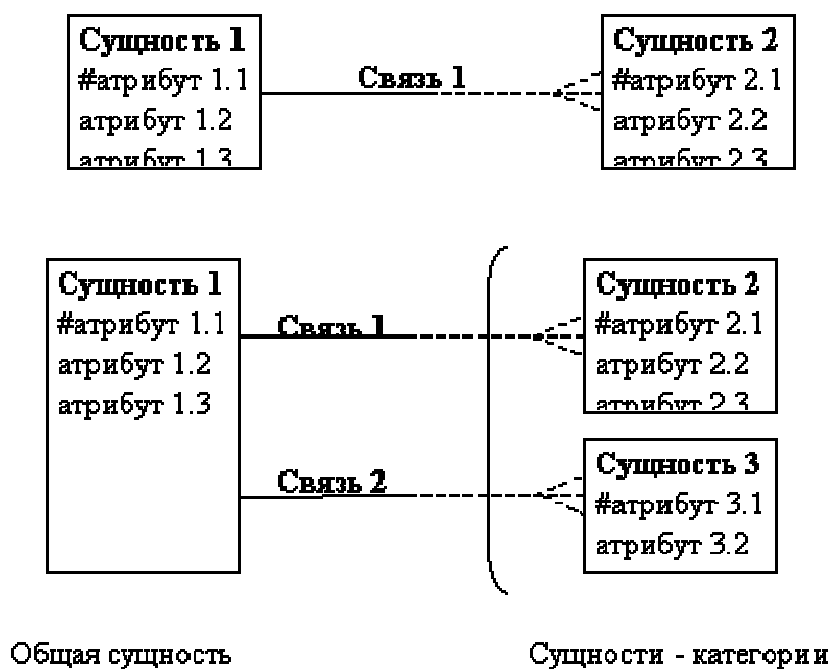


Рисунок 6 – ER-модель по нотации Баркера

Итак, можно выделить несколько категорий различий в изображении ER-моделей. Несущественные различия, связанные с использованием разных условных обозначений для отображения одних и тех же сущностей. Так, для

обозначения объекта могут использоваться прямоугольники, блоки с закругленными углами, овалы и т. д.

Следующая совокупность различий связана со способом изображения связей между объектами и заданием имен связей. Так, в некоторых методиках для изображения связи в разъеме линии, отображающей эту связь, предлагается изображать ромб и внутри него или рядом с ним писать название связи (модель Чена).

Разные условные обозначения используются и для изображения типа связи (1:1, 1 : M, M : M).

Таким образом, обзор литературных источников по проблеме исследования позволяет сделать следующие выводы:

- среди выделяемых подходов к анализу описаний и спецификаций предметных областей в деятельности архитектора баз данных следует отметить подход, основанный на использовании грамматического анализа для описания системы на естественном языке. Данный подход предполагает проведение текстового анализа на основе выделения существительных и глаголов (метод Дж. Аббота);

- для построения концептуальной модели, описывающей особенности предметной области, следует использовать модель «сущность-связь», разработанную Питером Ченом, в которой предлагается извлекать именные и глагольные группы для построения моделей данных. Согласно нотации Питера Чена, предметная область состоит из отдельных объектов, находящихся друг с другом в определенных связях, отношениях. Объекты, описываются различными параметрами или атрибутами; однотипные объекты описываются одним и тем же набором параметров и объединяются в множества или классы; такие классы называются сущностями. Конкретные объекты, составляющие класс, называются экземплярами соответствующей сущности. Между сущностями специфицируются взаимосвязи различного вида: один к одному, один ко многим и др. Отношение в самом общем виде представляет собой связь

между двумя и более сущностями. Именованье отношения осуществляется с помощью грамматического оборота глагола.

Библиографический список:

1. Буч Г. Объектно-ориентированный анализ и проектирование. // http://programming-lang.com/ru/comp_programming/buch/0/
2. Коннолли Т., Бегг К. Базы данных. Проектирование, реализация и сопровождение. Теория и практика = Database Systems: A Practical Approach to Design, Implementation, and Management. – 3-е изд. – М.: Вильямс, 2003. – 1436 с.
3. Кузнецов С. Д. Основы баз данных. – 2-е изд. – М.: Интернет-Университет Информационных Технологий, 2007. – 484 с.
4. Чен П. Модель «сущность – связь» - шаг к единому представлению данных // Системы управления базами данных. – 1995. - №3. – С.137–158.
5. Abbott R. Program Design by Informal English Descriptions. // Communications of the ACM. – Vol. 26(11). –November, 1983. – Pp. 882-894.
6. Brown A. Modelling a Real-World System and Designing a Schema to Represent It // Data Base Description. – North-Holland, 1975. – 382 p.