

07, июль 2017

УДК 519.816

Сравнение нотаций разработки инфологических схем в проектировании баз данных

*Лукьянова И.В., магистр
кафедра «Системы обработки информации и управления»,
Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана*

*Джанкулаева З.А., магистр
кафедра «Системы обработки информации и управления»,
Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана*

*Научный руководитель: Постников В.М., к.т.н., доцент
Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана,
кафедра «Системы обработки информации и управления»
chernen@bmstu.ru*

В процессе создания программного обеспечения, проектировщику приходится решать различные задачи. Не последней из них является проектирование базы данных. Первым и одним из важнейших этапов разработки качественной базы данных является этап инфологического (или концептуального) проектирования.

Целью данной работы является определение наиболее удобной нотации для разработки инфологической схемы базы данных (БД), что поможет сократить сроки проектирования и обеспечит лучшее взаимопонимание между пользователем и разработчиком.

Инфологическое (концептуальное) проектирование — построение семантической модели предметной области, то есть информационной модели наиболее высокого уровня абстракции. Основными задачами инфологического проектирования являются определение предметной области (ПО) системы и формирование взгляда на ПО с позиций сообщества будущих пользователей БД, т.е. инфологической модели ПО. Такая модель создаётся без ориентации на какую-либо конкретную СУБД и модель данных. [1, 2]

Исходя из изложенных выше требований, в описываемой методике проектирования используется модель, названная «сущность-связь» или ER-модель. Эта модель позволяет описывать структуру предметной области в виде объектов и связей между ними. [3,4]

При этом следует учитывать следующие основные понятия, используемые при разработке инфологической модели БД.[5-7]

- **Предметная область** - часть реального мира, подлежащая изучению с целью организации управления и, в конечном счете, автоматизации.
- **Сущность** - представление (абстракция) реально существующего объекта, процесса или явления. Наименование сущности должно быть уникально во всей модели.
- **Атрибут** - свойство сущности (объекта). Его имя должно быть уникально в рамках одной сущности.
- **Ключ** – минимальный набор атрибутов, по значениям которых можно однозначно найти требуемый экземпляр сущности.
- **Связь** позволяет моделировать отношения между объектами предметной области. Наименование связи должно быть уникально во всей модели.

Основные преимущества инфологических моделей, т.е. ER-моделей (eternity-relationship models):

- наглядность;
- возможность проектирования баз данных с большим количеством объектов и атрибутов;
- широкое применение во многих системах автоматизированного проектирования баз данных (например, ERWin, Oracle Designer и т.д.).

Существуют различные нотации разработки инфологических моделей, а также огромное количество CASE-средств для их реализации.

На начальных этапах проектирования разработчик сталкивается с проблемой выбора инструментария для создания программного обеспечения, а также нотации в которой будет написана проектная документация.

Для сравнения были выбраны следующие нотации, являющиеся наиболее широко используемыми [8,9]:

- В1 - Нотация Чена
- В2 - Нотация Баркера
- В3 - Нотация IDEFx1

В данных нотациях были созданы инфологические схемы по предметной области «Call-центр».

Нотация Чена

- Сущность обозначается как прямоугольник
- Связь соединяется с ассоциируемыми сущностями линиями.
- Возле каждой сущности на линии, соединяющей ее со связью, цифрами указывается класс принадлежности.

Элемент диаграммы	Обозначает		
	независимая сущность		Атрибут
	зависимая сущность		первичный ключ
	родительская сущность иерархической связи	в 	внешний ключ (понятие внешнего ключа вводится в реляционной модели данных)
	Связь		многозначный атрибут
	идентифицирующая связь		получаемый (наследуемый) атрибут иерархических связях

Рис. 1. Виды сущностей и связей в нотации Чена

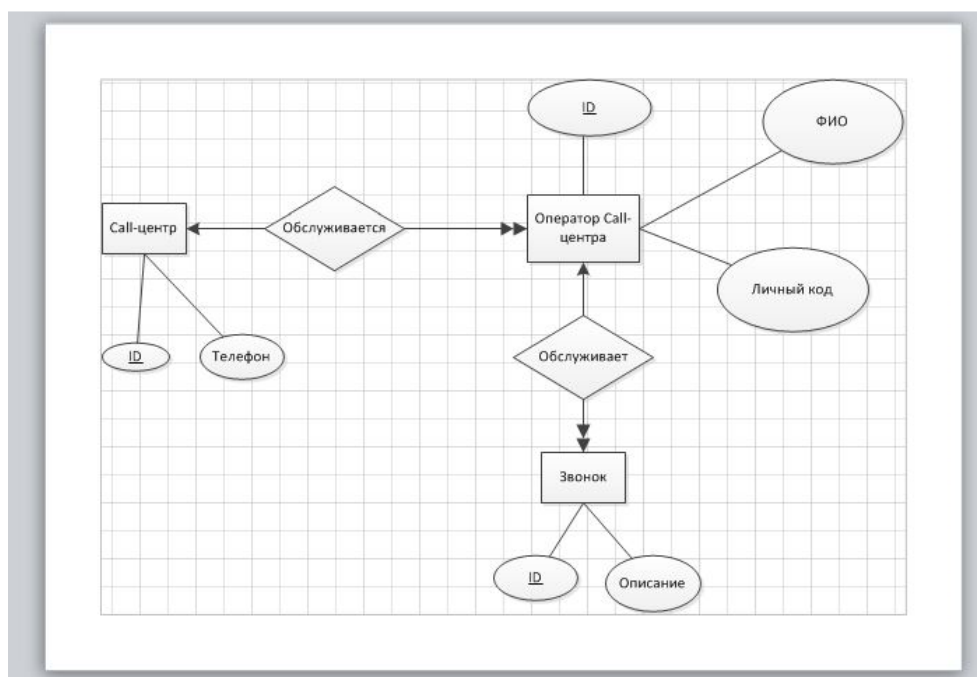


Рис. 2. Фрагмент инфологической модели ИС «Call-центр» в нотации Чена

Нотация Баркера

- Сущности обозначаются прямоугольниками, внутри которых приводится список атрибутов.
- Ключевые атрибуты отмечаются символом # (решетка).
- Связи обозначаются линиями с именами, место соединения связи и сущности определяет кардинальность связи:

Обозначение	Кардинальность
-----	0,1
_____	1,1
----->=	0,N
_____>=	1,N

Рис. 3. Виды связей в нотации Баркера

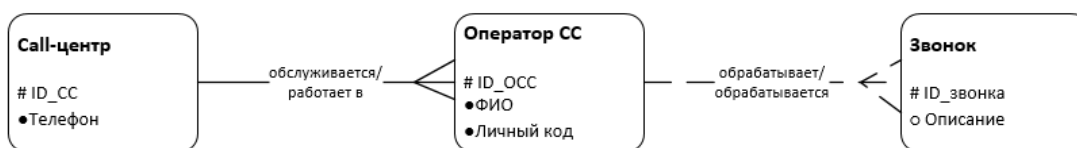


Рис. 4. Фрагмент инфологической модели ИС «Call-центр»

Нотация IDEFx1

- Список атрибутов приводится внутри прямоугольника, обозначающего сущность.
- Атрибуты, составляющие ключ сущности, группируются в верхней части прямоугольника и отделяются горизонтальной чертой.



Рис. 5. Виды сущностей и связей в нотации IDEFx1

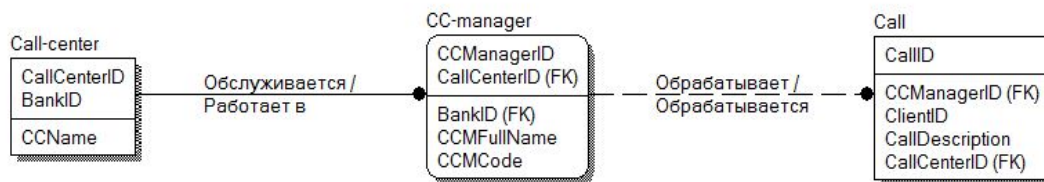


Рис. 6. Фрагмент инфологической модели ИС «Call-центр» в нотации IDEFx1

Приведенные нотации сравнивались как с точки зрения пользователя, так и с точки зрения разработчика. Для сравнительного анализа были выбраны наборы локальных критериев, приведенные в таблице 1.

Таблица 1

Критерии сравнения нотаций

Номер критерия	С точки зрения пользователя	С точки зрения разработчика
K1	Информативность модели	Легкость освоения и развития модели в данной нотации
K2	Компактность представления	Доступность Case-средств
K3	Наглядность представления	Возможность создания агрегированного объекта
K4	Избыточность	Простота перехода к даталогической модели и программированию
K5	Количество обозначений в нотации	Возможность расширения

Для выбора наиболее удобной нотации с учетом предложенного набора локальных критериев целесообразно использовать метод аналитической иерархии (МАИ). [10]

Вербально-числовая шкала для сравнения вариантов по локальным критериям приведена в табл. 2, а парное сравнение локальных критериев приведено в таблице 3.

Таблица 2

Перевод качественных показателей критериев в количественные

	К1 Информативность модели	К2 Компактность представления	К3 Наглядность представления	К4 Избыточность	К5 Количество обозначений в нотации
5	Модель информативна. Уточнений не требуется.	13 и более сущностей на лист	Незнакомому с предметной областью пользователю практически не требуется времени для понимания модели	Избыточность невозможна.	Менее 10
4	Модель информативна. Требуется небольшие уточнения.	От 10 до 12 сущностей на лист	Незнакомому с предметной областью пользователю требуется около получаса для понимания модели	Избыточность маловероятна.	От 10 до 15
3	В целом модель информативна, но требуются серьезные уточнения деталей.	8 – 9 сущностей на лист	Незнакомому с предметной областью пользователю требуется около часа для понимания модели	Модель может быть избыточна, но это не мешает в работе с ней	От 15 до 19
2	Модель мало информативна. Требуется большое количество уточнений как деталей, так и в целом.	От 4 до 7 сущностей на лист	Эксперту требуется около получаса, для понимания модели	Модель избыточна. Дублирование небольшой части информации.	От 20 до 24

Таблица 2 (Продолжение)

1	Модель абсолютно неинформативна.	Менее 4 сущностей на лист	Эксперту требуется более часа, чтобы понять модель	Модель избыточна. Для достижения нужной степени информативности, необходимо дублирование большей части информации	Более 25
---	----------------------------------	---------------------------	--	---	----------

Таблица 3

Матрица сравнения локальных критериев выбора нотаций БД для пользователя

П	K1	K2	K3	K4	K5	C_i	α_i
K1	1	3	1	0,33	5	1,377	0,242
K2	0,33	1	0,33	1	1	0,643	0,113
K3	1	3	1	3	5	2,145	0,377
K4	3	1	0,33	1	1	1,002	0,176
K5	0,2	1	0,2	1	1	0,525	0,092
K3>K1>K4>K4>K5							

Определяем собственные значения векторов локальных критериев C_i . Затем вычисляем коэффициент важности α_i каждого критерия:

$$C_i = (k_{i1}k_{i2} \dots k_{in})^{1/n}$$

$$\alpha_i = \frac{C_i}{\sum_{i=1}^n C_i}$$

Далее проводим попарное сравнение вариантов нотаций по каждому критерию и рассчитываем собственные значения векторов C_{ij} и коэффициенты важности β_i по следующим формулам, согласно [10].

$$C_{ij} = (k_{ij1}k_{ij2} \dots k_{ijm})^{1/m}$$

$$\beta_i = \frac{C_{ji}}{\sum_{i=1}^m C_{ij}}$$

Матрицы парного сравнения вариантов нотаций по критериям с позиции пользователя приведены в таблицах 4 - 8.

Таблица 4

Матрица сравнения вариантов по критерию К1

	В1	В2	В3	Собств. вектор	Вес варианта по критерию К1
В1	1	0,8	2	1,17	0,357
В2	1,25	1	3	1,554	0,475
В3	0,5	0,333	1	0,55	0,168
			<i>Сумма:</i>	3,274	1

Таблица 5

Матрица сравнения вариантов по критерию К2

	В1	В2	В3	Собств. вектор	Вес варианта по критерию К2
В1	1	0,5	0,3	0,531	0,161
В2	2	1	2	1,587	0,48
В3	3,333	0,5	1	1,186	0,359
			<i>Сумма:</i>	3,304	1

Таблица 6

Матрица сравнения вариантов по критерию К3

	В1	В2	В3	Собств. вектор	Вес варианта по критерию К3
В1	1	0,8	0,8	0,862	0,286
В2	1,25	1	1	1,077	0,357
В3	1,25	1	1	1,077	0,357
			<i>Сумма:</i>	3,016	1

Таблица 7

Матрица сравнения вариантов по критерию К4

	В1	В2	В3	Собств. вектор	Вес варианта по критерию К4
В1	1	0,6	0,6	0,711	0,231
В2	1,666	1	1	1,186	0,385
В3	1,666	1	1	1,186	0,385
			<i>Сумма:</i>	3,083	1

Матрица сравнения вариантов по критерию К5

	B1	B2	B3	Собств. вектор	Вес варианта по критерию К5
B1	1	0,8	0,8	0,862	0,286
B2	1,25	1	1	1,077	0,357
B3	1,25	1	1	1,077	0,357
			<i>Сумма:</i>	<i>3,016</i>	<i>1</i>

Согласно [10] определяем наилучшую нотацию создания инфологических схем для пользователей, используя выражения (1) и (2):

$$Y_j = \sum_{i=1}^n \alpha_i \beta_{ij}, j=1, \dots, m \quad (1)$$

$$Y_l = \max_j Y_j \quad (2)$$

Подставляя численные значения из табл.4-8 в выражение (1) получаем:

$$Y_1 = 0.242 * 0.357 + 0.113 * 0.161 + 0.377 * 0.286 + 0.176 * 0.231 + 0.092 * 0.286 = 0.279$$

$$Y_2 = 0.242 * 0.475 + 0.113 * 0.48 + 0.377 * 0.357 + 0.176 * 0.385 + 0.092 * 0.357 = 0.404$$

$$Y_3 = 0.242 * 0.168 + 0.113 * 0.359 + 0.377 * 0.357 + 0.176 * 0.385 + 0.092 * 0.357 = 0.316$$

$$Y_1 = Y_2$$

Получаем, что вариант 2, соответствующий нотации Баркера, является наилучшим для пользователей.

Далее был проведен опрос среди разработчиков БД. Сравнительный анализ нотаций разработки инфологической схемы БД с точки зрения разработчика с использованием метода аналитической иерархии приведен в таблицах 9 - 17.

Таблица 9

Перевод качественных показателей критериев в количественные.

К1	К2	К3	К4	К5
Легкость освоения и развития модели в данной нотации	Доступность Case-средств	Возможность создания агрегированного объекта	Простота перехода к дата-логической модели и программированию	Возможность расширения

Таблица 9 (Продолжение)

5	Модель в данной нотации легка в освоении и развитии. Разработчику практически не требуется времени на изучение	Широкий выбор case-средств, находящихся в свободном доступе	Создание агрегированного объекта изначально заложено в нотации	Переход с помощью case-средств	Возможности расширения заложены в нотации и case-средствах
4	Освоение и развитие модели достаточно легко, но требует времени.	Несколько case-средств за небольшую плату	Создание агрегированного объекта не предусмотрено в нотации, но возможно, с использованием усложненных конструкций	Переход вручную, требуется небольшая доработка модели	Расширение нотации возможно, но не все изменения предусмотрены в существующих case - средствах
3	Модель легка в освоении, но ее развитие вызывает затруднения.	Существуют case-средства, но их мало, и они дороги	Создание агрегированного объекта возможно, но требует корректировки модели	Переход вручную, требуется серьезная	Расширение нотации возможно, но ограничено
2	Освоение и развитие модели затруднительно, но возможно.	Единственное case-средство, количество лицензий ограничено 1-2 компаниями	Создание агрегированного объекта затруднительно	Переход затруднителен	Расширение практически невозможно, но существ
1	Освоить и развить модель в данной нотации практически невозможно	Case – средств не существует	Создание агрегированного объекта невозможно	Переход невозможен	Расширение невозможно

Проведем парное сравнение локальных критериев.

Матрицы сравнения локальных критериев для разработчика

P	K1	K2	K3	K4	K5	C_i	α_i
K1	1	0,33	3	0,33	0,33	0,641	0,108
K2	3	1	3	0,33	3	1,552	0,261
K3	0,33	0,33	1	0,33	1	0,516	0,087
K4	3	3	3	1	3	2,423	0,407
K5	3	0,33	1	0,33	1	0,804	0,135
K4 >> K2 > K5 > K1 > K3							

Определяем собственные значения векторов локальных критериев C_i . Затем вычисляем коэффициент важности α_i каждого критерия, которые приведены в последнем столбце в табл. 10:

$$C_i = (k_{i1} k_{i2} \dots k_{in})^{1/n} .$$

$$\alpha_i = \frac{C_i}{\sum_{i=1}^n C_i} .$$

Сравним варианты попарно по каждому критерию и рассчитаем собственные значения векторов C_{ij} и коэффициенты важности β_i по следующим формулам:

$$C_{ij} = (k_{ij1} k_{ij2} \dots k_{ijm})^{\frac{1}{m}}$$

$$\beta_i = \frac{C_{ji}}{\sum_{i=1}^m C_{ij}} .$$

Матрицы парного сравнения вариантов нотаций по критериям с позиции разработчика приведены в таблицах 11 - 15.

Таблица 11

Матрица сравнения вариантов по критерию К1

	B1	B2	B3	Собств. вектор	Вес варианта по критерию К1
B1	1	0,8	0,6	0,783	0,256
B2	1,25	1	0,8	1	0,327
B3	1,666	1,25	1	1,277	0,417
			<i>Сумма:</i>	<i>3,06</i>	<i>1</i>

Таблица 12

Матрица сравнения вариантов по критерию К2

	B1	B2	B3	Собств. вектор	Вес варианта по критерию К2
B1	1	0,8	0,5	0,737	0,238
B2	1,25	1	0,8	1	0,323
B3	2	1,25	1	1,357	0,439
			<i>Сумма:</i>	<i>3,094</i>	<i>1</i>

Таблица 13

Матрица сравнения вариантов по критерию К3

	B1	B2	B3	Собств. вектор	Вес варианта по критерию К3
B1	1	0,6	0,8	0,783	0,25
B2	1,666	1	2	1,494	0,477
B3	1,25	0,5	1	0,855	0,273
			<i>Сумма:</i>	<i>3,132</i>	<i>1</i>

Таблица 14

Матрица сравнения вариантов по критерию К4

	B1	B2	B3	Собств. вектор	Вес варианта по критерию К4
B1	1	0,8	0,8	0,862	0,286
B2	1,25	1	1	1,077	0,357
B3	1,25	1	1	1,077	0,357
			<i>Сумма:</i>	<i>3,016</i>	<i>1</i>

Матрица сравнения вариантов по критерию K5

	B1	B2	B3	Собств. вектор	Вес варианта по критерию K5
B1	1	2	3	1,817	0,54
B2	0,5	1	2	1	0,297
B3	0,333	0,5	1	0,55	0,163
			<i>Сумма:</i>	<i>3,367</i>	<i>1</i>

Согласно [10] определяем наилучшую нотацию создания инфологических схем для пользователей, используя выражения (3) и (4):

$$:Y_j = \sum_{i=1}^n \alpha_i \beta_{ij}, j=1, \dots, m \quad (3)$$

$$Y_l = \max_j Y_j. \quad (4)$$

После подстановки данных из табл. 10-15 получаем:

$$Y_1 = 0.108 * 0.256 + 0.261 * 0.238 + 0.087 * 0.25 + 0.407 * 0.286 + 0.235 * 0.54 = 0.289$$

$$Y_2 = 0.108 * 0.327 + 0.261 * 0.323 + 0.087 * 0.477 + 0.407 * 0.357 + 0.235 * 0.297 = 0.347$$

$$Y_3 = 0.108 * 0.417 + 0.261 * 0.439 + 0.087 * 0.273 + 0.407 * 0.357 + 0.235 * 0.163 = 0.351$$

$$Y_l = Y_3.$$

Получаем, что вариант 3, соответствующий нотации IDEFx1, является наилучшим для разработчиков.

Выводы

Наиболее удобной в практической работе, т.е. для разработчиков баз данных, является представление инфологических схем в нотации IDEFx1, а для пользователей – в нотации Баркера.

Однако следует иметь в виду, что каждая из рассмотренных нотаций обладает рядом преимуществ и ориентирована на определенные цели и предпочтения разработчиков.

Также следует учитывать, что нотация Чена до сих пор широко используется на первоначальных этапах инфологического проектирования баз данных благодаря простоте и доступности CASE-средств.

Список литературы

- [1]. Проектирование баз данных. Режим доступа: https://ru.wikipedia.org/wiki/Проектирование_баз_данных (дата обращения 10.03.2016).
- [2]. Теоретические основы построения баз данных. Режим доступа: <http://lib.ssga.ru/fulltext/УМК/ЭУМК%20ФГОС-%203%2013-14/ЭУМК%20080501.62%20МЕНЕДЖМЕНТ/Семестр%202/Информационные%20технологии%20в%20менеджменте/ТЕОРЕТИЧЕСКИЕ%20ОСНОВЫ%20БАЗ%20ДАНЫХ.pdf> (дата обращения: 16.03.2016)
- [3]. Петер Пин-Шен Чен. Модель «сущность-связь» – шаг к единому представлению о данных // Системы Управления Базами Данных #3/1995. 2009. С. 3-12
- [4]. Построение диаграммы «сущность-связь» в различных нотациях. Режим доступа: <https://sites.google.com/site/gosyvmkss12/bazy-dannyh/15-postroenie-diagrammy-susnost-svaz-v-razlicnyh-notaciah> (дата обращения 16.03.2016).
- [5]. Диго С.М. Базы данных. Проектирование и создание: учебно-методический комплекс. М.: Изд. центр ЕАОИ, 2008. 171 с.
- [6]. Хомопепко А.Д., Цыганков В.М., Мальцев М.Г. Базы данных: учебник для высших учебных заведений / под ред. А.Д. Хомопепко. 6-е изд., доп. СПб.: КОРОНА-Век, 2009. 736 с.
- [7]. Харрингтон Д. Проектирование объектно-ориентированных баз данных: пер. с англ. М.: ДМК Пресс, 2012. 272 с.: ил. (Серия «Для программистов»).
- [8]. Рыбанов А.А. Инструментальные средства автоматизированного проектирования баз данных: учебное пособие и варианты заданий к лабораторным работам по дисциплине «Базы данных». Волгоград: ВолгГТУ, 2007. 96 с.
- [9]. Швецов В.И. Базы данных. М.: НОУ «Интуит», 2016. 218 с.
- [10]. Постников В.М., Черненький В.М. Методы принятия решений в системах организационного управления. М.: Изд-во МГТУ им. Н.Э. Баумана, 2014. 205 с.