

## Проектирование баз данных

Проектирование базы данных (БД) – одна из наиболее сложных и ответственных задач, связанных с созданием информационной системы (ИС). В результате её решения должны быть определены содержание БД, эффективный для всех её будущих пользователей способ организации данных и инструментальные средства управления данными.

Процесс проектирования БД представляет собой последовательность переходов от неформального словесного описания информационной структуры предметной области к формализованному описанию объектов предметной области в терминах некоторой модели. В общем случае можно выделить следующие этапы проектирования:

1. Системный анализ и словесное описание информационных объектов предметной области.
2. Концептуальное (инфологическое) проектирование — построение семантической модели предметной области, то есть информационной модели наиболее высокого уровня абстракции. Такая модель создаётся без ориентации на какую-либо конкретную СУБД и модель данных. Термины «семантическая модель», «концептуальная модель» и «инфологическая модель» являются синонимами. Кроме того, в этом контексте равноправно могут использоваться слова «модель базы данных» и «модель предметной области» (например, «концептуальная модель базы данных» и «концептуальная модель предметной области»), поскольку такая модель является как образом реальности, так и образом проектируемой базы данных для этой реальности. Конкретный вид и содержание концептуальной модели базы данных определяется выбранным для этого формальным аппаратом. Обычно используются графические нотации, подобные ER-диаграммам.

Чаще всего концептуальная модель базы данных включает в себя:

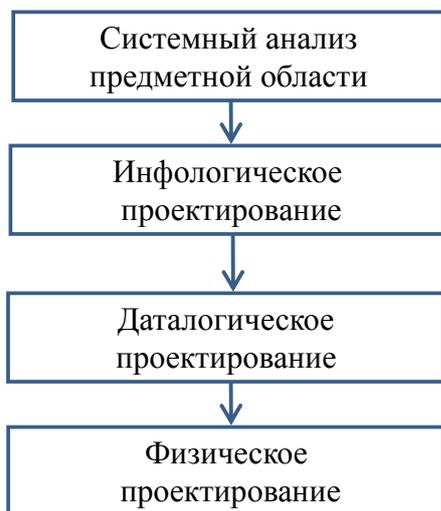
- описание информационных объектов, или понятий предметной области и связей между ними.

- описание ограничений целостности, т.е. требований к допустимым значениям данных и к связям между ними.

3. Даталогическое (логическое) проектирование — создание схемы базы данных на основе конкретной модели данных, например, реляционной модели данных. Для реляционной модели данных даталогическая модель — набор схем отношений, обычно с указанием первичных ключей, а также «связей» между отношениями, представляющих собой внешние ключи.

4. Физическое проектирование — создание схемы базы данных для конкретной СУБД. Специфика конкретной СУБД может включать в себя ограничения на именование объектов базы данных, ограничения на поддерживаемые типы данных и т.п. Кроме того, специфика конкретной СУБД при физическом проектировании включает выбор решений, связанных с физической средой хранения данных (выбор методов управления дисковой памятью, разделение БД по файлам и устройствам, методов доступа к данным), создание индексов и т.д.

Условно процесс проектирования БД можно представить последовательностью выполнения пяти соответствующих этапов.



### **Системный анализ предметной области**

Первый этап проектирования БД это системный анализ , то есть подробное словесное описание объектов предметной области и реальных связей, которые присутствуют между описываемыми объектами.

## 1. Функциональный подход к проектированию БД.

Этот метод является наиболее распространённым. Он реализует принцип "от задач" и применяется в том случае, когда известны функции некоторой группы лиц и комплекса задач, для обслуживания информационных потребностей которых создаётся рассматриваемая БД.

## 2. Предметный подход к проектированию БД.

Предметный подход применяется в тех случаях, когда у разработчиков есть чёткое представление о самой ПО и о том, какую именно информацию они хотели бы хранить в БД, а структура запросов не определена или определена не полностью. Тогда основное внимание уделяется исследованию ПО и наиболее адекватному её отображению в БД с учётом самого широкого спектра информационных запросов к ней.

Чаще всего на практике рекомендуется использовать некоторый компромиссный вариант, который, с одной стороны, ориентирован на конкретные задачи или функциональные потребности пользователей, а с другой стороны, учитывает возможность наращивания новых приложений.

Системный анализ должен заканчиваться подробным описанием информации об объектах предметной области, которая требуется для решения конкретных задач и которая должна храниться в БД, формулировкой конкретных задач, которые будут решаться с использованием данной БД с кратким описанием алгоритмов их решения, описанием выходных документов, которые должны генерироваться в системе, описанием входных документов, которые служат основанием для заполнения данными БД.

### **Пример описания предметной области**

Пусть требуется разработать информационную систему для автоматизации учета получения и выдачи книг в библиотеке. Система должна предусматривать режимы ведения системного каталога, отражающего перечень областей знаний, по которым имеются книги в библиотеке. Внутри

библиотеки области знаний в систематическом каталоге могут иметь уникальный внутренний номер и полное наименование.

Каждая книга может содержать сведения из нескольких областей знаний.

Каждая книга в библиотеке может присутствовать в нескольких экземплярах.

Каждая книга, хранящаяся в библиотеке, характеризуется следующими параметрами:

- уникальный шифр;
- название;
- фамилии авторов;
- место издания (город);
- издательство;
- год издания;
- количество страниц;
- количество экземпляров;
- номер области знаний;
- название области знаний.

Книги могут иметь одинаковые названия, но они различаются по своему уникальному шифру (ISBN).

В библиотеке ведется картотека читателей.

На каждого читателя в картотеку заносятся следующие сведения:

- фамилия, имя, отчество;
- пол
- домашний адрес;
- телефон;
- дата рождения.

Каждому читателю присваивается уникальный номер читательского билета.

Каждый читатель может одновременно держать на руках не более 5 книг.

Читатель не должен одновременно держать более одного экземпляра книги одного названия.

Каждая книга в библиотеке может присутствовать в нескольких экземплярах.

Каждый экземпляр имеет следующие характеристики:

- уникальный инвентарный номер;
- шифр книги, который совпадает с уникальным шифром из описания книг;
- место размещения в библиотеке.

В случае выдачи экземпляра книги читателю в библиотеке хранится специальный вкладыш, в котором должны быть записаны следующие сведения:

- номер билета читателя, который взял книгу;
- дата выдачи книги;
- дата возврата.

При работе с системой библиотекарь должен иметь возможность решать следующие задачи:

1. Принимать новые книги и регистрировать их в библиотеке.
2. Относить книги к одной или к нескольким областям знаний.
3. Проводить каталогизацию книг, то есть назначение новых инвентарных номеров вновь принятым книгам, и, помещая их на полки библиотеки, запоминать место размещения каждого экземпляра.
4. Проводить дополнительную каталогизацию, если поступило несколько экземпляров книги, которая уже есть в библиотеке, при этом информация о книге в предметный каталог не вносится, а каждому новому экземпляру присваивается новый инвентарный номер и для него определяется место на полке библиотеки.
5. Проводить списание старых и не пользующихся спросом книг. Списывать можно только книги, ни один экземпляр которых не находится у читателей. Списание проводится по специальному акту списания, который утверждается администрацией библиотеки.
6. Вести учет выданных книг читателям, при этом предполагается два режима работы: выдача книг читателю и прием от него

возвращаемых им книг обратно в библиотеку. При выдаче книг фиксируется, когда и какой экземпляр книги был выдан данному читателю и к какому сроку читатель должен вернуть этот экземпляр книги. При выдаче книг наличие свободного экземпляра и его конкретный номер могут определяться по заданному уникальному шифру книги или инвентарный номер может быть известен заранее. Не требуется вести «историю» чтения книг, то есть требуется отражать только текущее состояние библиотеки.

7. Проводить списание утерянных читателем книг.

8. Проводить закрытие абонемента читателя, то есть уничтожение данных о нем, если читатель хочет выписаться из библиотеки и не является ее должником, то есть за ним не числится ни одной библиотечной книги.

Читатель должен иметь возможность решать следующие задачи:

1. Просматривать системный каталог, то есть перечень всех областей знаний, книги по которым есть в библиотеке.

2. По выбранной области знаний получить полный перечень книг, которые числятся в библиотеке.

3. Для выбранной книги получить инвентарный номер свободного экземпляра книги или сообщение о том, что свободных экземпляров книги нет. В случае отсутствия свободных экземпляров книги читатель должен иметь возможность узнать дату ближайшего предполагаемого возврата экземпляра данной книги. Читатель не может узнать данные о том, у кого в настоящий момент экземпляры данной книги находятся на руках (в целях обеспечения личной безопасности держателей требуемой книги).

4. Для выбранного автора получить список книг, которые числятся в библиотеке.

Этот пример показывает, что перед началом разработки необходимо иметь точное представление о том, что же должно выполняться в системе, какие пользователи в ней будут работать, какие задачи будет решать каждый пользователь. Отсутствие четких целей создания БД может свести на нет все усилия разработчиков, и проект БД получится «плохим», неудобным, не соответствующим ни реально моделируемому объекту, ни задачам, которые должны решаться с использованием данной БД.

## Инфологическое проектирование

Инфологическая модель применяется на втором этапе проектирования БД, то есть после словесного описания предметной области и постановки задачи. Инфологическая модель должна включать такое формализованное описание предметной области, которое будет «читабельно» не только для специалистов по базам данных, но и сторонних людей. Описание должно быть настолько емким, чтобы можно было оценить глубину и корректность проработки проекта БД, и не должно быть привязано к конкретной СУБД. ER-диаграмма (entity-relationship diagram, диаграмма «сущность-связь») – это графическое представление инфологической модели.

Основные элементы ER-моделей:

- сущности (объекты);
- атрибуты сущностей;
- ключ сущности
- связи между сущностями.

**Сущность** - это класс однотипных объектов, информация о которых должна быть учтена в модели. Каждая сущность должна иметь наименование, выраженное существительным в единственном числе.

Каждая сущность в модели изображается в виде прямоугольника с наименованием.



**Атрибут сущности** - это именованная характеристика, являющаяся некоторым свойством сущности.

Наименование атрибута должно быть выражено существительным в единственном числе (возможно, с характеризующими прилагательными).

Примерами атрибутов сущности "Книга" могут быть такие атрибуты как "ISBN", "Название", "Автор", "Издательство", "Место издания", "Год издания", "Количество страниц", "Количество экземпляров" . Атрибуты изображаются в пределах прямоугольника, определяющего сущность:

Книга
ISBN Название Автор Издательство Место издания Год издания Количество страниц Количество экземпляров Номер области знаний Название области знаний

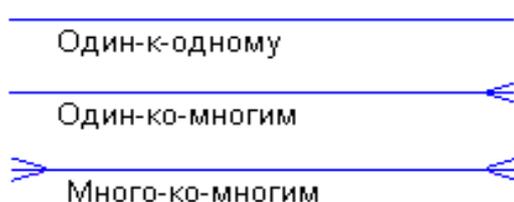
**Ключ сущности** - это избыточный набор атрибутов, значения которых в совокупности являются уникальными для каждого экземпляра сущности. Избыточность заключается в том, что удаление любого атрибута из ключа нарушается его уникальность. Сущность может иметь несколько различных ключей. Ключевые атрибуты изображаются на диаграмме подчеркиванием:

Книга
<u>ISBN</u> Название Автор Издательство Место издания Год издания Количество страниц Количество экземпляров Номер области знаний Название области знаний

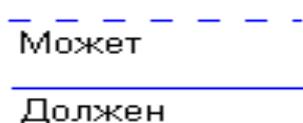
**Связь** - это некоторая ассоциация между двумя сущностями. Одна сущность может быть связана с другой сущностью или сама с собою.

Связи позволяют по одной сущности находить другие сущности, связанные с нею. Графически связь изображается линией, соединяющей две сущности. (рис ) Каждая связь имеет два конца и одно или два наименования. Наименование обычно выражается в неопределенной глагольной форме: "иметь", "принадлежать" и т.п. Каждое из наименований относится к своему концу связи. Иногда наименования не пишутся ввиду их очевидности.

Каждая связь может иметь один из следующих типов связи :



Каждая связь может иметь одну из двух модальностей связи:



Модальность "**может**" означает, что экземпляр одной сущности может быть связан с одним или несколькими экземплярами другой сущности, а может быть и не связан ни с одним экземпляром.

Модальность "**должен**" означает, что экземпляр одной сущности обязан быть связан не менее чем с одним экземпляром другой сущности

Например между сущностями «Читатель» и «Экземпляр» установлена связь «один-ко-многим», и при этом она не обязательная с двух сторон. Читатель в данный момент может не держать ни одной книги на руках, а с другой стороны, данный экземпляр книги может не находиться ни у одного читателя, а просто стоять на полке в библиотеке.

Так как мы выделили три основные сущности «Книга», «Экземпляр» и «Читатель» получилась инфологическая модель представленная на Рис.1.1.

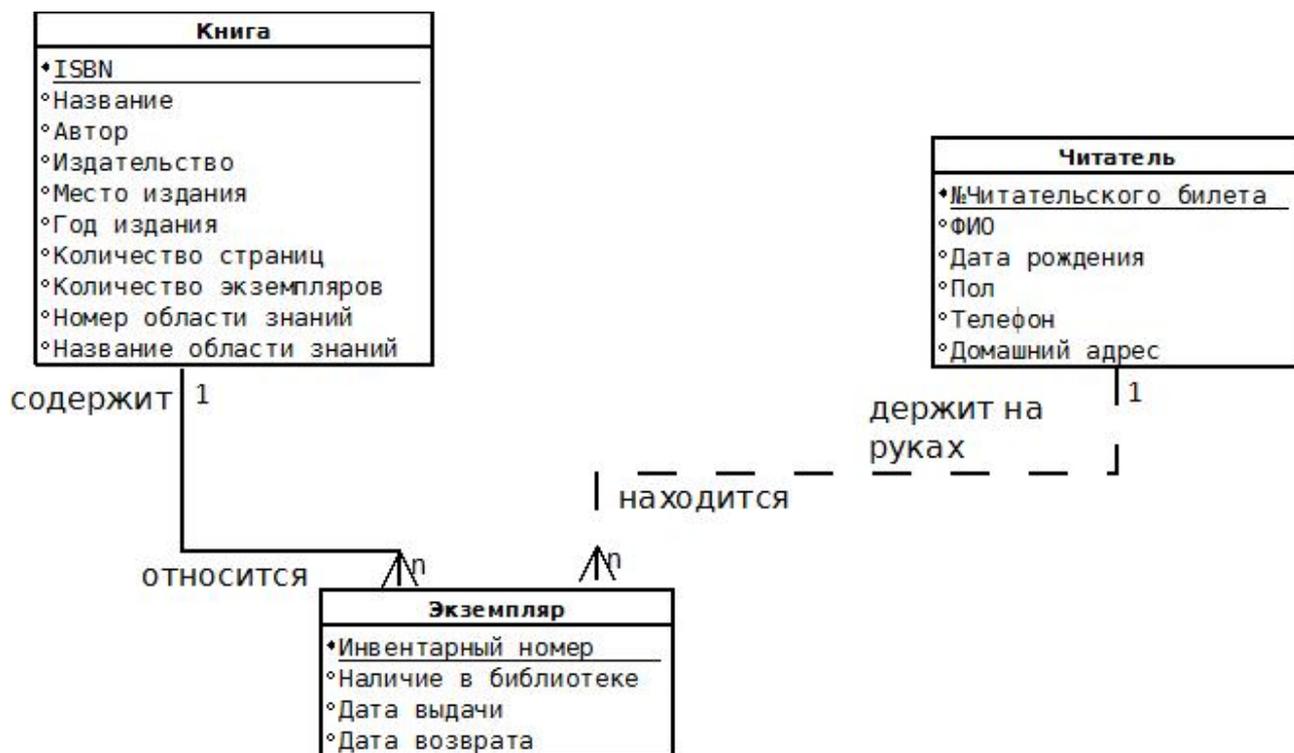


Рис.1.1 Инфологическая модель предметной области «Библиотека»

### *Даталогическое проектирование*

Следующим шагом является выбор конкретной СУБД и отображение в ее среду спецификаций инфологической модели предметной области. Эту стадию называют даталогическим (логическим) проектированием БД.

Существуют реляционные, иерархические и сетевые модели данных. Наибольшее распространение при разработке БД получили реляционные.

Реляционная модель данных является удобной и наиболее привычной формой представления данных в виде таблицы. В отличие от других моделей, такой способ представления:

- 1) понятен пользователю-непрограммисту;
- 2) позволяет легко изменить схему – присоединять новые элементы данных и записи без изменения соответствующих подсхем;
- 3) обеспечивает необходимую гибкость при обработке непредвиденных запросов.

Получение реляционной схемы из ER-диаграммы.

1. Каждая простая сущность превращается в таблицу (отношение). Имя сущности становится именем таблицы.

2. Каждый атрибут становится возможным столбцом с тем же именем.

3. Компоненты уникального идентификатора сущности превращаются в первичный ключ.

4. Связи «многие к одному» и «один к одному» становятся внешними ключами. Т.е. создается копия первичного ключа(уникального атрибута) с конца связи «один», и соответствующие столбцы составляют внешний ключ.

В результате получили следующие отношения:

Книга (ISBN, Название, Автор, Издательство, Место издания, Год издания, Количество страниц, Количество экземпляров, Номер области знаний, Название области знаний )

Экземпляр (Инвентарный номер книги, ISBN , Наличие в библиотеке, № Читательского билета, Дата выдачи, Дата возврата)

Читатель (№ Читательского билета, Фамилия И.О., Дата рождения, Домашний адрес, Пол, Телефон).

Основная цель логического проектирования базы данных — устранение избыточности и дублирования информации. Для удовлетворения этих требований Э.Коддом предложен аппарат нормализации отношений. База данных считается нормализованной, если ее таблицы представлены как минимум в третьей нормальной форме.

### **Первая нормальная форма**

1НФ требует атомарности данных в таблицах, т.е. данные в таблицах должны быть представлены в виде минимально возможных и далее неделимых кусочков информации.

### **Вторая нормальная форма**

2НФ требует, чтобы данные во всех неключевых столбцах полностью зависели от первичного ключа.

### **Третья нормальная форма**

Третья нормальная форма, известная также как 3НФ, требует от нас, чтобы структура базы данных Удовлетворяла требованиям 1НФ и 2НФ. Все неключевые столбцы таблицы зависели от первичного ключа, но были независимы друг от друга.

Отношения «Читатель» и «Экземпляр» находятся в 1-ой нормальной форме, т.к. не имеют сложных атрибутов. Атрибуты таблицы «Читатель» и таблицы «Экземпляр» зависят от первичных ключей (2НФ) и независят друг от друга что соответствует 3НФ.

Рассмотрим отношение «Книги». Каждая книга может относиться к многим областям знаний , т.е. атрибуты номер области знаний и название области знаний – сложные, а это значит, что нарушена 1НФ. Чтобы привести к 1-ой нормальной форме добавим к ключу еще один атрибут – номер области знаний.

Книга (ISBN, Название, Автор, Издательство, Место издания, Год издания, Количество страниц, Количество экземпляров, Номер области знаний, Название области знаний )

В отношении «Книга» 2-ая нормальная форма нарушена, т.к. есть неключевые атрибуты, зависящие только от части ключа, а не от всего составного ключа. Приведем это отношение ко 2-ой форме, разделив отношение на три отношения по зависимости от ключа или части ключа.

Назовем новые отношения: «Книга», «Область знаний», «Принадлежность книги к области знаний». В результате мы получили 5 отношений: «Книга», «Экземпляр», «Область знаний», «Принадлежность книги к области знаний», «Читатель» .

Даталогическая модель нормализованных отношений представлена на Рис.1.2.

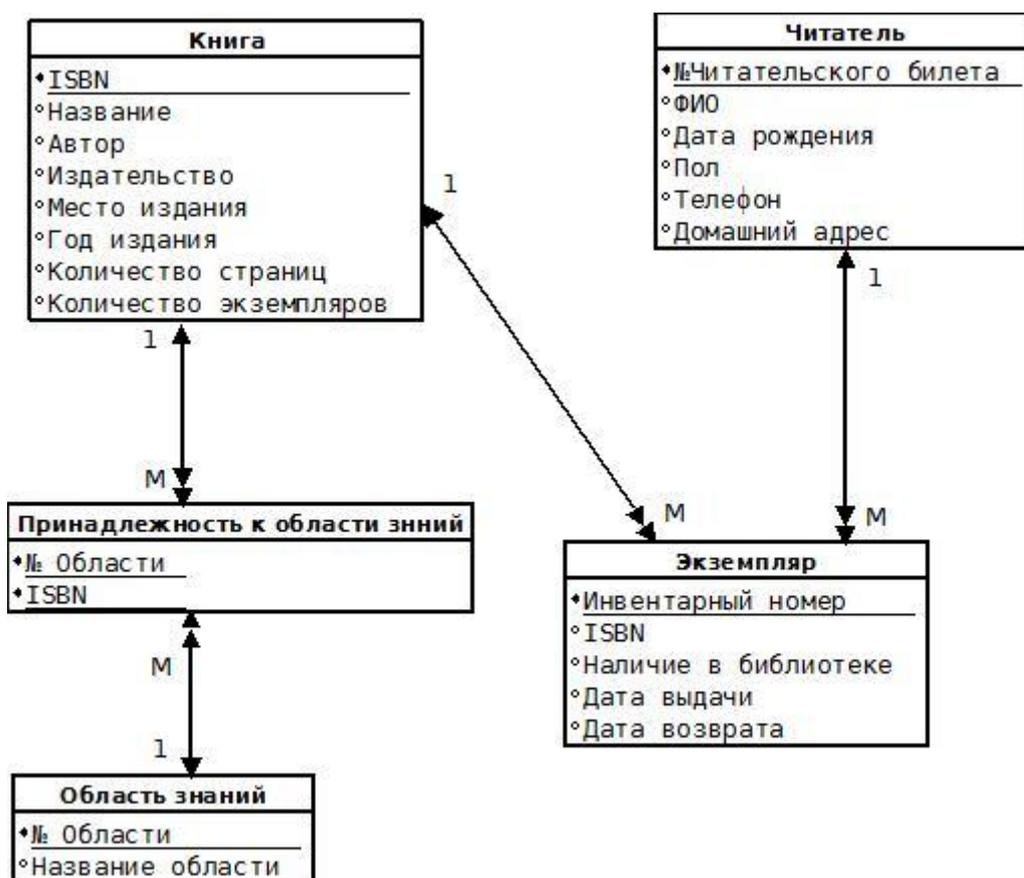


Рис.1.2 Дatalogическая модель предметной области «Библиотека»

## Физическое проектирование

Основой для физического проектирования является схема БД, полученная на предыдущем этапе. Физическое проектирование заключается в увязке логической структуры БД и физической среды хранения с целью наиболее эффективного размещения данных.

Определим структуры таблиц в среде СУБД OpenOffice.org Base . Дадим названия таблицам и атрибутам, определим типы данных и размерность атрибутов. В таблицах выберем первичные ключи и создадим внешний. (описания структуры таблицы Книги представлен в таблице 1).

Название таблицы	Имя поля	Тип данных	Размер поля	Ключ
Книга	ISBN	Текстовый	15	Первичный
	Название	Текстовый	50	
	Автор	Текстовый	50	

	Издательство	Текстовый	20	
	Место издания	Текстовый	15	
	Год издания	Текстовый	20	
	Количество страниц	Числовой		
	Количество экземпляров	Числовой		

Аналогичным образом определяется структура остальных таблиц.

Связи между таблицами в базе данных «Библиотека» представлены на

Рис.1.3.

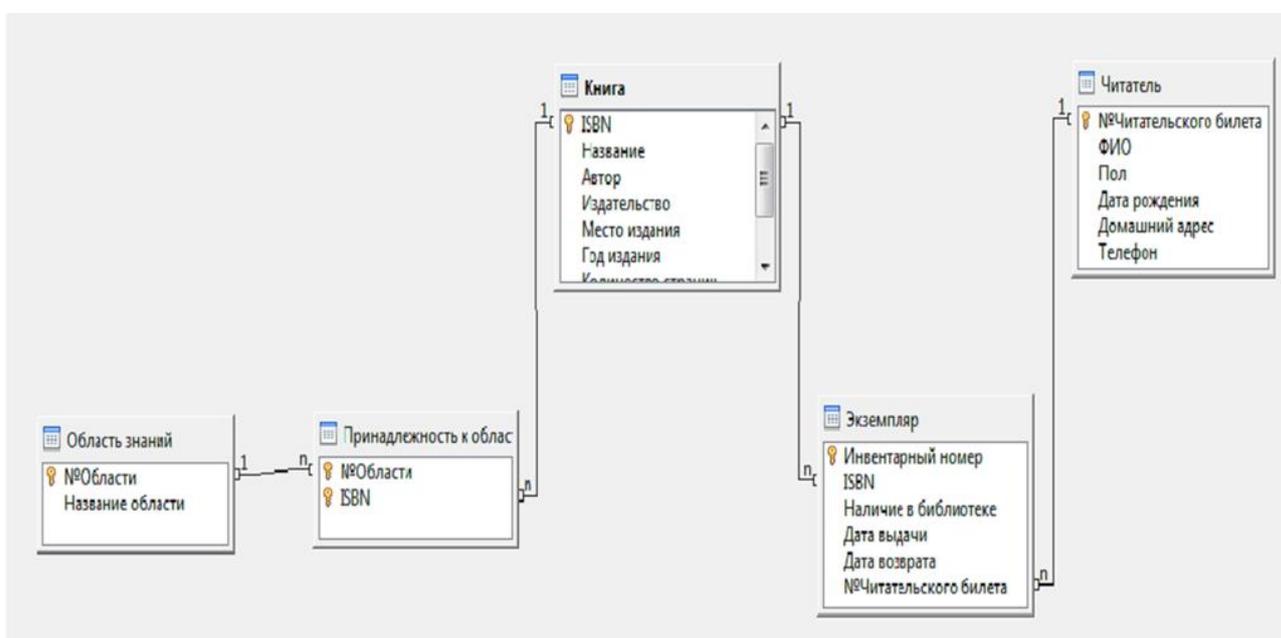


Рис.1.3 Физическое проектирование предметной области «Библиотека»

### Самостоятельная работа 1.1.

Для формирования отчета по самостоятельной работе рекомендуется использовать текстовый редактор Writer. Для составления схем и диаграмм использовать программу Dia. А для создания базы данных рекомендуется использовать СУБД Base.

1. Узнать у преподавателя тему для проектирования базы данных.
2. На основе анализа материалов Интернета составить описание предметной области.

3. Построить ER-диаграмму своей предметной области.
4. Сделать даталогическое проектирование БД т.е. перевести ER-диаграмму в реляционную схему и нормализовать отношения (таблицы).
5. Определить структуры таблиц и их связи, первичные и внешние ключи  
Реализовать базу данных в СУБД.
6. Заполнить таблицы, спроектировать запросы, формы и отчеты.

**Примерные темы для разработки баз данных:**

1. Школа
2. Детский сад
3. Университет
4. Компьютерная фирма
5. Кинотеатр
6. Аптека
7. Спортивный комплекс
8. Служба занятости
9. Больница
10. Автосалон
11. Бюро недвижимости
12. Туристическое агентство
13. Прокат
14. Военкомат
15. Салон красоты