

Моделирование данных

Введение и концептуальные модели данных



Московский государственный университет
имени Н.Э. Баумана

Факультет ИБМ

Июль 2024 года

Москва

Артемьев Валерий Иванович © 2024

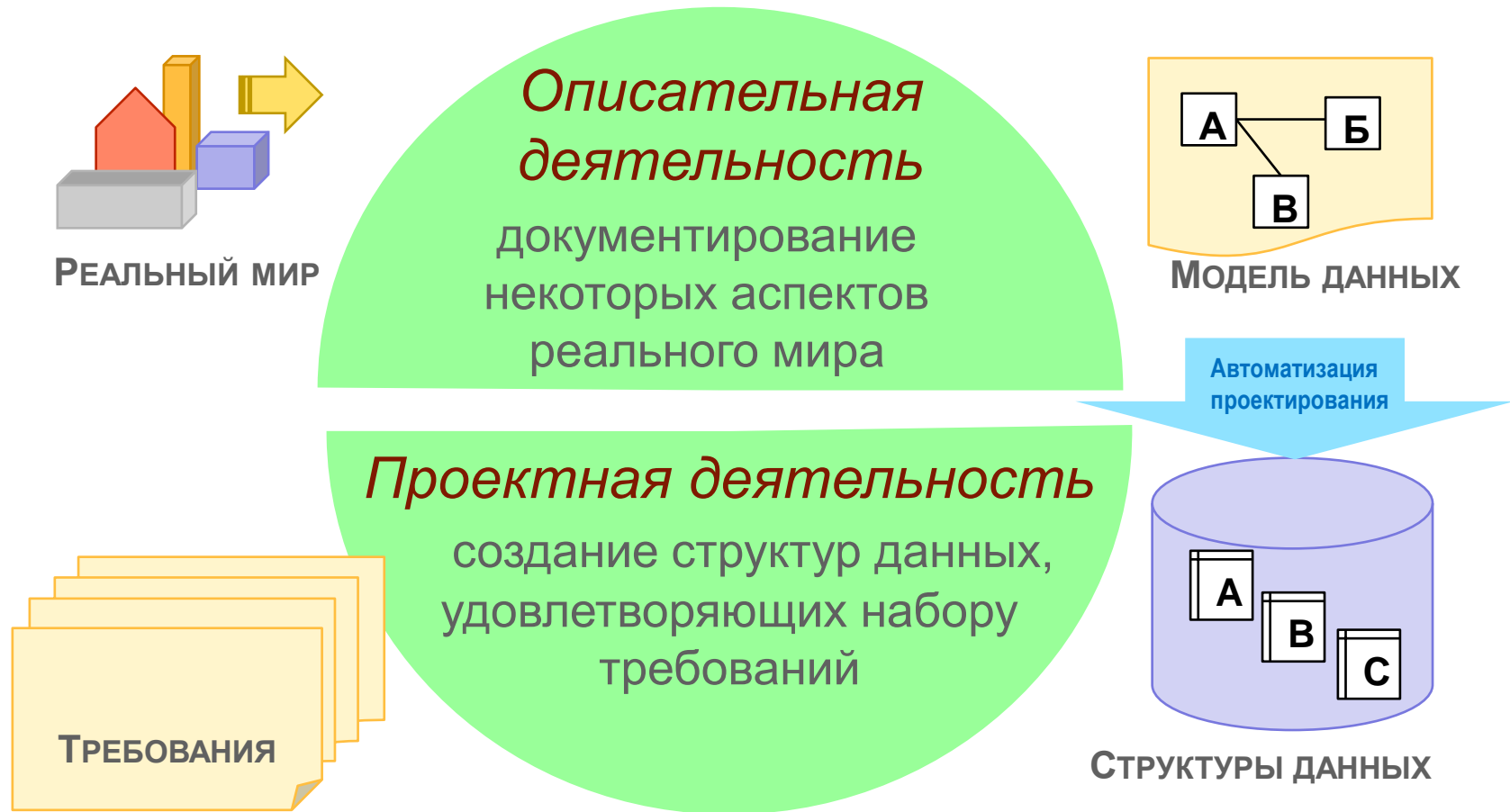
Модель данных

Описание структуры и содержания данных, а также бизнес-правил для представления реального объекта, процесса или концепции

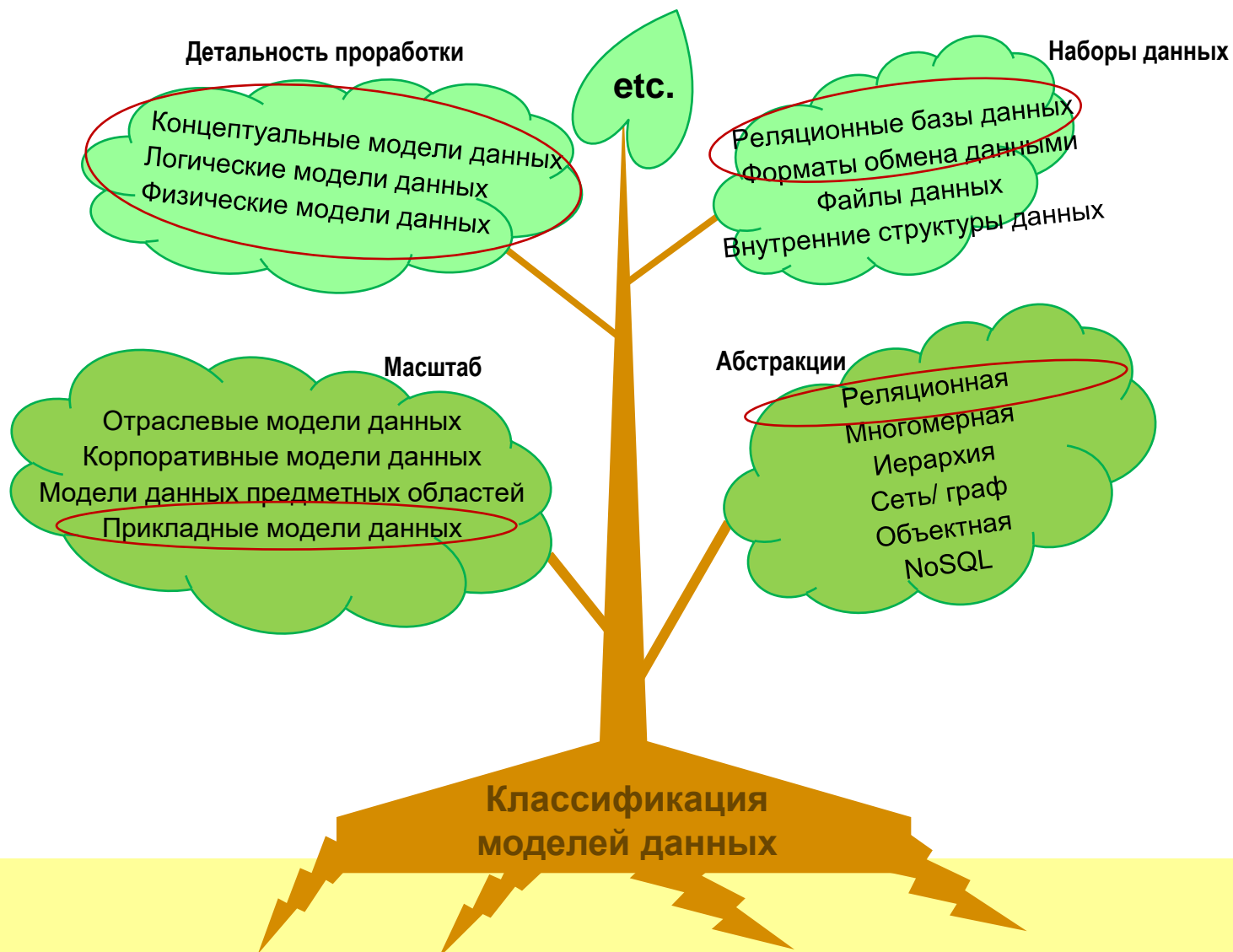


Модели данных от разработчиков СУБД включают ещё операции с данными!

Моделирование и проектирование данных



Какие бывают модели данных?



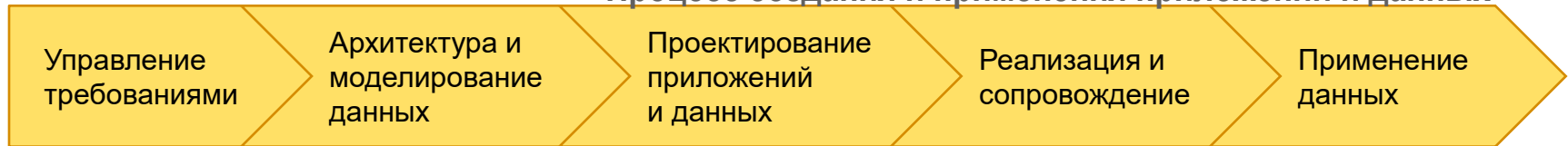
Для чего нужны модели данных?

Вид деятельности		Операции
Применение данных	Оценка пригодности данных для решения задач	R
	Связывание источников данных для анализа	R
	Формирование требований к данным	CUR
	Разработка/ сопровождение приложений	RU
	Тестирование баз данных и форматов	R
Проектирование данных	Разработка/ сопровождение баз данных и форматов	CUR
	Реинжиниринг баз данных и форматов	RCU
	Оценка влияния изменения структуры и состава данных	R
	Разработка/ сопровождение внутренней структуры данных	CUR
	Документирование информационного и программного обеспечения	CU
Архитектура данных	Семантическое моделирование предметной области	CU
	Разработка/ сопровождение корпоративной модели данных	CU
	Разработка/ сопровождение канонической модели обмена данными	CU
	Ведение репозитория моделей данных, бизнес-гlossария данных	CU

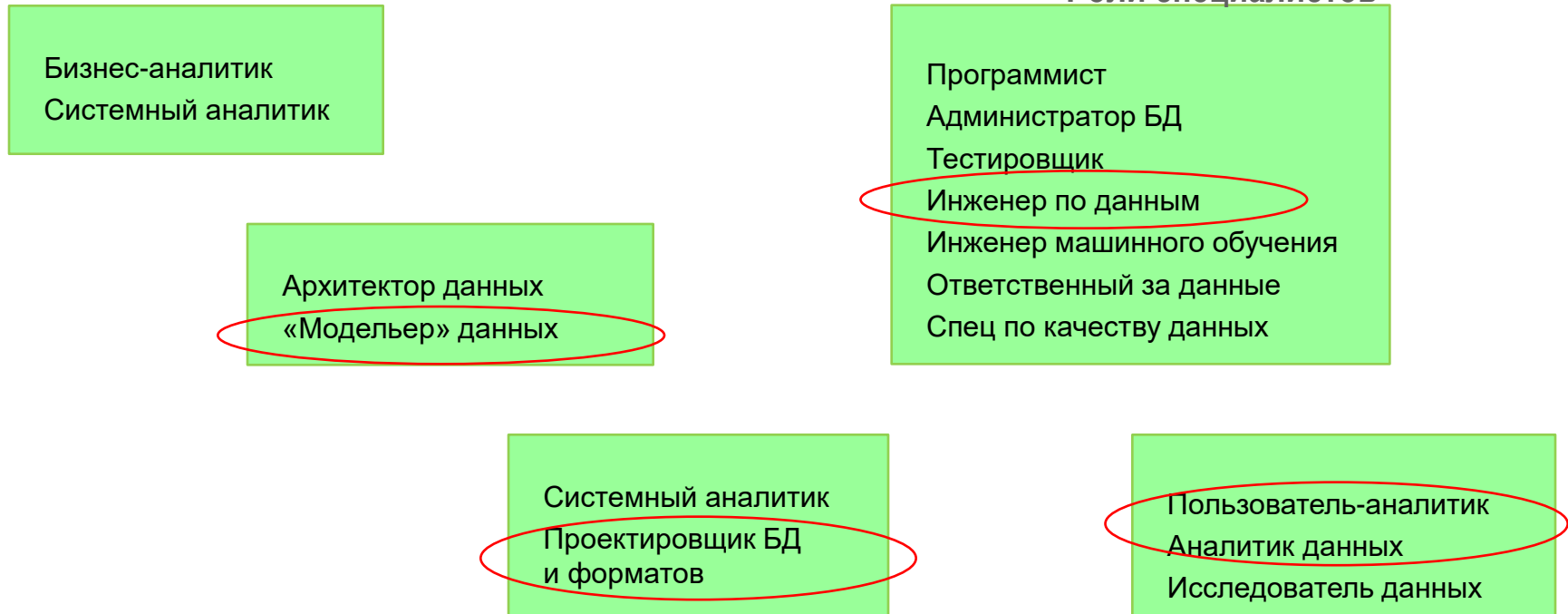
R – чтение, C – создание, U – изменение моделей

Кто имеет дело с моделями данных?

Процесс создания и применения приложений и данных



Роли специалистов

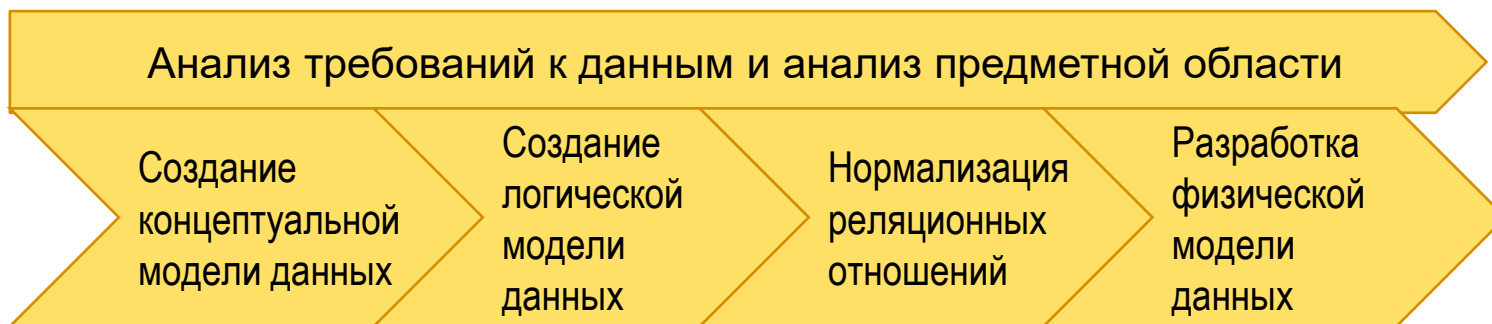


Выделены роли, которые мы будем играть на занятиях

Что мы должны получить в итоге моделирования данных?

Вид модели данных	Графическое представление	Таблица описания	Языковое описание
Концептуальная модель данных (КМД)	ER-диаграмма КМД Расширенная диаграмма КМД	Описание сущностей Описание связей	Таксономия КМД
Логическая модель данных (ЛМД)	ER-диаграмма ЛМД	Описание атрибутов	Функциональные зависимости для нормализации
Физическая модель данных (ФМД)	ER-диаграмма ФМД	Описание таблиц, колонок и ограничений целостности ФМД	Определения схем данных (DDL/SQL)

Процесс моделирования данных



Сущности и связи
ER-диаграмма КМД
Таксономия
Табличное описание

Ключи и атрибуты
ER-диаграмма ЛМД
Табличное описание

Функциональные зависимости
Нормальные формы

Домены и типы данных
Ограничения целостности данных
ER-диаграмма ФМД
Табличное описание

Концептуальная модель данных «сущность–связь»

Укрупнённая модель данных на основе реляционной абстракции «сущность-связь» для предметной области или приложения, которая отражает бизнес–сущности и связи между ними.

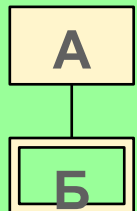
- Укрупнённое бизнес-описание структуры и состава данных
- Реляционная абстракция
- Документация по информационному обеспечению
- Средство коммуникации заказчиков и разработчиков приложений
- Средство проектирования данных
- Создание КМД позволяет избежать ошибок при реализации БД
- Не отражает специфику конкретной системы управления БД

О бизнес-сущностях

Бизнес–сущность (*entity*)– набор сведений о чём-либо из предметной области , что важно для бизнеса и должно фиксироваться и обрабатываться. Характеризуется самостоятельным существованием.

Выявляется при ответах на вопросы «кто», «что», «когда», «где», «зачем» и «как» (схема Бахмана).

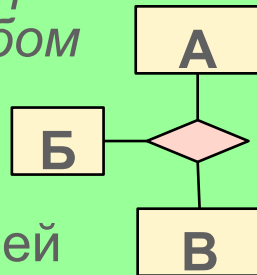
- Сущность должна иметь *уникальное имя* в рамках предметной области
- *Сильная (родительская) сущность* – существование её не зависит от какой-либо другой сущности
- *Слабая (дочерняя) сущность* – существование её зависит от какой-либо другой сущности
- Сущность представляет *набор экземпляров сущности*
- *Экземпляры бизнес–сущности* должны быть идентифицированы уникальным образом.



О связях сущностей

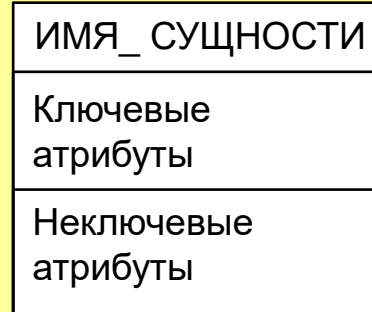
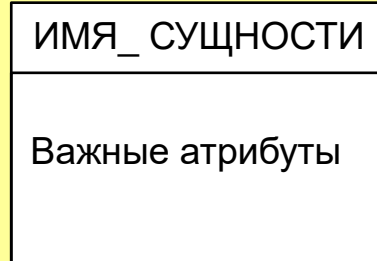
Связь (relationship) сущностей – осмысленная ассоциация между бизнес-сущностями, описывает взаимоотношение (правила) и возможность использования экземпляров смежных сущностей.

- *Степень связи* – число сущностей, охваченных связью. Различают унарные (рекурсивные), бинарные, тернарные и *n*-арные связи. В общем случае обозначаются ромбом
- *Экземпляр связи* – ассоциация между экземплярами сущностей
- *Кардинальность связи* – количество возможных связей экземпляров связанных сущностей. Бинарные связи 1:1, 1:N, M:N имеют спец. обозначения или указание явных значений (min; max)
- *Обязательные и необязательные связи.*



Нотация Мартина («вороньи лапки»)

Оформление блоков сущности

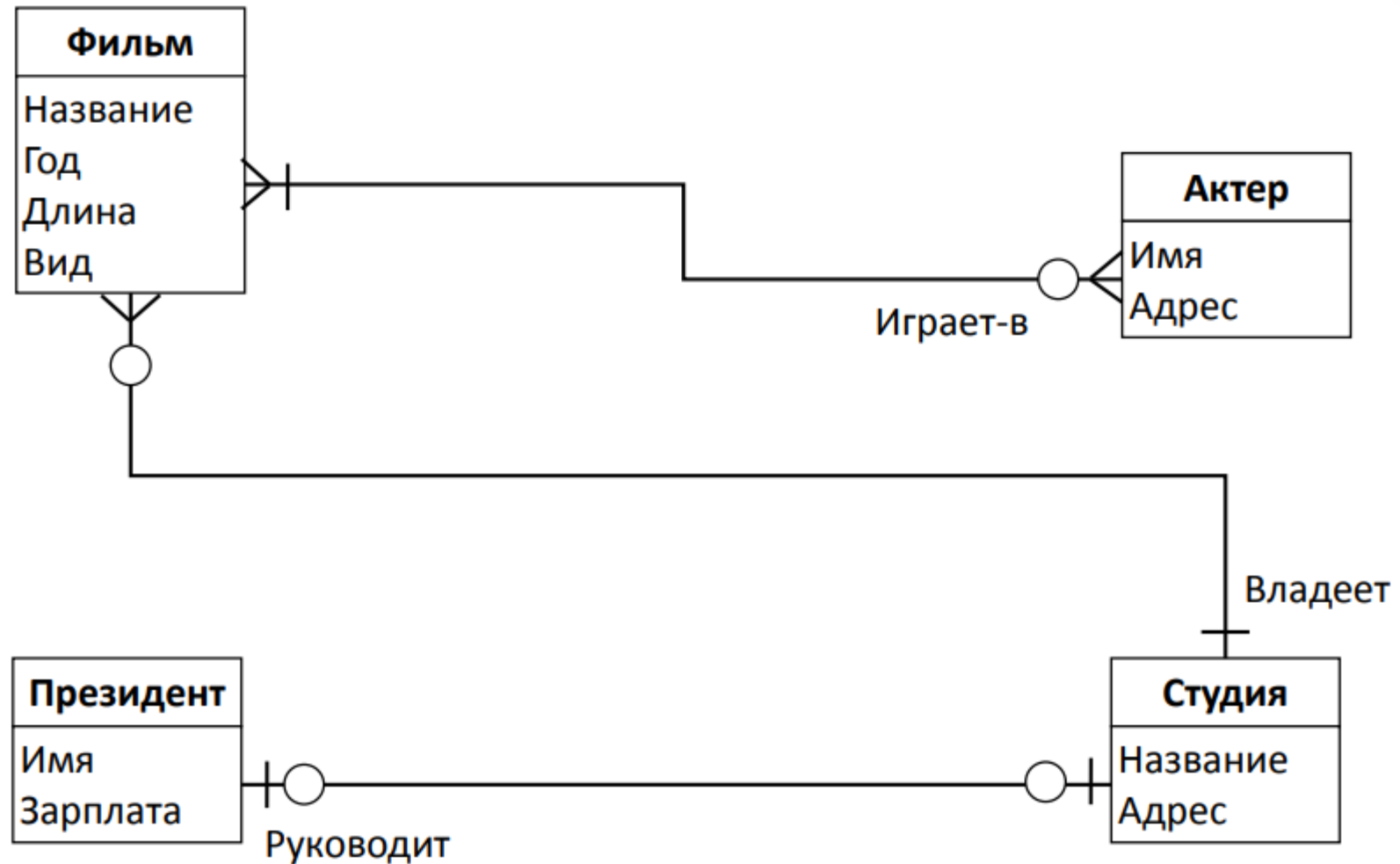


Оформление связей



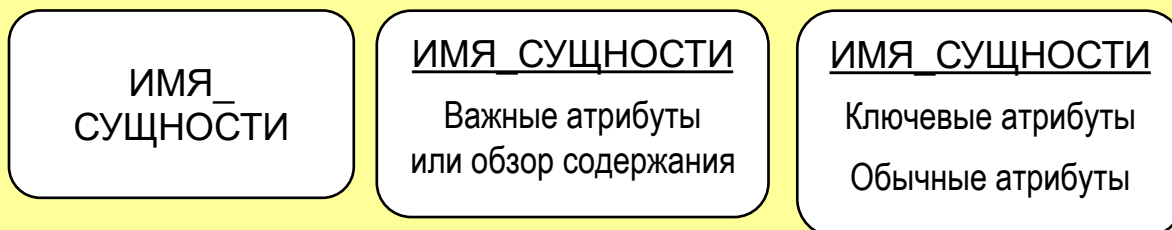
- Блок сущности – прямоугольник
- Сущности **не дублируются**
- Имя сущности – существительное **в единственном числе**
- **Атрибуты не указываются** или указаны **важные атрибуты** или **обзор содержания**
- На линии связи символически обозначаются кратности **появления экземпляров у каждой связанной сущности**
- На каждой стороне **штриховая линия задаёт необязательную связь** (МОЖЕТ), **сплошная – обязательную связь** (ДОЛЖЕН)
- Связь может **именоваться глаголами** по-русски с двух сторон, реже с одной стороны

Пример ER-диаграммы в нотации Мартина «вороньи лапки»



Нотация Баркера

Оформление блоков сущности



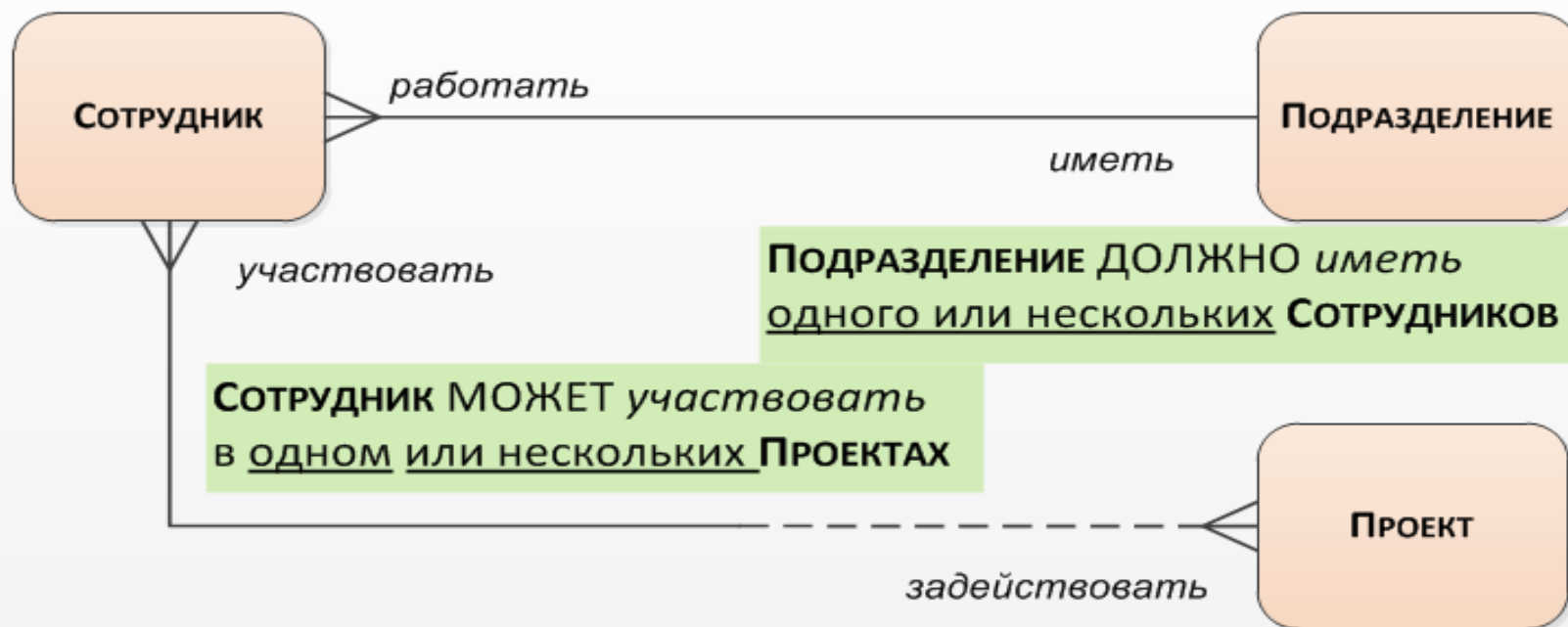
Оформление связей



- Блок сущности – прямоугольник со скруглёнными углами
- Сущности **не дублируются**
- Имя сущности – существительное **в единственном числе**
- **Атрибуты не указываются** или указаны **важные атрибуты** или **обзор содержания**
- На линии связи символически обозначаются кратности **появления экземпляров у каждой связанной сущности**
- На каждой стороне **штриховая линия задаёт необязательную связь** (МОЖЕТ), **сплошная – обязательную связь** (ДОЛЖЕН)
- Связь может **именоваться глаголами** по-русски с двух сторон, реже с одной стороны

Как читать связи на ER-диаграмме?

СОТРУДНИК ДОЛЖЕН *работать* в одном ПОДРАЗДЕЛЕНИИ



Табличное описание концептуальной модели данных

Описание сущностей

- Предметная область
- Имя сущности
- Описание сущности
- Тип сущности
(реестр, справочник, факты, кросс-сущность)
- Имя супертипа
- Важные атрибуты

Описание связей

- Предметная область
- Имя связи
- Связанные сущности
- Описание связи
- Тип связи (один к одному, один ко многим, многие ко многим и др.)
- Обязательность связи

Таксономия концептуальной модели данных

Вовлечённая сторона – *концепт верхнего уровня*

- + Вид легальности
- + Вид вовлечённой стороны

+ Организация

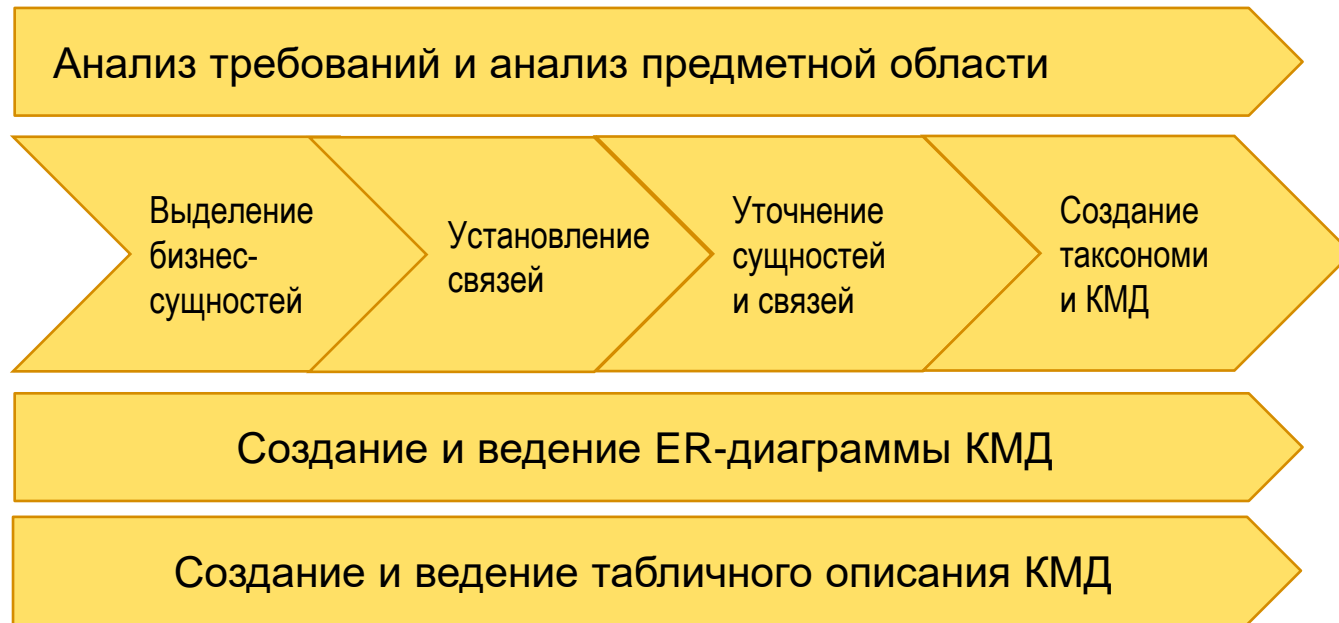
- + Вид структуры организации
- + Назначение организации
- + Состояние ЖЦ организации
- + Экономическая ориентация

+ Индивид – *концепт более низкого уровня*

- + Род индивида – *признак классификации*
 - мужской – *значение признака классификации*
 - женский
- + Состояние ЖЦ индивида
- + Состояние здоровья
- + Состояние занятости
- + Род занятий
- + Семейное положение

Иерархическая классификация данных
Позволяет раскрывать/ свёртывать уровни
Основа для разработки справочников
Создаётся в MS Word в режиме структуры

Процедуры создания концептуальной модели данных



Рекомендуемые инструменты моделирования данных

■ Средства моделирования

- CA Erwin Data Modeler
- MS Visio
- DBeaver

■ Редакторы диаграмм

- Diagrams (draw.io)
- yEd

■ Офисные приложения

- MS PowerPoint
- MS Excel
- MS Word

■ Карты идей

- Freemind
- Mindmup
- Xmind

Пример: Домашняя библиотека

Создать каталог домашней библиотеки

■ Атрибуты поиска книги:

- Тема, вид издания
- Название издания и/ или ключевые слова
- Фамилия, имя (инициалы) автора, редактора, составителя, переводчика или художника
- Название и место издательства
- Год выпуска.

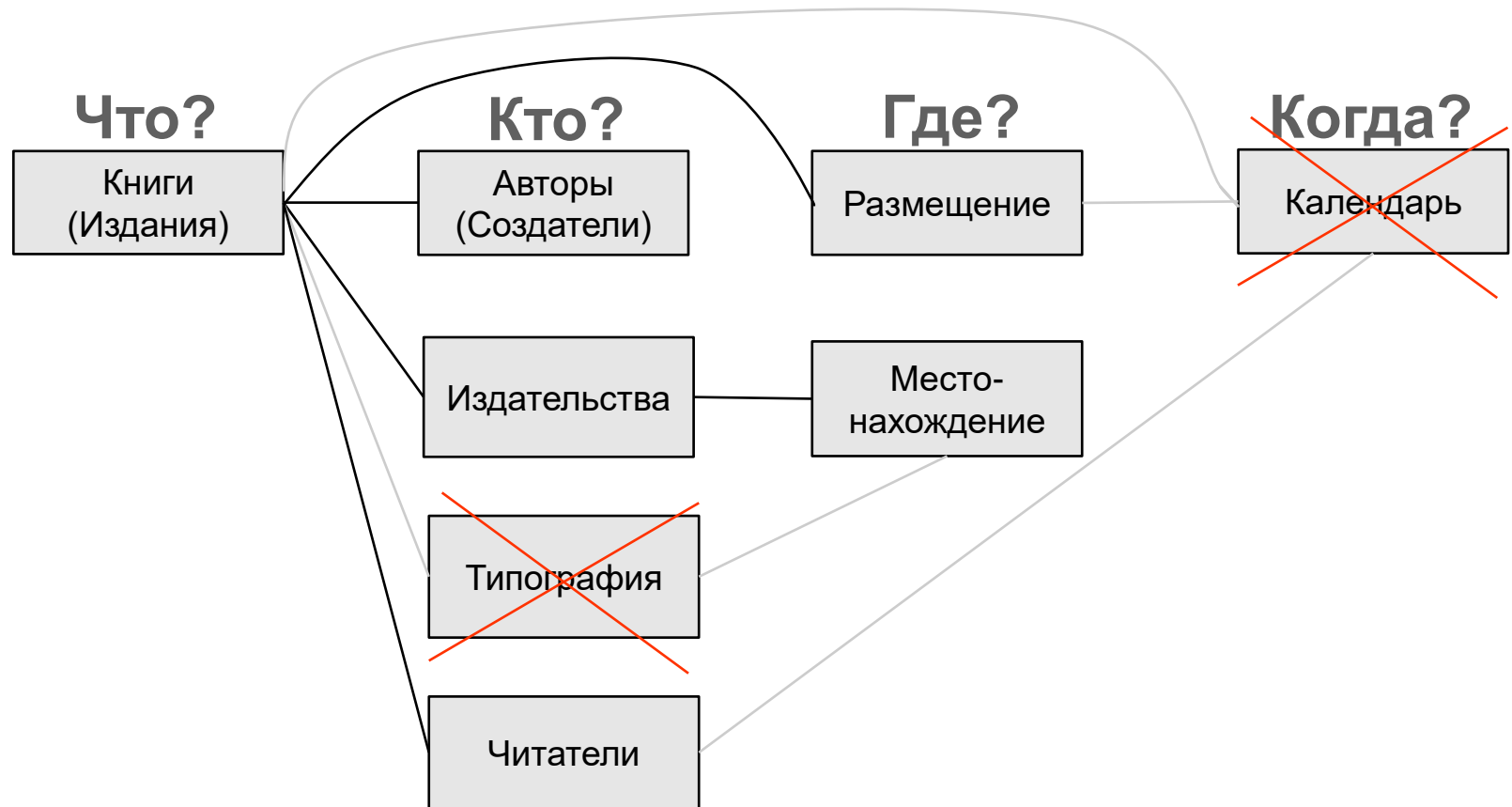
■ Результаты поиска:

- Хранимые атрибуты книги (включая аннотацию, сведения об оригинале перевода)
- Место хранения издания (шкаф, полка)
- Текущий держатель (читатель) издания, его контакты

■ В библиотеке каждая книга присутствует в *одном экземпляре*

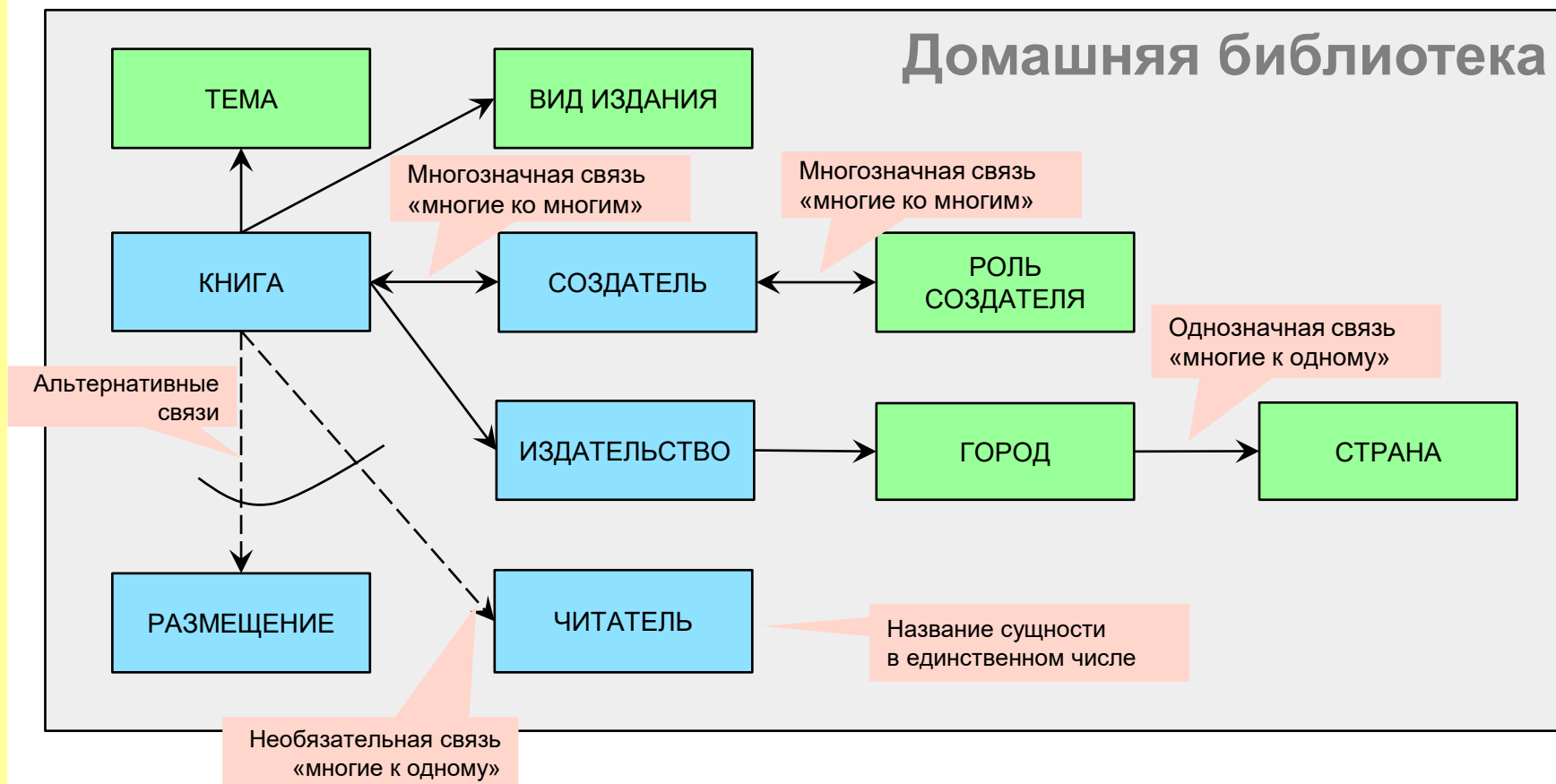
Создание концептуальной модели данных «Домашняя библиотека»

Анализ предметной области. Выделение сущностей

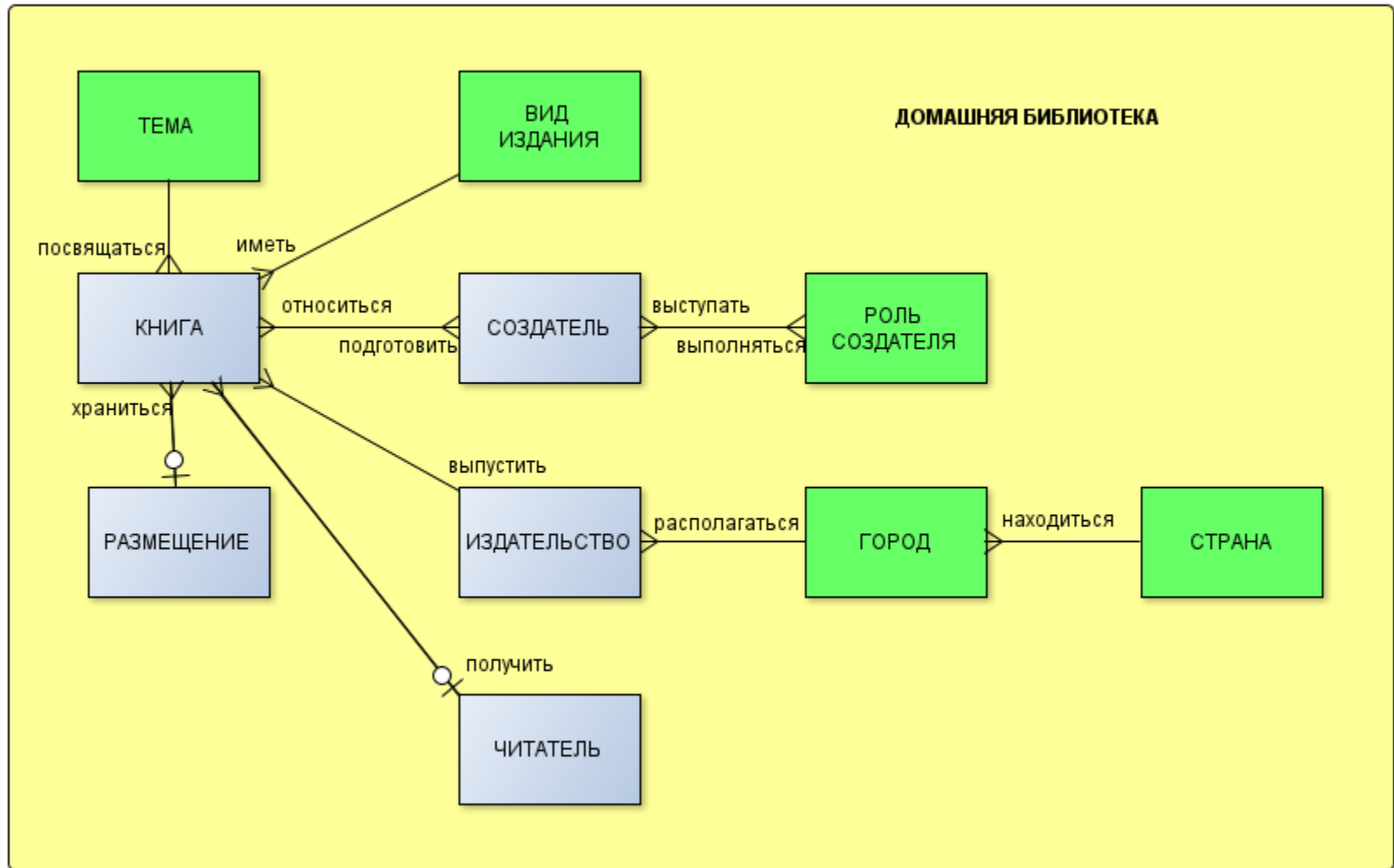


ER-диаграмма КМД «Домашняя библиотека»

Создание КМД. Уточнение сущностей и связей



ER-диаграмма КМД, подготовленная в приложении уЕd



Табличное описание КМД «Домашняя библиотека»

Описание сущностей КМД предметной области "Домашняя библиотека"

Имя сущности	Описание сущности	Тип сущности	Имя супертипа	Важные атрибуты
КНИГА	Единственный экземпляр книги в бумажном переплёте	Реестр	ТЕМА	Название, аннотация, год выпуска
СОЗДАТЕЛЬ	Имя создателя книги в определённой роли	Реестр		Фамилия, имя (инициалы) автора, редактора и т.п.
ИЗДАТЕЛЬСТВО	Название издательства, выпустившего книгу	Реестр	ГОРОД	
РАЗМЕЩЕНИЕ	Место хранения книги	Реестр		Обозначение шкафа и полки
ЧИТАТЕЛЬ	Имя читателя книги	Реестр		
ТЕМА	Тема, к которой относится книга	Классификатор		
ВИД_ИЗДАНИЯ	Вид издания книги: учебник, роман			
РОЛЬ_СОЗДАТЕЛЯ	Роль создателя книги: автор, редактор			
ГОРОД	Название города, в котором находится издательство			
СТРАНА	Название страны, где находится издательство			

Описание связей КМД предметной области "Домашняя библиотека"

Глагол прямой (обратной) связи	Связанные сущности	Описание связи	Тип связи	Обязательность связи
относиться	КНИГА - ТЕМА	Одна или несколько книг ДОЛЖНЫ относиться к одной теме	многие к одному	да
иметь	КНИГА - ВИД_ИЗДАНИЯ	Одна или несколько книг ДОЛЖНЫ иметь один вид издания	многие к одному	да
подготовить (иметь отношение)	КНИГА - СОЗДАТЕЛЬ	Книгу ДОЛЖНЫ подготовить один или несколько создателей. Создатель ДОЛЖЕН иметь отношение к одной или нескольким книгам	многие ко многим	да
(выпустить)	КНИГА - ИЗДАТЕЛЬСТВО	Издательство ДОЛЖНО выпустить одну или несколько книг	многие к одному	да
храниться	КНИГА - РАЗМЕЩЕНИЕ	В одном месте МОЖЕТ храниться несколько книг. Выданная книга не имеет места	многие к одному	нет
(взять)	КНИГА - ЧИТАТЕЛЬ	Читатель МОЖЕТ взять несколько книг. Хранимая книга не имеет читателя	многие к одному	нет
выступать в (выполняться)	СОЗДАТЕЛЬ - РОЛЬ_СОЗДАТЕЛЯ	Создатель ДОЛЖЕН выступать в одной или нескольких ролях. Определённая роль ДОЛЖНА выполняться одним или несколькими создателями	многие ко многим	да
иметь место	ИЗДАТЕЛЬСТВО - ГОРОД	Одно или несколько издательств ДОЛЖНЫ иметь место в одном городе	многие к одному	да

Таксономия КМД

«Домашняя библиотека»

- ⊕ ДОМАШНЯЯ БИБЛИОТЕКА
 - ⊕ КНИГА
 - ⊕ ТЕМА
 - Информационные технологии
 - Архитектура
 - Живопись
 - Лингвистика
 - История
 - Путешествия
 - Художественная литература
 - ⊕ ВИД ИЗДАНИЯ
 - Роман
 - Сборник стихов
 - Собрание сочинений
 - Учебник
 - Путеводитель
 - Разговорник
 - Словарь
 - Альбом
 - Сказки
 - ⊕ СОЗДАТЕЛЬ
 - ⊕ РОЛЬ СОЗДАТЕЛЯ
 - Редактор
 - Составитель
 - Автор
 - Переводчик
 - Художник

Терпения и удачи всем, кто связан
с моделированием данных

Спасибо за внимание!

Валерий Иванович Артемьев

Департамент статистики и управления данными
Банк России

Тел.: +7(495) 753-96-25

e-mail: avi@cbr.ru