

Модели данных

А1. Введение и концептуальные модели данных



Московский государственный технический университет
имени Н.Э. Баумана

Факультет ИБМ

Июль 2024 года

Москва

Артемьев Валерий Иванович © 2024

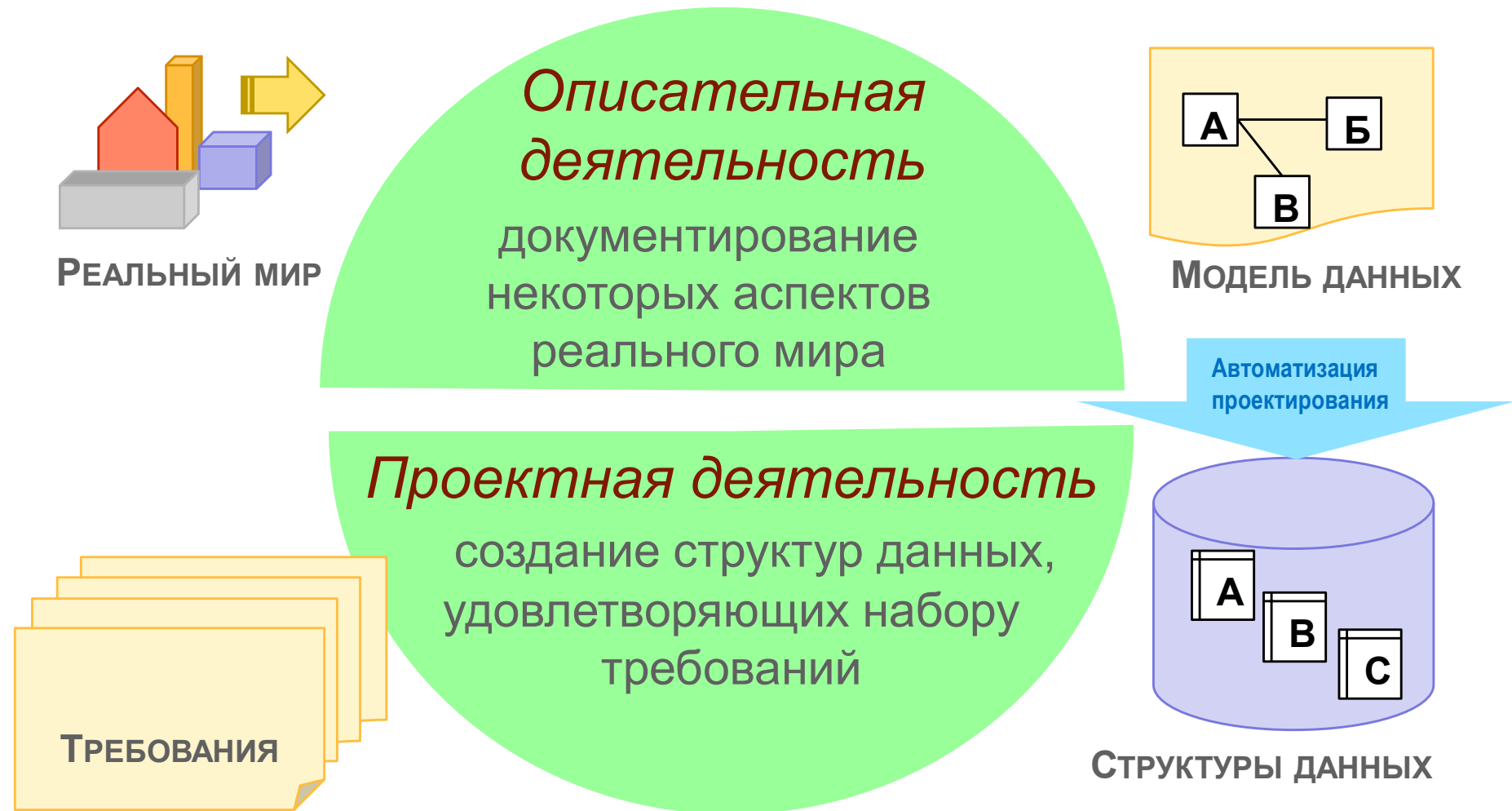
Модель данных

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?



Выделены роли, которые мы будем играть на лабораторных

Подходы к моделированию и проектированию данных



Сверху вниз (up down)

- Анализ на верхнем уровне требований и предметной области
- Выделение бизнес-сущностей и связей по схеме Бахмана (ответы на вопросы «кто», «что», «когда», «где», «зачем» и «как»)
- Детализация описаний сущностей и связей: атрибуты, ключи, типы данных, ограничения целостности.



Снизу вверх (bottom up)

- Анализ на нижнем уровне требований и предметной области
- Группировка свойств для выявления сущностей и связей
- Декомпозиция при нормализации
- Детализация описаний сущностей и связей: атрибуты, ключи, типы данных, ограничения целостности

Процесс моделирования данных сверху вниз



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.

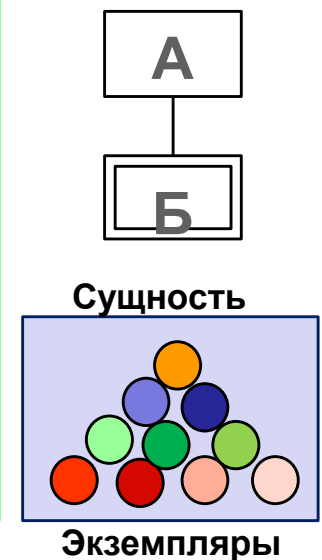
- Укрупнённое бизнес-описание структуры и состава данных
- Реляционная абстракция
- Документация по информационному обеспечению
- Средство коммуникации заказчиков и разработчиков приложений
- Создание КМД позволяет избежать ошибок при реализации БД
- Не отражает специфику конкретной системы управления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

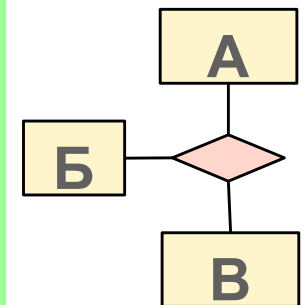
- Сущность должна иметь *уникальное имя* в рамках предметной области
- *Сильная (родительская) сущность* – существование её не зависит от какой-либо другой сущности
- *Слабая (дочерняя) сущность* – существование её зависит от какой-либо другой сущности
- Сущность представляет *набор экземпляров сущности*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

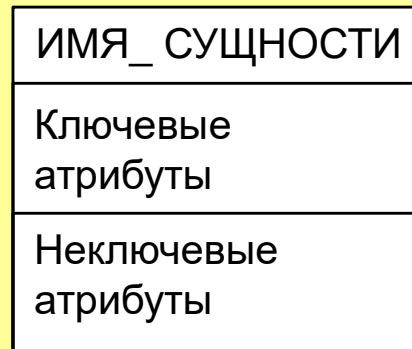
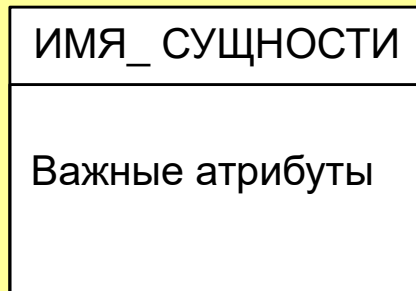


Какие артефакты мы получим при моделировании данных?

Вид модели данных	Графическое представление	Табличное описание	Языковое описание
Концептуальная модель данных (КМД)	ER-диаграмма КМД	Описание сущностей Описание связей	Таксономия КМД
Логическая модель данных (ЛМД)	ER-диаграмма ЛМД	Описание атрибутов	Функциональные зависимости
Физическая модель данных (ФМД)	ER-диаграмма ФМД	Описание таблиц БД, колонок и ограничений целостности ФМД	Сценарии определения схем данных (DDL/SQL)

Нотация Мартина («вороньи лапки»)

Оформление блоков сущности

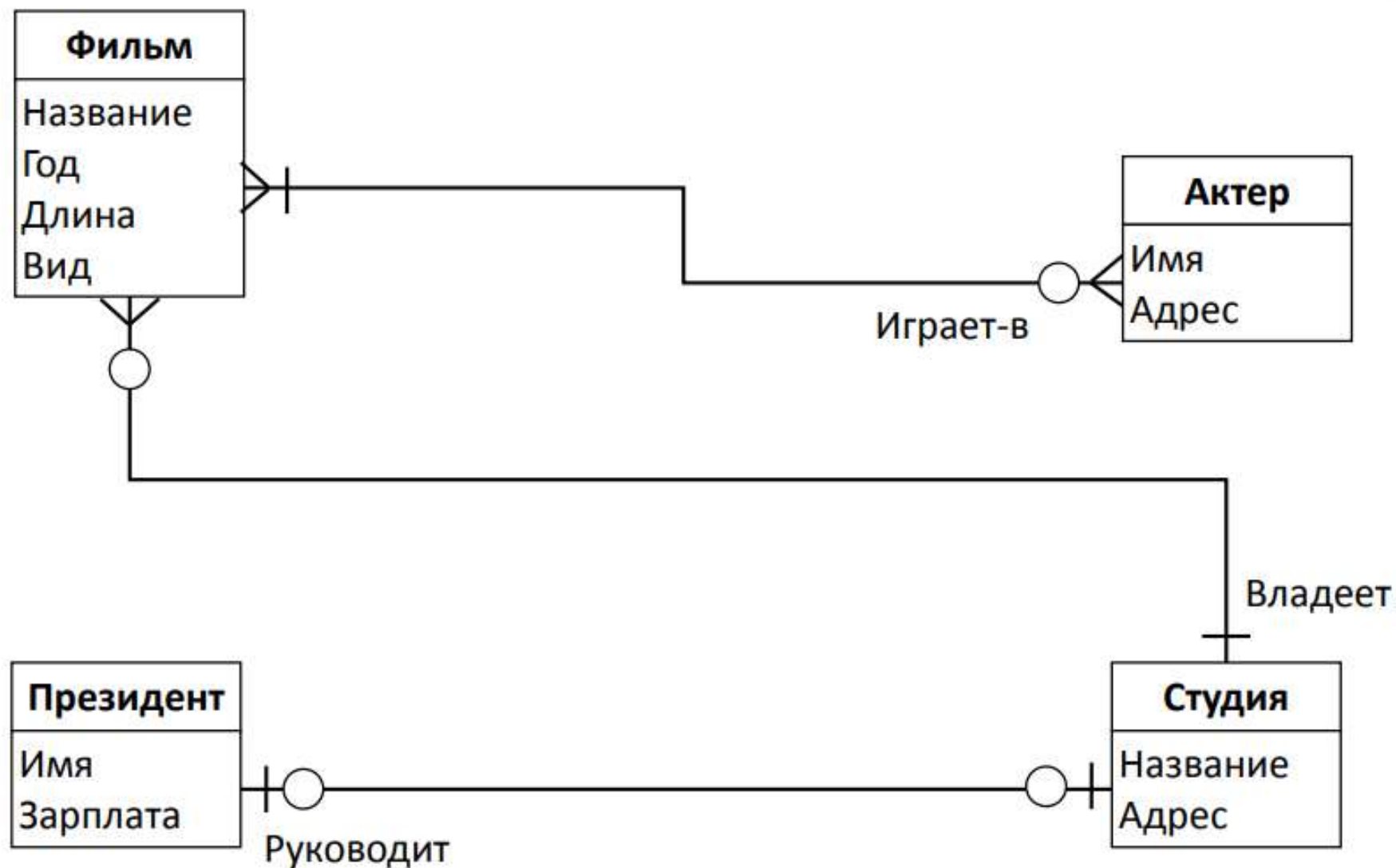


Оформление связей



- 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

ИМЯ_
СУЩНОСТИ

ИМЯ СУЩНОСТИ

Важные атрибуты
или обзор содержания

ИМЯ СУЩНОСТИ

Ключевые атрибуты
Обычные атрибуты

Оформление связей



один к одному



многие к одному



многие ко многим



необязательная
связь

- 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

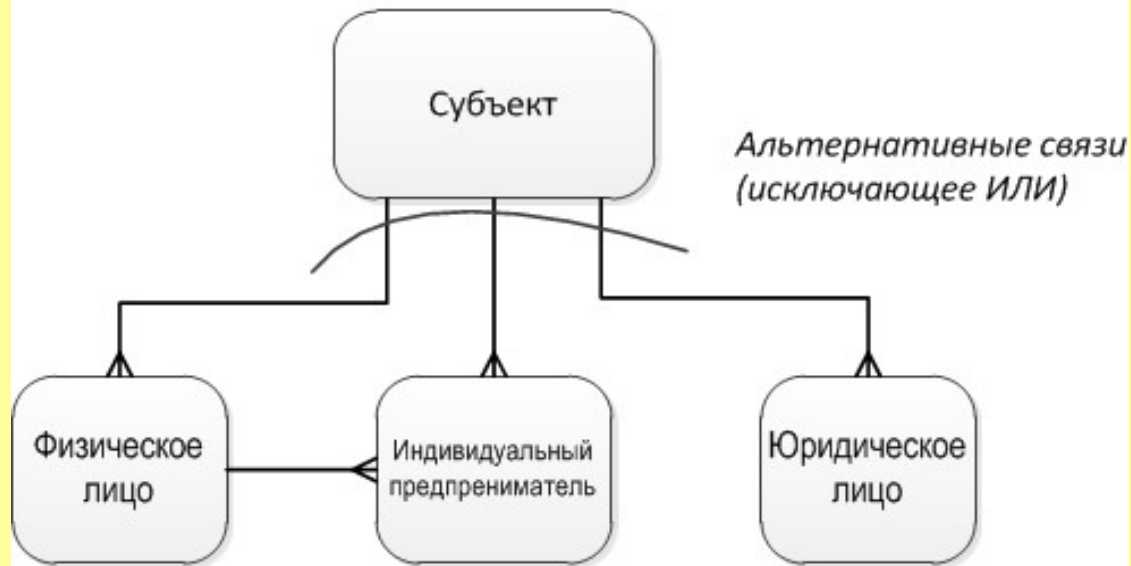
Как читать связи на ER-диаграмме?

Сотрудник ДОЛЖЕН работать в одном Подразделении

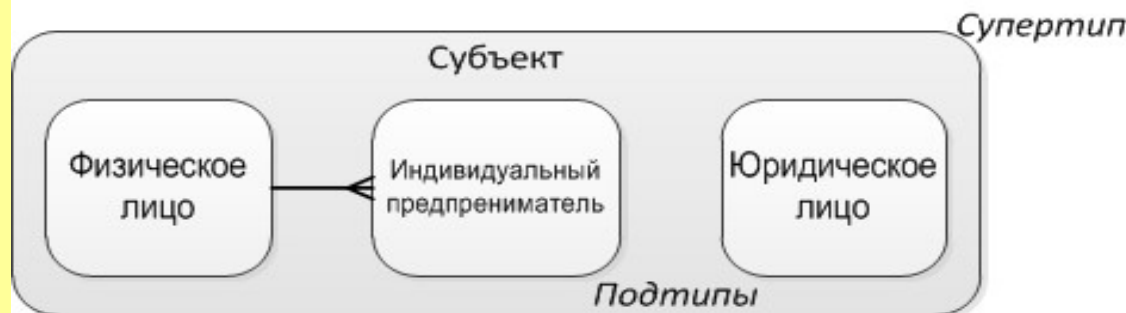


Пример расширенной ER-диаграммы КМД с супертипами и подтипами

Вариант с исключаящими связями



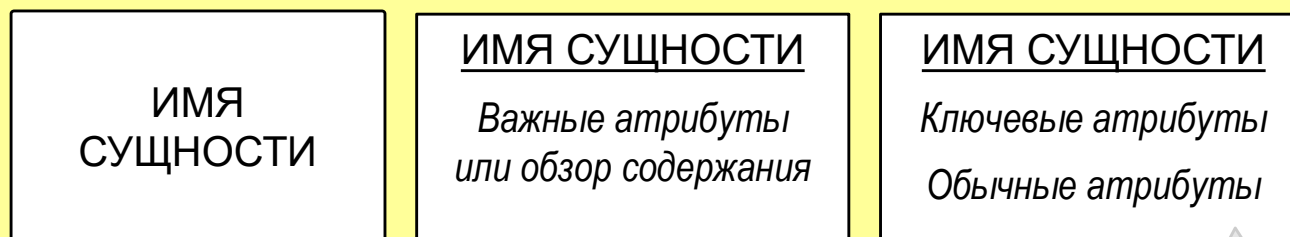
Вариант с супертипами и подтипами



- В нотации Баркера альтернативные связи обозначены пересекающей их дугой
- Расширенная диаграмма – вложение блоков подтипов в блок супертипа
- Супертип – родительская сущность.
- Супертип предоставляет общие свойства для подтипов.
- Подтип – дочерняя зависимая сущность.
- Подтипы *наследуют общие свойства* супертипа, также имеют свои свойства.
- Связи подтипов с супертипами – альтернативные (исключающее ИЛИ).
- *Компактный способ представления.* Как он поддержан в инструментах?

Упрощенная нотация

Оформление блоков сущности



для ЛМД

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 стрелками обозначаются кратности *появления экземпляров у каждой связанной сущности*
- На каждой стороне можно задать *необязательную связь* (МОЖЕТ) с помощью указания диапазона кратности 0-1 или 0-N
- Связь может *именоваться глаголами* по-русски с двух сторон, реже с одной стороны

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

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

Процедуры создания концептуальной модели «сверху вниз»



Рекомендуемые инструменты моделирования данных

- **CASE-средства:** построение ER-диаграмм и генерация DDL/SQL для создания БД
 - CA Erwin Data Modeler
 - MS Visio
 - DBeaver
- **Графические редакторы:** построение ER-диаграмм
 - Diagrams (draw.io)
 - yEd
 - MS PowerPoint
- **Офисные приложения:** создание табличных описаний и таксономий
 - MS Excel
 - MS Word

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**

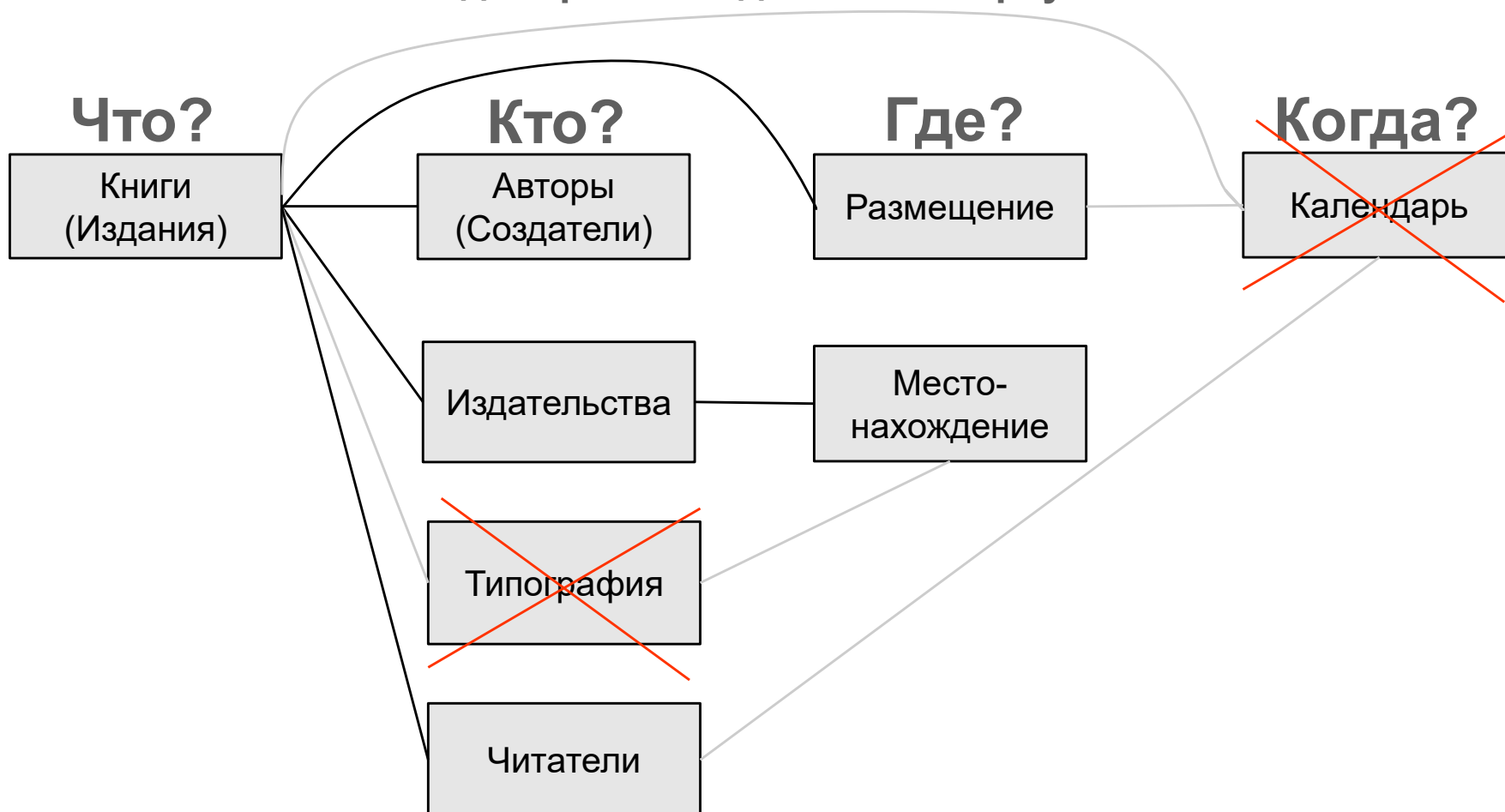
Общеупотребительные категории бизнес-сущностей (схема Бахмана)

Категория	Определение	Примеры
Кто?	Персона или организация, представляющие интерес для бизнеса. Часто это однородные группы или роли. Лицо или организация включается в различные группы и играет различные роли	Работодатель, Пациент, Игрок, Подозреваемый, Клиент, Студент, Поставщик, Пассажир
Что?	Предмет интереса предприятия. Обычно используется для определения категорий продукции или услуг. Важно чётко определять атрибуты категорий, типов и т. п.	Продукт, Услуга, Книга, Поставка, Товар, Курс, Саундтрек, Авто, Фотография
Когда?	Календарные даты, сроки или периоды, интересующие организацию. Когда именно действует то, что важно для бизнеса	Время, Дата, Месяц, Квартал, Год, Семестр, Расписание, Срок
Где?	Места локализации интересов организации, включая фактические физические, почтовые и электронные адреса, где ведётся бизнес	Почтовый адрес, Пункт выдачи, URL веб-сайта, IP-адрес
Почему?	События или операции, представляющие интерес для организации и держащие бизнес на плаву. Почему и для чего делается то, что делается	Заказ, Возврат, Претензия, Жалоба, Депозит, Отзыв, Запрос, Замена, Рекламация
Как?	Документы, относящиеся к интересующим событиям, служат подтверждением факта того или иного события. Как мы определяем, имело ли место то или иное событие	Счет-фактура, Договор, Контракт, Заказ, Квитанция об уплате, Товарная накладная
Сколько?	Итоги, суммы и т. п. Сводные и контрольные данные по всем остальным категориям	Продажи, Количество единиц товара, Платежи, Баланс

Создание КМД «Домашняя библиотека».

Выделение бизнес-сущностей

Моделирование данных «сверху вниз» по схеме Бахмана



Создание КМД «Домашняя библиотека».

Выделение сущностей как групп свойств

№ п/п	Название книги	Аннотация	Авторы и др. создатели	Тема	Ключевые слова	Вид издания	Издательство	Город	Год	Страниц	№ шкафа	№ полки	Читатель	Контакты	Дата выдач	Срок
1	Заставьте данные говорить	Это книга о создании дашбордов в Excel, но приемы визуализации из нее актуальны и для BI-платформ. Дашборд заменяет десятки страниц и слайдов. В основе – 13-летний опыт разработки интерактивных дашбордов для бизнеса. А еще – проблемы клиентов и студентов.	Алексей Колоколов	Информационные технологии	Визуализация данных, дашборды, Excel	Практическое руководство	Институт бизнес-аналитики	Москва	2022	223	1	3				
2	Data Quality. The Accuracy Dimension	Understanding the concepts of accurate data is fundamental to improving the ways we collect and use data.	Jack E. Olson	Информационные технологии	Качество данных	Монография	Morgan Kaufmann Publishers	San Francisco	2003	294	1	4				
3	Data Modeling Made Simple	Practical Guide on Data Modelling for Business and IT Professionals	Steve Hoberman, Bill Inmon, Michael Blaha, Graeme Simsion	Информационные технологии	Моделирование данных, модели данных	Практическое руководство	Technics Publications, LLC	Basking Ridge	2016	234	1	4				
4	Модель "сущность - связь"	В учебном пособии обсуждаются основные концепции моделирования данных, основанного на методе «сущность – связь». Демонстрируется технология применения методов разработки информационной модели в инструментальной среде популярной CASE-системы ERwin. На конкретном содержательном примере закрепляются навыки практической работы по построению информационной модели.	В.Ю. Кара-Ушанов	Информационные технологии	Моделирование данных, модели данных "сущность - связь", CASE-система, Erwin	Учебное пособие	Уральский федеральный университет	Екатеринбург	2017	63	1	4				
4	Дашборд для директора. Как делать управленческие отчеты красивыми и понятными	Книга полезна тем, кто готовит отчеты своим руководителям, клиентам и партнерам. Обычно это экономисты, финансисты, маркетологи, менеджеры проектов и продуктов. Должность может называться как угодно, главное – это роль бизнес-аналитика, того, кто превращает сырые данные в информацию для принятия решений.	Алексей Колоколов	Информационные технологии	Визуализация данных, дашборды, Excel	Учебное пособие	Институт бизнес-аналитики		2019	129	1	3				

Книги

ИЗДАТЕЛЬСТВА

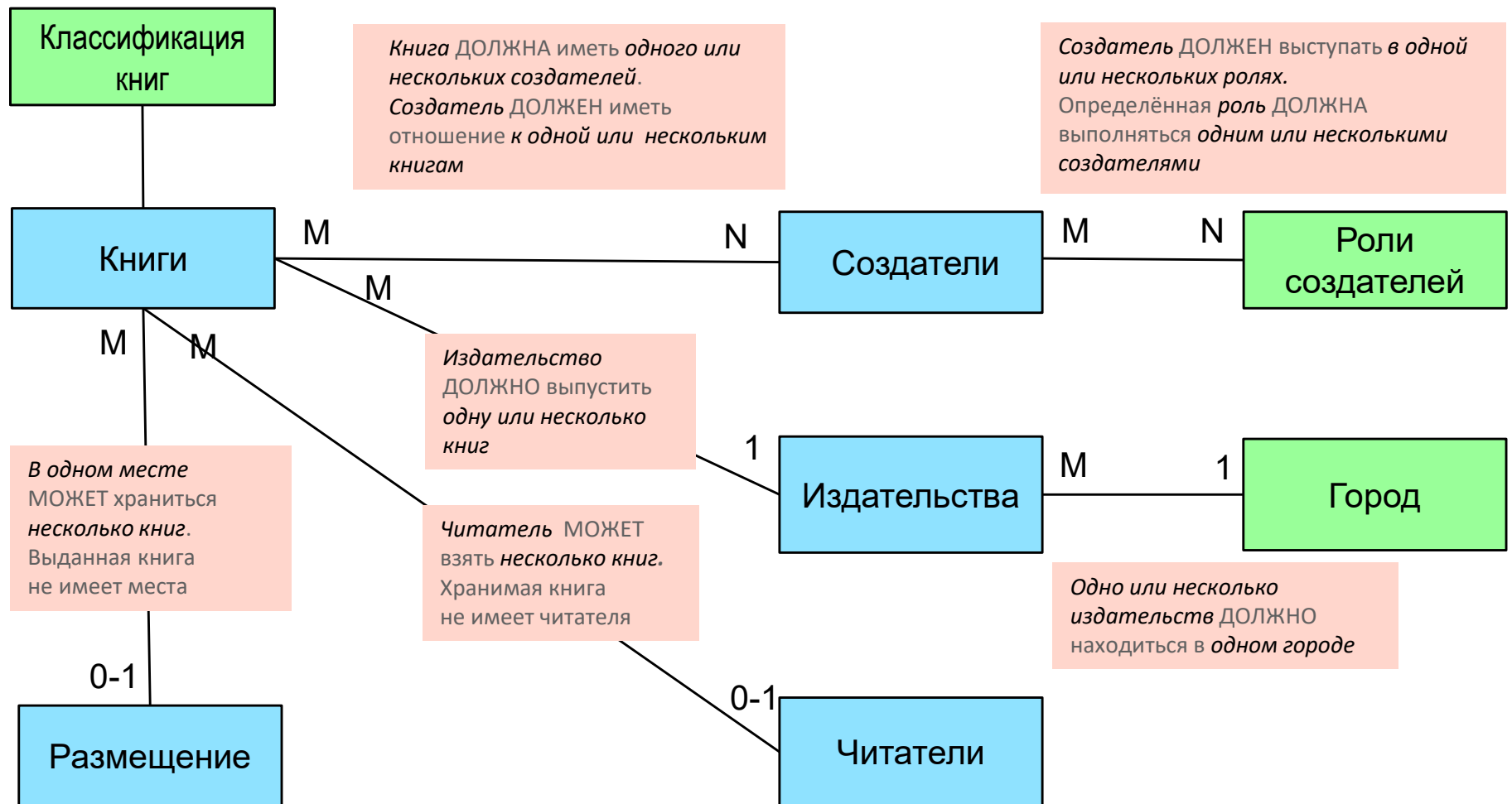
РАЗМЕЩЕНИЕ

Читатель

Моделирование данных «снизу вверх» обобщения свойств

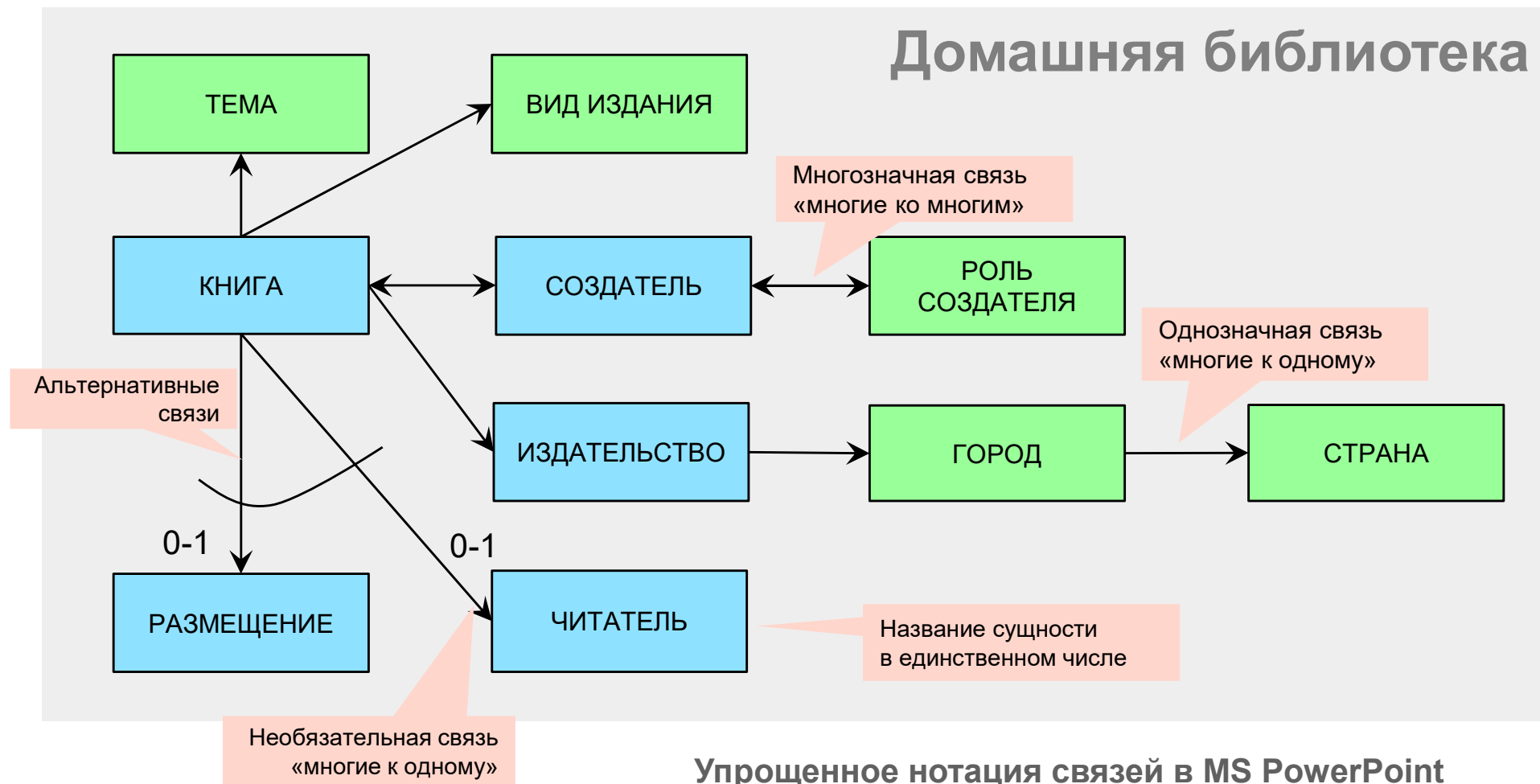
Создание КМД «Домашняя библиотека».

Установление видов связей

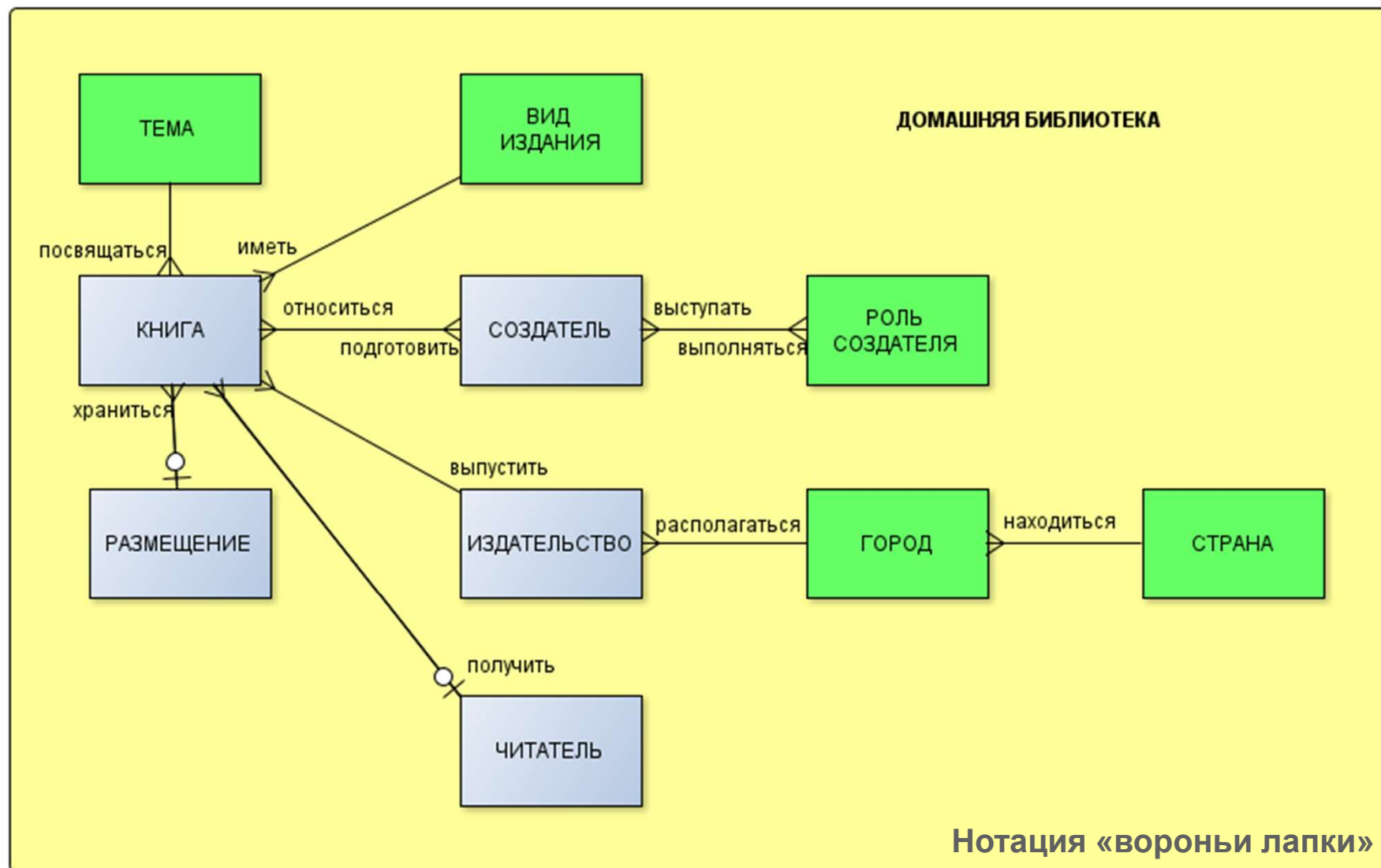


Создание КМД «Домашняя библиотека».

Уточнение сущностей и связей

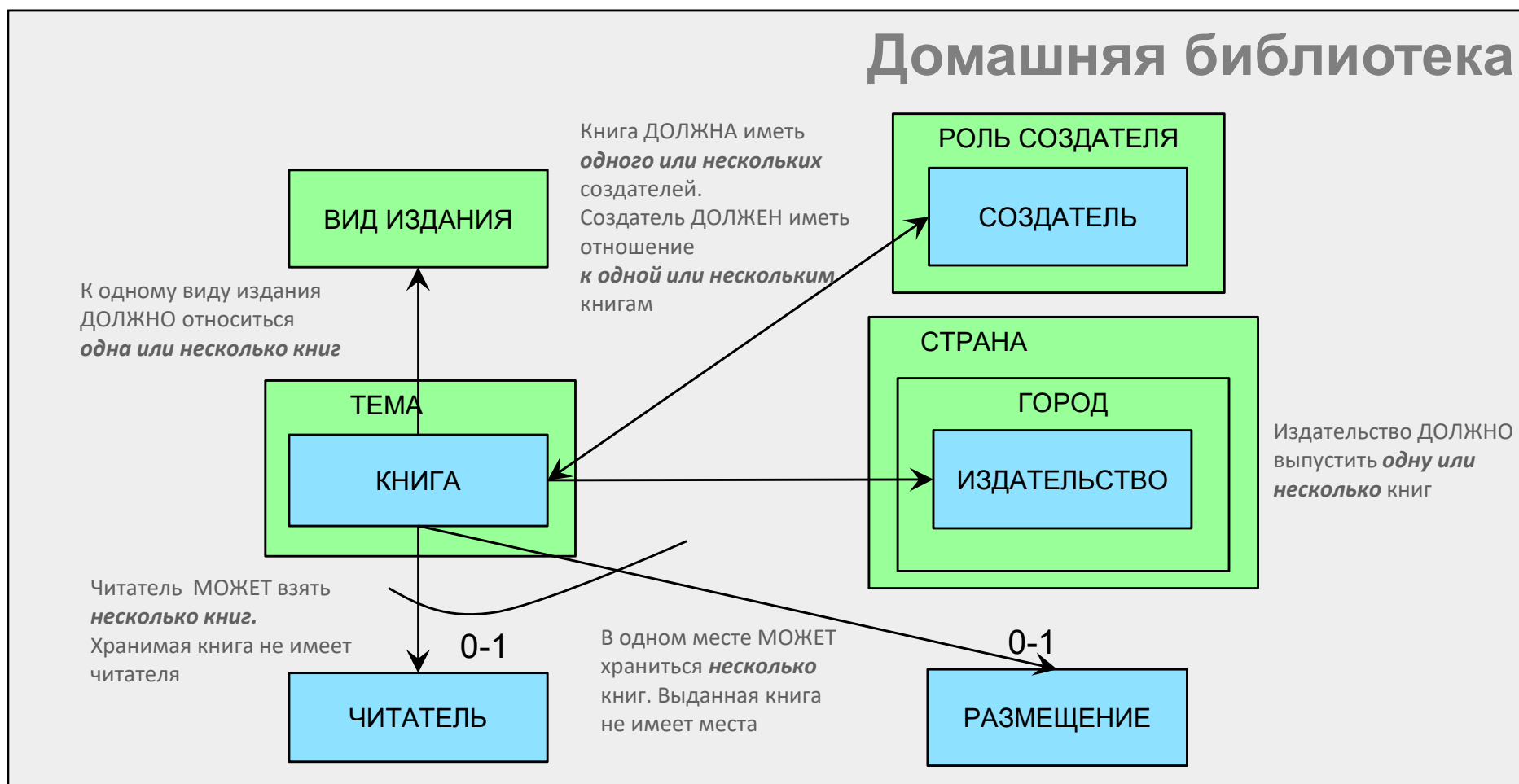


Создание КМД «Домашняя библиотека». ER-диаграмма, подготовленная в уЕd



Расширенная ER-диаграмма КМД «Домашняя библиотека»

КМД в виде супертипов и подтипов



Создание КМД «Домашняя библиотека».

Табличное описание сущностей и связей

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			

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	

Создание КМД «Домашняя библиотека».

Таксономия данных

- ⊕ 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

МГТУ имени Н.Э. Баумана, кафедра ИУ-5

Банк России

Департамент данных, проектов и процессов

Тел.: +7(495) 753-96-25

e-mail: viart@bmstu.ru