

Разработка базы данных для хранения  
информации, полученной в абазинских  
экспедициях

Соня Дурнева, БКЛ-151

26 декабря 2017 г.

# Глава 1

## Задание и анализ предметной области

### 1.1 Задание

Разработать БД для хранения информации, полученной от информантов в абазинских экспедициях.

### 1.2 Описание

База данных предназначена для хранения языкового материала, полученного от носителей абазинского языка в течение экспедиций. Данные выглядят как предложения на абазинском (примеры) с переводом на русский язык. Иногда на одно предложение на русском приходится несколько вариантов перевода от разных информантов. Для каждого предложения на абазинском может/должна быть также указана следующая информация:

- способ получения предложения: прямой перевод с русского либо конструирование опрашивающим с подтверждением информанта;
- информант, от которого получен пример;
- опрашивающий;
- тема исследования;
- комментарий любого характера

- отрицательный пример или положительный (признак бинарный: отрицательный пример – 0, положительный – 1);
- дата получения примера;
- место получения примера.

## 1.3 Анализ предметной области

Сущности:

- пример;
- информант;
- опрашивающий.

Атрибуты (необязательные выделены курсивом):

**пример:** перевод на русский, *способ получения*, дата получения, *тема исследования*, *комментарий*

**информант:** ФИО, пол, год рождения, место рождения, образование, род занятий, родной язык, владеет с детства, место рождения отца, родной язык отца, место рождения матери, родной язык матери, владеет ли мать абазинским, язык общения в семье в детстве, язык общения в семье сейчас, владение черкесским, часто ли говорит по-черкесски

**опрашивающий:** ФИО, академический статус, e-mail, университет, *научный руководитель*.

Связи:

**пример** – **пример** может иметь альтернативный(-ые) вариант(-ы)

**пример** – **информант** дан

**пример** – **опрашивающий** получен

**опрашивающий** – **опрашивающий** может иметь/быть академического(-их)/(-им(и))

руководителя(-ей)/(-ем/-ями)

## Глава 2

### ER-модель

В предложенной ER-модели (рис. 1) изображены все три сущности, выделенные выше, а также их связи. Линия с вилочкой на конце означает кратность связи  $N$ , обычная линия, перечеркнутая одной чертой, кратность 1. Пунктирная линия обозначает необязательность наличия связи. Таким образом, на предложенной схеме имеем две обязательные связи с кратностью  $1:N$ , и две необязательные с кратностью  $N:N$ .

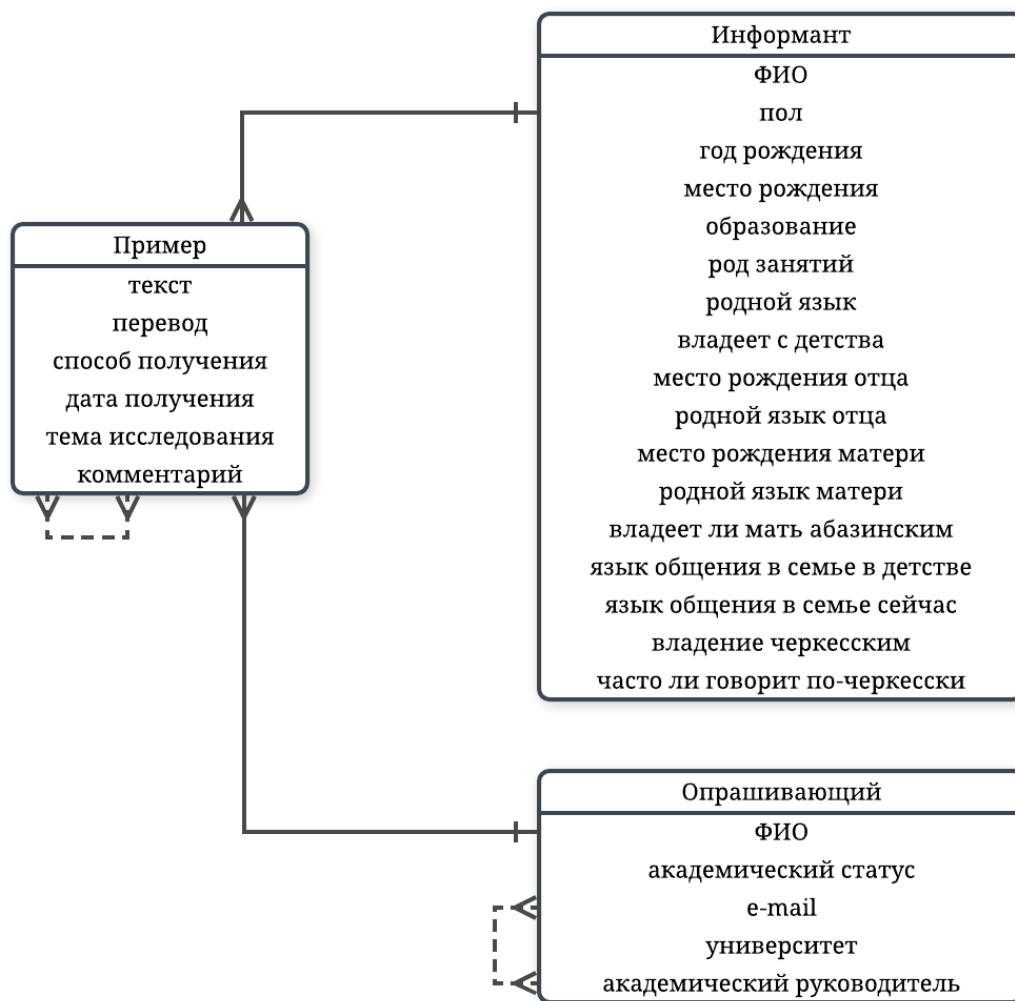


Рис. 1: ER-модель базы данных

# Глава 3

## Модель БД

### 3.1 Объединение отношений

На основании кратности связей определим количество таблиц и дадим им названия:

- "[пример – пример] альтернативный вариант кратность N:N, преобразуется в две таблицы Example и Alternatives;
- "[пример – информант] дан кратность N:1, преобразуется в две таблицы Example и Informant;
- "[пример – опрашивающий] получен кратность N:1, преобразуется в две таблицы Example и Interviewer;
- "[опрашивающий – опрашивающий] руководитель" с кратностью N:N, преобразуется в две таблицы Interviewer и Advisor.

### 3.2 Таблицы и атрибуты

Ниже приведён перечень таблиц БД с данными в скобках атрибутами. Перед названиями атрибутов из соседних таблиц стоит знак #:

Example (text, translation, obtaining\_method, datetime, subject, comment, #informant, #interviewer)

Alternatives (#example, #example)

Informant (full\_name, gender, year\_of\_birth, birthplace, education, occupation, native\_language, speaks\_since\_childhood, fathers\_birthplace, fathers\_native\_language, mothers\_birthplace, mothers\_native\_language, mother\_speaks\_Abaza, fam\_lang\_childh,

fam\_lang\_now, speaks\_Circaccian, speaks\_Circ\_freq)  
Interviewer (full\_name, academic\_stat, email, university)  
Advisor (#interviewer, #interviewer)

### 3.3 Ключевые поля

Можно было бы в качестве естественных ключей использовать первый атрибут у каждой из сущностей в ER-модели (рис.1): у информанта и опрашиваемого – ФИО, а у примера – текст. Тем не менее, было решено ввести для каждой из сущностей искусственный ключ – id, так как все вышеупомянутые потенциальные естественные ключи содержат слишком длинные текстовые значения.

### 3.4 Нормализация

В результате моделирования таблицы находятся в третьей нормальной форме (3НФ) или в нормальной форме Бойса-Кодда (НФБК).