**НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ**

**ЯДЕРНЫЙ УНИВЕРСИТЕТ**

**«МИФИ»**

**ѕѕѕѕѕѕѕѕѕѕѕѕѕѕѕѕѕѕѕѕѕѕѕѕѕѕѕѕ**

#### Институт интеллектуальных кибернетических систем

#### Кафедра № 22

**Курсовая работа по проектированию информационных систем**

Вариант № *1 «Словарь для перевода слов с иностранного языка»*

|  |  |
| --- | --- |
| Выполнил | Саутов Д. С. |
| Группа | М22-514 |
| Вариант | 1 |
| Преподаватель | Тихомирова Д.В. |
| Проверяющий |  |
| Оценка |  |

Оглавление

[1. Описание предметной области 3](#_Toc121506369)

[1.1. Формулировка задания 3](#_Toc121506370)

[1.2. Конкретизация предметной области 3](#_Toc121506371)

[1.3. Пользователи системы 4](#_Toc121506372)

[1.4. Сроки хранения информации 4](#_Toc121506373)

[1.5. События, изменяющие состояние Базы Данных 4](#_Toc121506374)

[1.6. Основные запросы к Базе Данных (на естественном языке) 5](#_Toc121506375)

[1.7. Требования к системе 5](#_Toc121506376)

[2. Концептуально-информационная модель предметной области 6](#_Toc121506377)

[2.1. ER-диаграмма модели (Draw.io) 6](#_Toc121506378)

[2.2. Оценка мощностных характеристик сущностей и связей 7](#_Toc121506379)

[3. Концептуальное проектирование 8](#_Toc121506380)

[3.1. Принятые проектные соглашения 8](#_Toc121506381)

[3.2. Обоснование выбора модели базы данных 8](#_Toc121506382)

[3.3. Используемые в системе кодификаторы 9](#_Toc121506383)

[3.4. Концептуальная модель Базы Данных (Draw.io) 10](#_Toc121506384)

[4. Логическое проектирование 12](#_Toc121506385)

[4.1. ER-диаграмма Базы Данных (ERWin Logical) 12](#_Toc121506386)

[4.2. Схемы отношений Базы Данных (ERWin Physical) 13](#_Toc121506387)

[4.3. Схема реляционной Базы Данных 13](#_Toc121506388)

[4.4. Схемы основных запросов на реляционной алгебре 13](#_Toc121506389)

[5. Физическое проектирования 17](#_Toc121506390)

[5.1. Обоснование выбора конкретной СУБД 17](#_Toc121506391)

[5.2. Создания БД 17](#_Toc121506392)

[5.3. Создание таблиц 17](#_Toc121506393)

[5.4. Заполнение таблиц 20](#_Toc121506394)

[5.5. Запросы в терминах SQL 22](#_Toc121506395)

[5.6. Оценка размера БД и каждого из файлов 23](#_Toc121506396)

[6. ПРИЛОЖЕНИЕ. ОТЧЕТЫ 25](#_Toc121506397)

# Описание предметной области

### Формулировка задания

Спроектировать базу данных для словаря иностранных слов. Словарь должен выдавать перевод и транскрипцию слов/выражений. Добавить возможность добавления в словарь новых слов и фраз вместе с их переводом. Все добавления в словарь должны будут проходить верификацию у лингвиста.

### Конкретизация предметной области

Словарь иностранных слов – портал, на котором люди, изучающие различные языки могут узнать перевод интересующих их слов и фраз. Все слова и фразы добавляются другими пользователями и проходят верификацию у лингвистов. Цель портала – предоставлять людям перевод, транскрипцию, а также развиваться вместе с языками, благодаря возможности добавления новых слов/фраз.

Каждое слово будет иметь хранить такую информацию, как какому языку оно принадлежит, какой частью речи является, его толкование, является ли оно устаревшим, а также будет хранить его транскрипцию и свой рейтинг.

Все языки будут иметь свое названия (к примеру английский язык будет именоваться как English)

Перевод слова с одного языка на другой будут хранится как пара (слово/фраза, слово/фраза), а также иметь свой собственный рейтинг.

Все фразы также будут также хранить язык, которому оно принадлежит, рейтинг среди пользователей, а также транскрипцию.

Основной функционал словаря:

* Предоставление различных перевод слова/фраз, отсортированных в зависимости от оценок пользователей
* Возможность оценивать один или несколько результатов перевода
* Возможность добавления слова в язык
* Возможность пользователям предлагать свои переводы фраз и слов на различные языки
* Возможность добавления различных языков для перевода
* Все предложенные варианты перевода фраз и слов встают в очередь проверки у лингвистов, в случае одобрения они сохраняются в словаре. В противном случае, предложенный вариант удаляется
* Предоставление списка самых популярных слов или фраз для определенного языка
* В случае если один из вариантов перевода получил много отрицательных отзывов, то он отправляется на проверку к лингвисту, который проверит актуальность перевода
* Возможность у лингвистов помечать фразы и слова, как устаревшие

### Пользователи системы

Пользователь, которому доступен весь функционал словаря, но имеет ограничение по невозможностью моментального добавления своего перевода фразы или слова в словарь, то есть он может только предложить перевод и ждать одобрения от лингвиста.

Лингвист, имеет доступ к списку предложенных переводов по известным ему языкам. Может одобрять предложенные переводы или же отклонять их. Могут изменять известные им языки, добавляя туда новый изученный язык или удалять их.

Администратор, может добавлять новые языки, а также давать пользователям права лингвиста.

### Сроки хранения информации

Перевод слов и фраз хранится в системе неограниченное количество времени, пока на них не придет большое количество отрицательных оценок со стороны пользователей и они не будет принято решение лингвистов, о том, что данный перевод не является правдивым и подлежит удалению.

Остальные сущности и связи хранятся неограниченное количество времени и не имеют срока годности.

### События, изменяющие состояние Базы Данных

Основные события в предметной области словаря иностранных слов, изменяющие хранимые в БД объекты:

* Добавление нового языка администратором
* Оценка перевода слова или фразы
* Добавление нового перевода слова или фразы
* Удаление перевода слова или фразы
* Указание слова как «устаревшие»
* Добавление слова в язык

### Основные запросы к Базе Данных (на естественном языке)

* выбрать 10 самых оценённых слов/фраз определенного языка
* Показать перевод фразы или слова, на выбранный пользователем язык
* Показать перевод фразы или слова, ***а также его транскрипцию***, на выбранный пользователем язык
* Выдача лингвисту предложенных переводов, в соответствии с выбранными им языками
* Выдача 10 самых оценённых слов из выбранной части речи определенного языка
* Выдача информации о словах/фразах для определенного языка (Его написание, значение, транскрипция, рейтинг)

### Требования к системе

#### Бизнес-требования

* Система должна иметь администратора
* Система должна иметь лингвиста
* Администратор может добавлять языки
* Лингвист имеет возможность добавлять слова для языков
* Слова могут иметь перевод с одного языка на другой
* Слова могут иметь поясняющую информацию о слове, такую как *является ли это слово устаревшим* или *какой части речи оно принадлежит* и его *значение/описание*

#### Требования пользователей

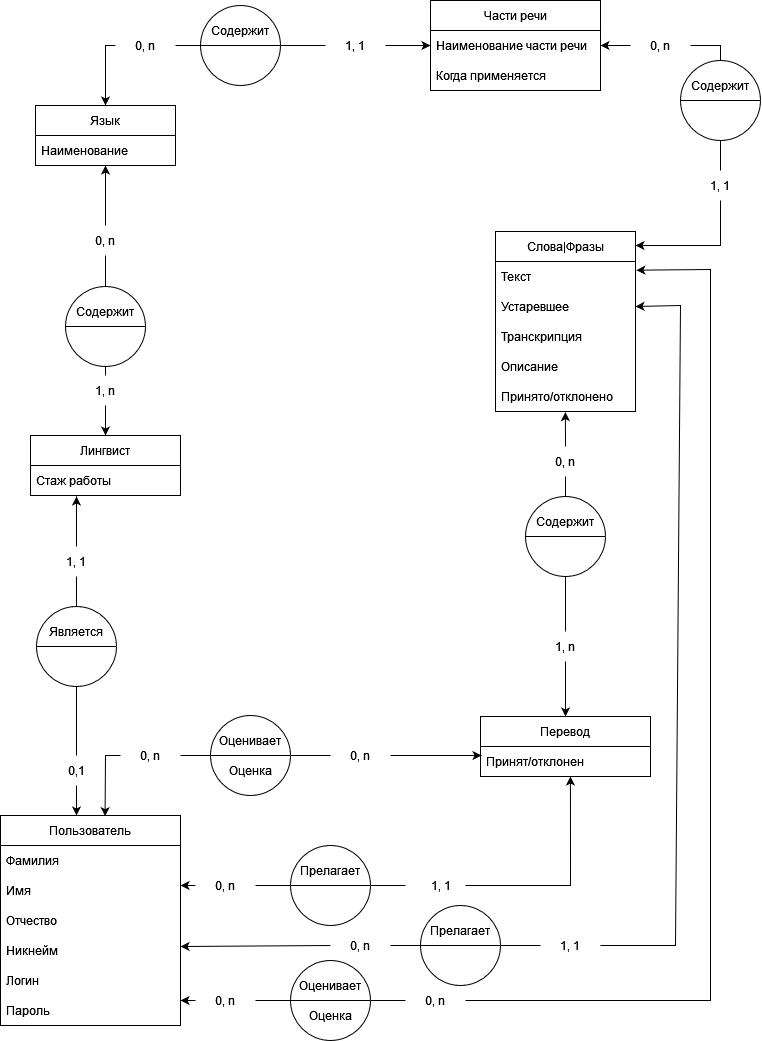
* Пользователи должны иметь возможность оценивать слова и фразы различных языков
* Пользователи должны иметь возможность оценивать переводы слов и фраз на другие языки
* Пользователи должны иметь возможность предлагать свои варианты перевода
* В случае если слово имеет низкий рейтинг, то оно идет на проверку к лингвисту, для его обновления или перевода в раздел *устаревших*
* В случае если перевод имеет низкий рейтинг, то он идет на проверку к лингвисту, который может удалить перевод, в случае его некорректности.

#### Функциональные требования

* Реализовать добавление языков
* Реализовать добавление частей речи к языкам
* Реализовать добавление слов к языку с выбранной частью речи (если она имеется)
* Реализовать добавление фраз к языку
* Каждое слово/фраза должно иметь свой рейтинг, который может выставляться пользователем
* Каждое слово/фраза должно иметь транскрипцию
* Каждое слово/фраза должно иметь свое описание
* Пользователи должны иметь возможность предлагать слова/фразы к определенным языкам
* Пользователи должны иметь возможность предлагать переводы к словам/фразам
* Пользователи должны иметь возможность оценивать переводы слов/фраз
* В системе должны работать лингвисты, которые будут принимать или отклонять предложенные слова/фразы пользователями
* В системе должны работать лингвисты, которые будут принимать или отклонять предложенные переводы слова/фразы пользователями
* Реализовать возможность проверки и изменения лингвистами слов/фраз с низкой оценкой

# Концептуально-информационная модель предметной области

### ER-диаграмма модели (Draw.io)



### Оценка мощностных характеристик сущностей и связей

Приводится таблица следующего вида.

|  |  |  |  |
| --- | --- | --- | --- |
| Сущность/связь | Мощность | | |
| минимальная | средняя | максимальная |
| Язык | 2 | 28 | 150 |
| Часть речи | 2 | 140 | 7500 |
| Слово|Фраза | 2 | 56000 | 300000 |
| Лингвист | 1 | 28 | 150 |
| Перевод | 1 | 28000 | 150000 |
| Пользователь | 0 | 100 | 15000 |
| Содержит | 2 | 56000 | 300000 |
| Предлагает | 2 | 28000 | 150000 |
| Подтверждает | 2 | 28000 | 150000 |
| Изменяет | 2 | 28000 | 150000 |
| Добавляет | 2 | 28 | 150 |
| Оценивает | 0 | 2000 | 300000 |
| Является | 1 | 28 | 150 |

В среднем в мире распространено около 28 языков. Всего в мире насчитывают около 142 различных языковых семейств. В каждом языке в среднем насчитывается 4-6 частей речи, если они там вообще присутствуют. В среднем каждый язык содержит под 100000 слов, которые не считаются устаревшими. Для среднего уровня владения языком нужно знать в среднем 2000-5000 слов. Так как будет хранится информация о нескольких языках, то это количество кратно больше. Для работоспособности системы необходим как минимум 1 лингвист. В среднем достаточно 28 лингвистов, для наполнения и обслуживания 28 языков. Так как перевод содержит фразу на 1 языке и ее перевод на другой язык, то дублируем все поля из «Слово|Фраза», только деленное на 2. Минимальное количество пользователей системы это 0. В среднем считается, что будет по 50-100 пользователей на язык. Будем считать, что пользователь знает 2 языка и может их оценивать, при этом он оценит только от 0 до 1% слов из двух языков и переводов.

# Концептуальное проектирование

### Принятые проектные соглашения

При переходе от КМПО к КМБД у всех сущностей, появляется новое свойство – **идентификатор**.

Новые сущности при переходе не добавляются.

### Обоснование выбора модели базы данных

Для разрабатываемой модели предлагается использовать реляционную модель базы данных. Это обуславливается наличием большого количества связей между сущностями в разрабатываемой базе данных.

### Используемые в системе кодификаторы

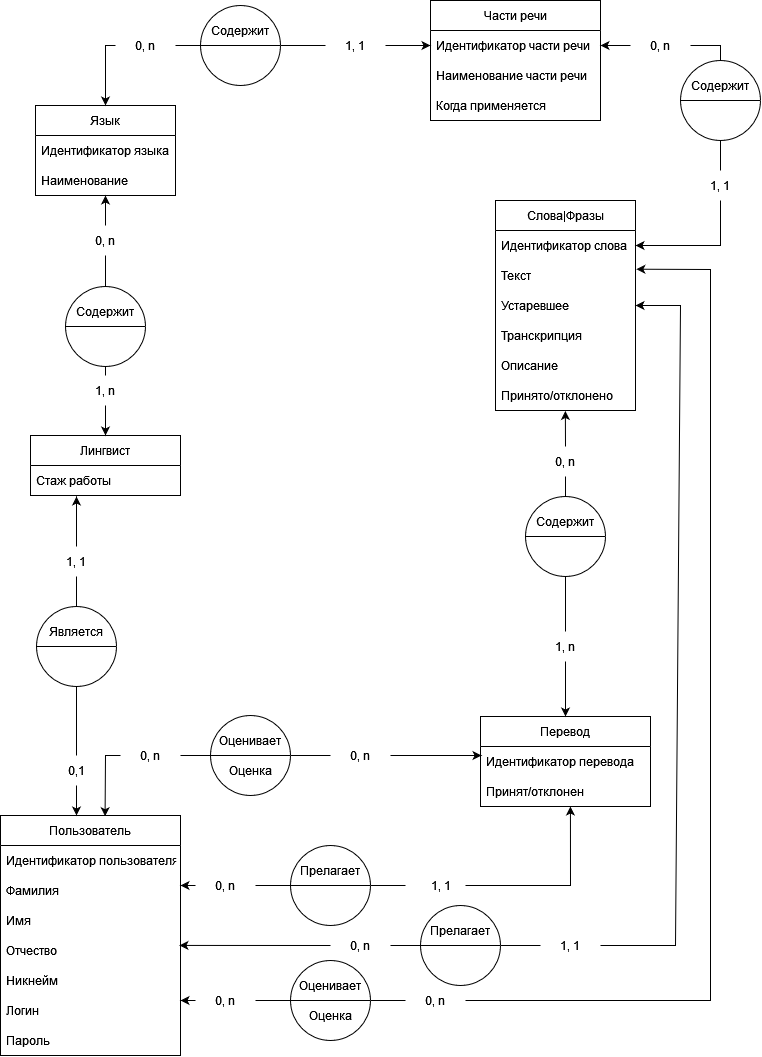
Добавим кодификатор – Устаревшее слово, которое будет определять, является слово или словосочетание устаревшим или нет. Данное поле содержит 2 значения: устаревшее, используется в современном языке.

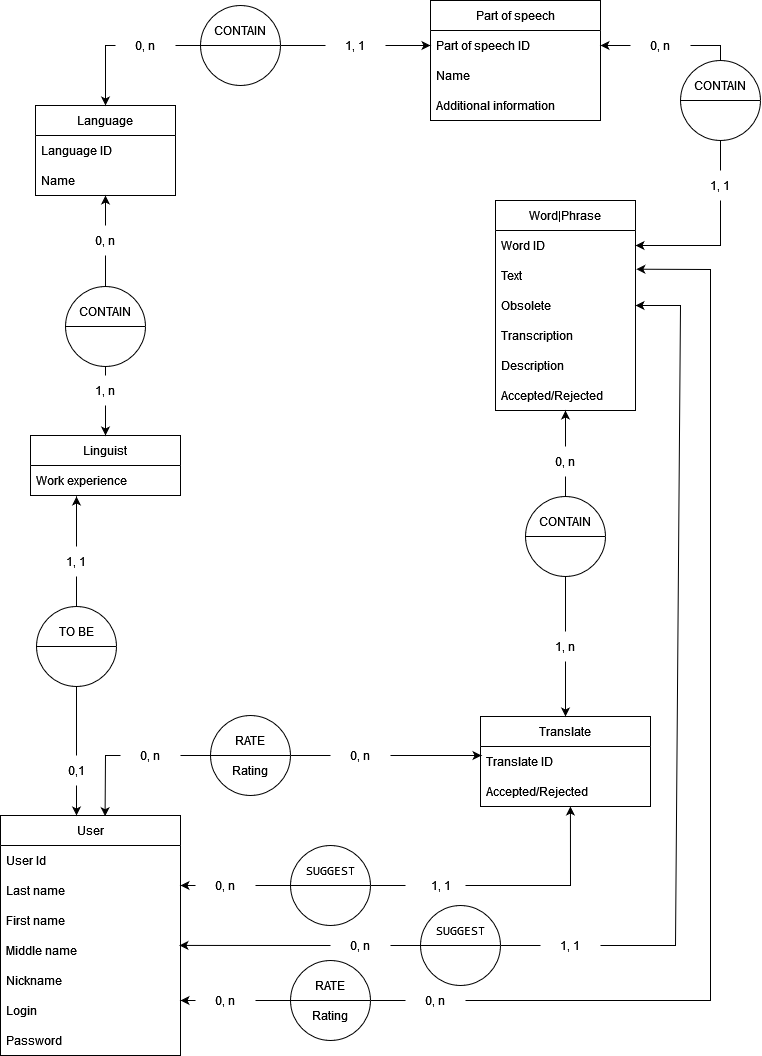
|  |  |  |
| --- | --- | --- |
| **Устаревшее слово** | | **ПО** |
| Код | Значение | |
| 0 | Современное слово | |
| 1 | Устаревшее слово | |

Добавим кодификатор – Принятый перевод, которое будет определять, прошло ли слово/фраза проверку лингвиста на добавление. Данное поле содержит 2 значения: Принято, отклонено.

|  |  |  |
| --- | --- | --- |
| **Принятый перевод** | | **ПО** |
| Код | Значение | |
| 0 | Отклонено | |
| 1 | Принято | |

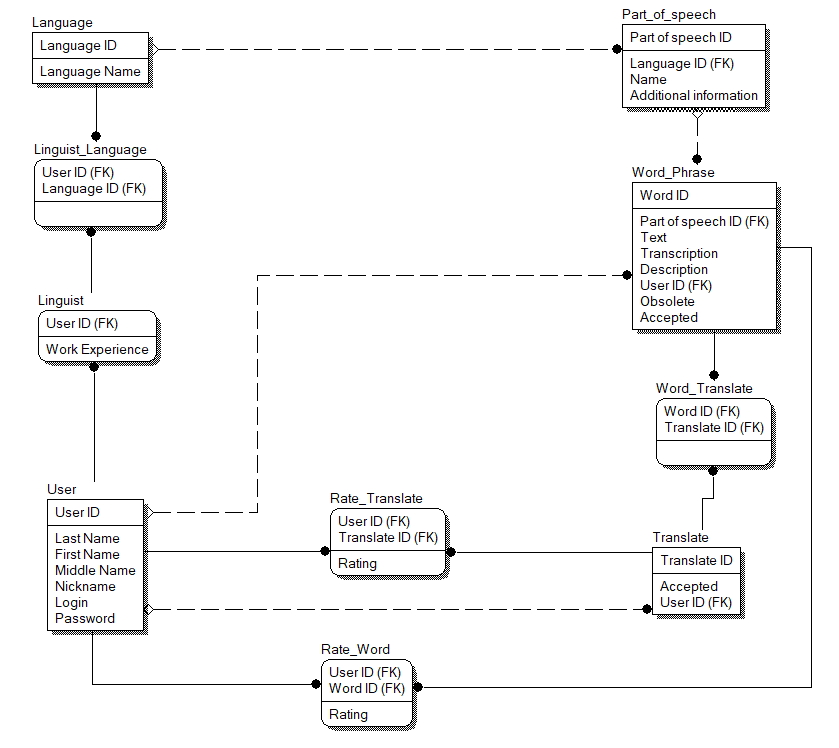
### Концептуальная модель Базы Данных (Draw.io)



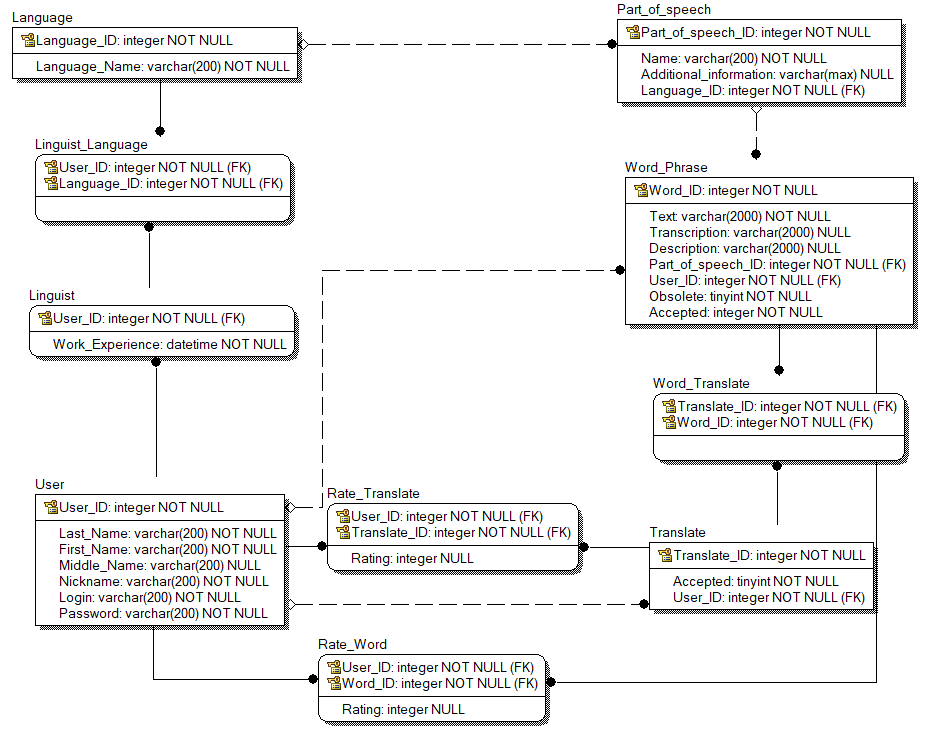


# Логическое проектирование

### ER-диаграмма Базы Данных (ERWin Logical)



### Схемы отношений Базы Данных (ERWin Physical)



### Схема реляционной Базы Данных

|  |  |
| --- | --- |
| R1 | Language(#Language\_ID, Language\_Name) |
| R2 | Part\_of\_speech(#Part\_of\_speech\_ID, Name, Additional\_information, Language\_ID) |
| R3 | Word\_Phrase(#Word\_ID, Text, Transcription, Description, Part\_of\_speech,\_ID, User\_ID, Obsolete, Accepted) |
| R4 | Word\_Translate(#Translate\_ID, #Word\_ID) |
| R5 | Translate(#Translate\_ID, Accepted, User\_ID) |
| R6 | User(#User\_ID, Last\_Name, First\_Name, Middle\_Name, Nickname, Login, Password) |
| R7 | Linguist(#User\_ID, Work\_Experience) |
| R8 | Rate\_Translate(#User\_ID, #Translate\_ID, Rating) |
| R9 | Rate\_Word(#User\_ID, #Word\_ID, Rating) |
| R10 | Linguist\_Language(#User\_ID, #Language\_ID) |

### Схемы основных запросов на реляционной алгебре

1. выбрать 10 самых оценённых слов/фраз определенного языка

* Дано: **Language\_ID1**

Считываем части речи выбранного языка

* **R1**(Part\_of\_speech\_ID) = Part\_of\_speech(Part\_of\_speech\_ID, Name, Additional\_information, Language\_ID) **[ Language\_ID1 = Part\_of\_speech.Language\_ID]** **[Part\_of\_speech\_ID]**

Считываем слова выбранного языка (слова должны быть приняты)

* **R2**(Word\_ID, Text) = **R1**(Part\_of\_speech\_ID) **[R1. Part\_of\_speech\_ID = Word\_Phrase. Part\_of\_speech\_ID AND Accepted = 1]** Word\_Phrase(Word\_ID, Text, Transcription, Description, Part\_of\_speech,\_ID, User\_ID, Obsolete, Accepted) **[Word\_ID, Text]**

Считываем рейтинг слов определённого языка

* **R3(**Word\_ID, Text, Rating**) = R2**(Word\_ID, Text) **[R2.Word\_ID = Rate\_Word.Word\_ID]** Rate\_Word(User\_ID, Word\_ID, Rating) **[R2.Word\_ID, Text, Rating]**
* Посчитать среднее значение поля Rating (Значения NULL пропускать)
* Отсортировать в порядке убывания и выбрать первые 10

1. Показать перевод фразы или слова, на выбранный пользователем язык

* Дано: Word\_ID1, Language\_ID1

Считываем **Translate\_ID** для того, чтобы найти все слова на которое это слово переводится

* **R1**(Translate\_ID) = **[Word\_ID = Word\_ID1]** Word\_Translate(Translate\_ID, Word\_ID) **[Translate\_ID]**

Проверяем что перевод принят

* **R2**(Translate\_ID) = **R1**(Translate\_ID) **[R1.Translate\_ID = Translate.Translate\_ID AND Translate.Accepted = 1]** Translate(Translate\_ID, Accepted, User\_ID) **[Translate\_ID]**

Считываем идентификаторы слов, на которое переводится данное слово

* **R3(**Translate\_ID, Word\_ID**) = R2(**Translate\_ID**) [R2.Translate\_ID = Word\_Translate.Translate\_ID AND NOT Word\_Translate.Word\_ID = Word\_ID1]** Word\_Translate(Translate\_ID, Word\_ID) **[R2.Translate\_ID, Word\_Translate.Word\_ID]**

Считываем часть речи слова

* **R4(**Translate\_ID, Word\_ID, Text, Part\_of\_speech**) = R3(**Translate\_ID, Word\_ID**) [R3.Word\_ID = Word\_Phrase .Word\_ID]** Word\_Phrase(Word\_ID, Text, Transcription, Description, Part\_of\_speech,\_ID, User\_ID, Obsolete, Accepted) **[R3.Translate\_ID, R3.Word\_ID, Text, Part\_of\_speech\_ID]**

Считываем язык, по части речи и выбираем только слова нужного языка

* **R5(**Translate\_ID, Word\_ID, Text**) = R4(**Translate\_ID, Word\_ID, Text, Part\_of\_speech**) [R4.Part\_of\_speech\_ID = Part\_of\_speech.Part\_of\_speech\_ID AND Language\_ID = Language\_ID1]** Part\_of\_speech(Part\_of\_speech\_ID, Name, Additional\_information, Language\_ID) [**R4.Translate\_ID, R4.Word\_ID, R4.Text**]

1. Показать перевод фразы или слова, ***а также его транскрипцию***, на выбранный пользователем язык

* Дано: Word\_ID1, Language\_ID1

Считываем Translate\_ID для того, чтобы найти все слова на которое это слово переводится

* **R1**(Translate\_ID) = **[Word\_ID = Word\_ID1]** Word\_Translate(Translate\_ID, Word\_ID) **[Translate\_ID]**

Проверяем что перевод принят

* **R2**(Translate\_ID) = **R1**(Translate\_ID) **[R1.Translate\_ID = Translate.Translate\_ID AND Translate.Accepted = 1]** Translate(Translate\_ID, Accepted, User\_ID) **[Translate\_ID]**

Считываем идентификаторы слов, на которое переводится данное слово

* **R3(**Translate\_ID, Word\_ID**) = R2(**Translate\_ID**) [R2.Translate\_ID = Word\_Translate.Translate\_ID AND NOT Word\_Translate.Word\_ID = Word\_ID1]** Word\_Translate(Translate\_ID, Word\_ID) **[R2.Translate\_ID, Word\_Translate.Word\_ID]**

Считываем часть речи слова

* **R4(**Translate\_ID, Word\_ID, Text, Transcription, Part\_of\_speech**) = R3(**Translate\_ID, Word\_ID**) [R3.Word\_ID = Word\_Phrase .Word\_ID]** Word\_Phrase(Word\_ID, Text, Transcription, Description, Part\_of\_speech,\_ID, User\_ID, Obsolete, Accepted) **[R3.Translate\_ID, R3.Word\_ID, Text,** **Transcription, Part\_of\_speech\_ID]**

Считываем язык, по части речи и выбираем только слова нужного языка

* **R5(**Translate\_ID, Word\_ID, Text, Transcription**) = R4(**Translate\_ID, Word\_ID, Text, Transcription, Part\_of\_speech**) [R4.Part\_of\_speech\_ID = Part\_of\_speech.Part\_of\_speech\_ID AND Language\_ID = Language\_ID1]** Part\_of\_speech(Part\_of\_speech\_ID, Name, Additional\_information, Language\_ID) [**R4.Translate\_ID, R4.Word\_ID, R4.Text**]

Считываем рейтинг переводов данного слова для выбранного языка

* **R6(**Translate\_ID, Word\_ID, Text, Transcription, Rating**) = R5(**Translate\_ID, Word\_ID, Text, Transcription**) [R5.Translate\_ID = Rate\_Translate.Translate\_ID]** Rate\_Translate(User\_ID, Translate\_ID, Rating) **[R5.Translate\_ID, R5.Word\_ID, R5.Text, R5.Transcription, Rate\_Translate.Rating]**
* Посчитать среднее значение поля Rating (Значения NULL пропускать)
* Отсортировать в порядке убывания

1. Выдача лингвисту предложенных переводов, в соответствии с выбранными им языками

* Дано User\_ID1

Считать все непринятые слова

* **R1**(Translate\_ID) = **[Accepted = 0]** Translate(Translate\_ID, Accepted, User\_ID) **[Translate\_ID]**

Перейти к слово-перевод

* **R2(**Translate\_ID, Word\_ID**) = R1(**Translate\_ID**) [R1.**Translate\_ID = **Word\_Translate.Translate\_ID]** Word\_Translate(Translate\_ID, Word\_ID) **[Translate\_ID, Word\_ID]**

Перейти к слову (+ вся инфа от слова)

* **R3(**Translate\_ID, Word\_ID, Text, Transcription, Description, Part\_of\_speech,\_ID, User\_ID, Obsolete**) = R2(**Translate\_ID, Word\_ID**) [R2.Word\_ID = Word\_Phrase.Word\_ID]** Word\_Phrase(Word\_ID, Text, Transcription, Description, Part\_of\_speech,\_ID, User\_ID, Obsolete, Accepted) **[Translate\_ID, R2.Word\_ID, Text, Transcription, Description, Part\_of\_speech,ID, User\_ID, Obsolete]**

Перейти к части речи

* **R3(**Translate\_ID, Word\_ID, Text, Transcription, Description, Part\_of\_speech,\_ID, User\_ID, Obsolete, Language\_ID**) = R2(**Translate\_ID, Word\_ID, Text, Transcription, Description, Part\_of\_speech,\_ID, User\_ID, Obsolete**) [R2.Part\_of\_speech,\_ID = Part\_of\_speech.Part\_of\_speech,\_ID]** Part\_of\_speech(Part\_of\_speech\_ID, Name, Additional\_information, Language\_ID) **[Translate\_ID, Word\_ID, Text, Transcription, Description, Part\_of\_speech,\_ID, User\_ID, Obsolete, Language\_ID]**

Перейти к языкам лингвиста

* **R4(**Language\_ID**) = [User\_ID = User\_ID1]** Linguist\_Language(User\_ID, Language\_ID) **[Language\_ID]**

Отсеять слова по языкам лингвиста R4 и отсортировать по **Translate\_ID**

* **R5(**Translate\_ID, Word\_ID, Text, Transcription, Description, Part\_of\_speech,\_ID, User\_ID, Obsolete, Language\_ID**) = R3(**Translate\_ID, Word\_ID, Text, Transcription, Description, Part\_of\_speech,\_ID, User\_ID, Obsolete, Language\_ID**) [R3.Language\_ID** = **R4. Language\_ID] R4(**Language\_ID**) [Translate\_ID, Word\_ID, Text, Transcription, Description, Part\_of\_speech,\_ID, User\_ID, Obsolete, Language\_ID]**

1. Выдача 10 самых оценённых слов из выбранной части речи определенного языка

* Дано Part\_of\_speech\_ID1

Перейти к словам

* R1(Word\_ID, Text) = **[Part\_of\_speech\_ID1 = Word\_Phrase.Part\_of\_speech\_ID AND Accepted = 1]** Word\_Phrase(Word\_ID, Text, Transcription, Description, Part\_of\_speech,\_ID, User\_ID, Obsolete, Accepted) **[Word\_ID, Text]**

Перейти в рейтинг слова

* **R2(**R1.Word\_ID, Text, Rating**)** = **(**Word\_ID, Text, Rating**) = R1**(Word\_ID, Text) **[R1.Word\_ID = Rate\_Word.Word\_ID]** Rate\_Word(User\_ID, Word\_ID, Rating) **[R1.Word\_ID, Text, Rating]**
* Посчитать среднее значение поля Rating (Значения NULL пропускать)
* Отсортировать в порядке убывания и выбрать первые 10

# Физическое проектирования

### Обоснование выбора конкретной СУБД

В качестве СУБД будет использоваться MS SQL Server. Данная СУБД позволит создать и разработать необходимую для данного проекта БД. Так же использование MS SQL Server’а имеет такие преимущества, как **возможность создания табличных функций,** которые будут полезны для проекта, так как многие запросы подразумевают входные данные, которые влияют на записи, в таблице, которая требуется в результате выполнения запроса.

### Создания БД

If(db\_id(N'PIS\_DB') IS NULL)

begin

CREATE DATABASE PIS\_DB

end

### Создание таблиц

USE PIS\_DB;

CREATE TABLE UserTable

(

UserID integer IDENTITY(1,1) PRIMARY KEY NOT NULL ,

Last\_Name varchar(200) NOT NULL ,

First\_Name varchar(200) NOT NULL ,

Middle\_Name varchar(200) NULL ,

Nickname varchar(200) NOT NULL ,

LoginUser varchar(200) NOT NULL ,

PasswordUser varchar(200) NOT NULL

)

go

CREATE TABLE Language

(

Language\_ID integer IDENTITY(1,1) PRIMARY KEY NOT NULL ,

Language\_Name varchar(200) NOT NULL

)

go

CREATE TABLE Linguist

(

Work\_Experience datetime NOT NULL ,

UserID integer PRIMARY KEY NOT NULL ,

FOREIGN KEY (UserID) REFERENCES UserTable(UserID)

ON DELETE CASCADE

ON UPDATE CASCADE

)

go

CREATE TABLE Linguist\_Language

(

UserID integer NOT NULL ,

Language\_ID integer NOT NULL ,

CONSTRAINT PK\_Linguist\_Language PRIMARY KEY CLUSTERED (UserID ASC,Language\_ID ASC),

FOREIGN KEY (UserID) REFERENCES Linguist(UserID)

ON DELETE CASCADE

ON UPDATE CASCADE,

CONSTRAINT R\_15 FOREIGN KEY (Language\_ID) REFERENCES Language(Language\_ID)

ON DELETE CASCADE

ON UPDATE CASCADE

)

go

CREATE TABLE Part\_of\_speech

(

Part\_of\_speech\_ID integer IDENTITY(1,1) PRIMARY KEY NOT NULL ,

NamePOF varchar(200) NOT NULL ,

Additional\_information varchar(2000) NULL ,

Language\_ID integer NOT NULL ,

CONSTRAINT R\_1 FOREIGN KEY (Language\_ID) REFERENCES Language(Language\_ID)

ON DELETE CASCADE

ON UPDATE CASCADE

)

go

CREATE TABLE Word\_Phrase

(

Word\_ID integer IDENTITY(1,1) PRIMARY KEY NOT NULL ,

TextWord varchar(2000) NOT NULL ,

Transcription varchar(2000) NULL ,

DescriptionWord varchar(2000) NULL ,

Part\_of\_speech\_ID integer NOT NULL ,

UserID integer NOT NULL ,

Obsolete tinyint NOT NULL ,

Accepted integer NOT NULL ,

CONSTRAINT R\_2 FOREIGN KEY (Part\_of\_speech\_ID) REFERENCES Part\_of\_speech(Part\_of\_speech\_ID)

ON DELETE CASCADE

ON UPDATE CASCADE,

CONSTRAINT R\_12 FOREIGN KEY (UserID) REFERENCES UserTable(UserID)

ON DELETE NO ACTION

ON UPDATE NO ACTION

)

go

CREATE TABLE Translation

(

Translate\_ID integer IDENTITY(1,1) PRIMARY KEY NOT NULL ,

Accepted tinyint NOT NULL ,

UserID integer NOT NULL ,

CONSTRAINT R\_11 FOREIGN KEY (UserID) REFERENCES UserTable(UserID)

ON DELETE NO ACTION

ON UPDATE NO ACTION

)

go

CREATE TABLE Rate\_Translate

(

Rating integer NULL ,

UserID integer NOT NULL ,

Translate\_ID integer NOT NULL ,

CONSTRAINT PK\_Rate\_Translate PRIMARY KEY CLUSTERED (UserID ASC,Translate\_ID ASC),

CONSTRAINT R\_7 FOREIGN KEY (UserID) REFERENCES UserTable(UserID)

ON DELETE CASCADE

ON UPDATE CASCADE,

CONSTRAINT R\_8 FOREIGN KEY (Translate\_ID) REFERENCES Translation(Translate\_ID)

ON DELETE CASCADE

ON UPDATE CASCADE

)

go

CREATE TABLE Rate\_Word

(

Rating integer NOT NULL ,

UserID integer NOT NULL ,

Word\_ID integer NOT NULL ,

CONSTRAINT PK\_Rate\_Word PRIMARY KEY CLUSTERED (UserID ASC,Word\_ID ASC),

CONSTRAINT R\_9 FOREIGN KEY (UserID) REFERENCES UserTable(UserID)

ON DELETE CASCADE

ON UPDATE CASCADE,

CONSTRAINT R\_10 FOREIGN KEY (Word\_ID) REFERENCES Word\_Phrase(Word\_ID)

ON DELETE CASCADE

ON UPDATE CASCADE

)

go

CREATE TABLE Word\_Translate

(

Word\_ID integer NOT NULL ,

Translate\_ID integer NOT NULL ,

CONSTRAINT PK\_Word\_Translate PRIMARY KEY CLUSTERED (Translate\_ID ASC,Word\_ID ASC),

CONSTRAINT R\_28 FOREIGN KEY (Word\_ID) REFERENCES Word\_Phrase(Word\_ID)

ON DELETE CASCADE

ON UPDATE CASCADE,

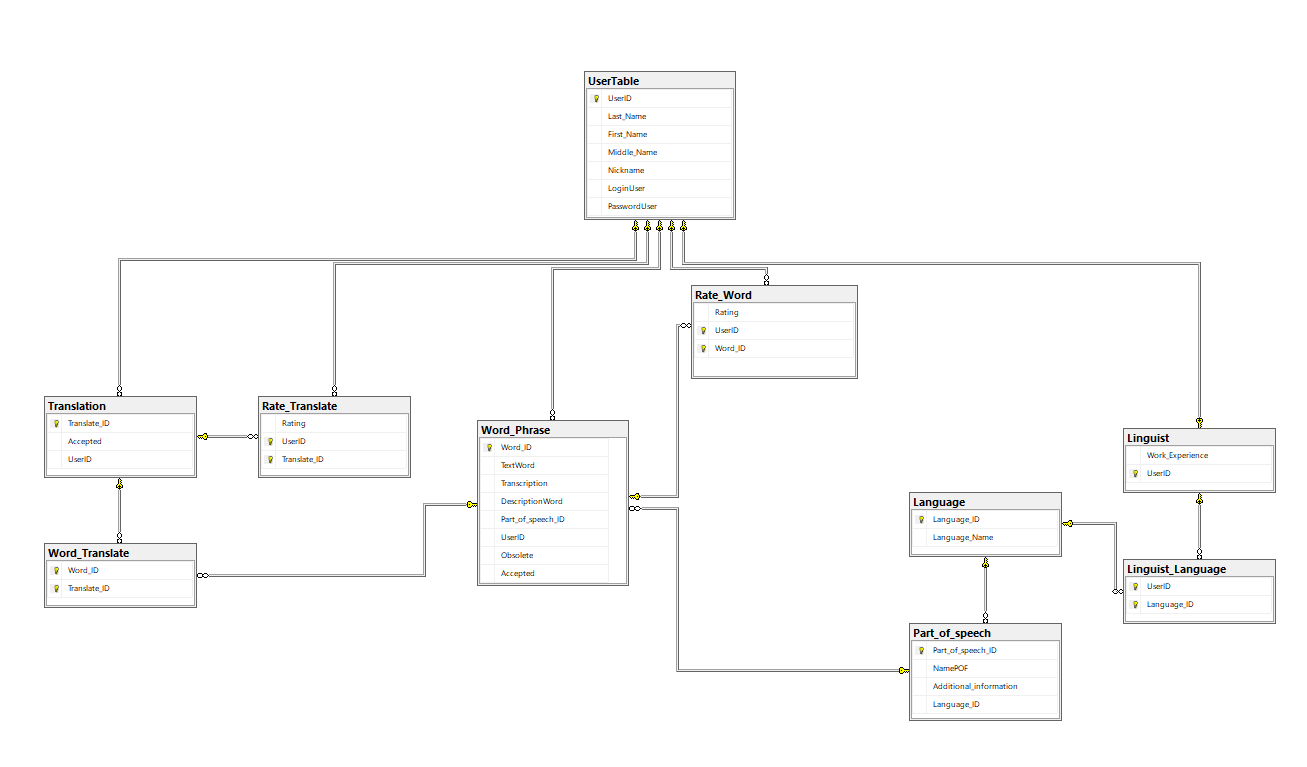
CONSTRAINT R\_30 FOREIGN KEY (Translate\_ID) REFERENCES Translation(Translate\_ID)

ON DELETE CASCADE

ON UPDATE CASCADE

)

go



### Заполнение таблиц. Запросы ETL

В качестве источников данных использовались слова из открытых источников, а именно:

Для русских слов использовался сайт:

<https://app.togetherdb.com/db/fwoedz5fvtwvq03v/russian3/words>

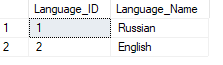
Для английских слов использовался сайт:

<https://www.bragitoff.com/2016/03/english-dictionary-in-csv-format/>

Открытых баз данных пользователей (с необходимой мне информацией) в интернете мною не было найдено

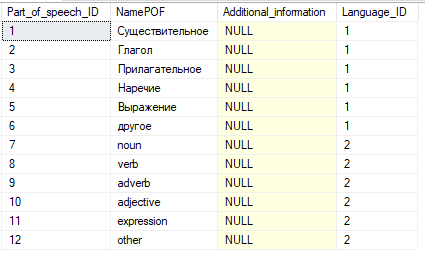
**Таблица Language**

Так как было найдено всего 2 языка, то они и будут созданы:



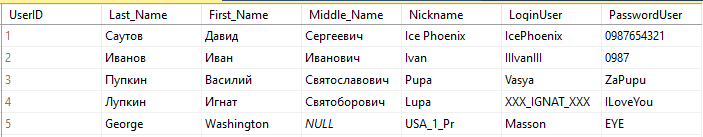
**Таблица Part\_of\_speech**

Для этих языков были созданы свои части речи



**Таблица UserTable**

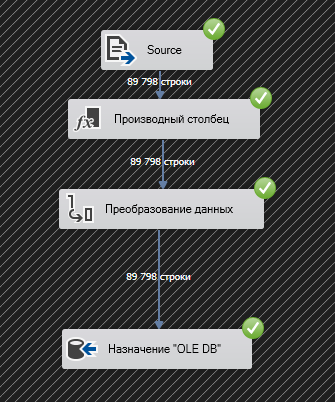
Было создано несколько тестовых пользователей для проверки работоспособности БД:

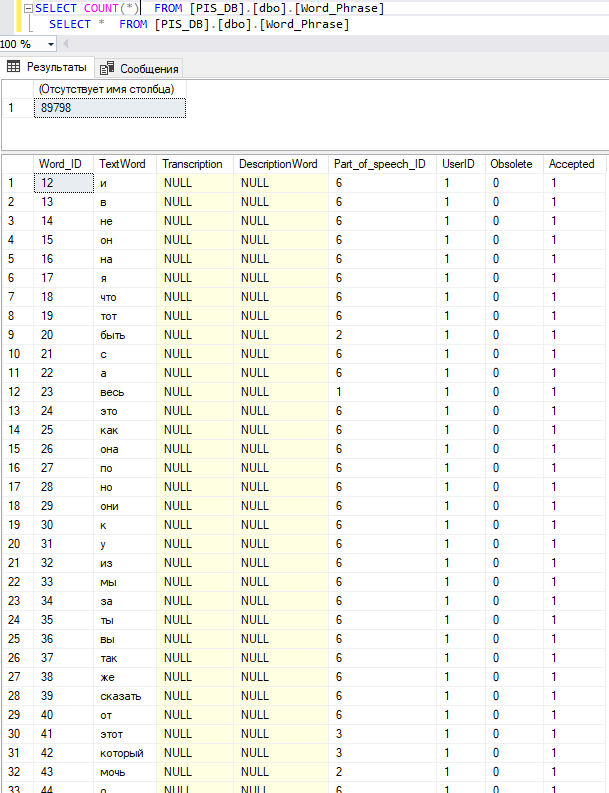


**Таблица Word\_Phrase**

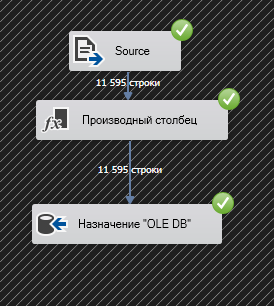
Для заполнения данной таблицы использовались данные, скачанные с указанных выше сайтов. Для этого использовалось расширение для VS, а также SISS. Эти инструменты позволили создать ETL-пакет, для обработки данных и ее загрузки обработанных данных в БД.

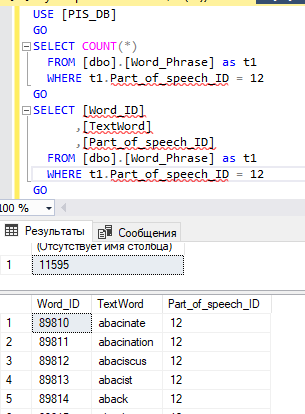
Русские слова:





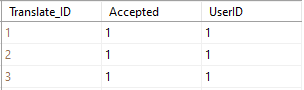
Английские слова на A:





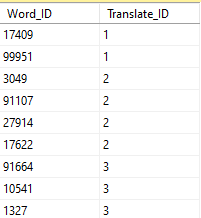
**Таблица Translation**

Так как слова скачивались с разных источников, то связать слова разного языка с правильным перевод будет весьма проблематично. Поэтому она будет напрямую заполнена минимальными данными.



**Таблица Word\_Translate**

Данная таблица будет напрямую заполнена минимальными данными.



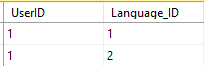
**Таблица Linguist**

Данная таблица будет напрямую заполнена минимальными данными.



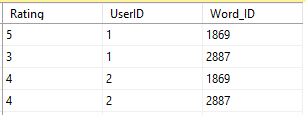
**Таблица Linguist\_Language**

Данная таблица будет напрямую заполнена минимальными данными.



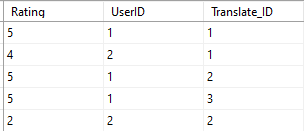
**Таблица Rate\_Word**

Данная таблица будет напрямую заполнена минимальными данными.



**Таблица Rate\_Translate**

Данная таблица будет напрямую заполнена минимальными данными.



### Запросы в терминах SQL

1. Выбрать 10 самых оценённых слов/фраз определенного языка

**Функция**

CREATE FUNCTION [dbo].[word\_rating] (

@wordID INT

)

RETURNS float AS

BEGIN

DECLARE

@rating1 float;

select @rating1 = AVG(cast(w1.Rating as float)) from Rate\_Word as w1 where w1.Word\_ID = @wordID;

RETURN @rating1;

END

**Запрос**

select top 10 wp.TextWord, dbo.word\_rating(wp.Word\_ID) as 'Рейтинг' from dbo.Word\_Phrase as wp

inner join Part\_of\_speech as pos on pos.Part\_of\_speech\_ID = wp.Part\_of\_speech\_ID

where wp.Accepted = 1 and pos.Language\_ID = @lang1 order by dbo.word\_rating(wp.Word\_ID) DESC



1. Показать перевод фразы или слова, на выбранный пользователем язык

**Запрос**

**для слова «accuracy» и на язык «Russian»**

select @word\_translateID=wt.Translate\_ID from Word\_Translate as wt where wt.Word\_ID = @word1;

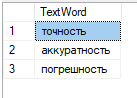
select wp.TextWord from Translation as t

inner join Word\_Translate as wt on t.Translate\_ID = wt.Translate\_ID

inner join Word\_Phrase as wp on wt.Word\_ID = wp.Word\_ID

inner join Part\_of\_speech as pof on wp.Part\_of\_speech\_ID = pof.Part\_of\_speech\_ID

where t.Accepted = 1 and t.Translate\_ID = @word\_translateID and pof.Language\_ID = @lang1



1. Показать перевод фразы или слова, ***а также его транскрипцию***, на выбранный пользователем язык

**Запрос**

**для слова «accuracy» и на язык «Russian»**

select @word\_translateID=wt.Translate\_ID from Word\_Translate as wt where wt.Word\_ID = @word1;

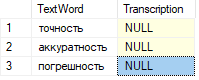
select wp.TextWord, wp.Transcription from Translation as t

inner join Word\_Translate as wt on t.Translate\_ID = wt.Translate\_ID

inner join Word\_Phrase as wp on wt.Word\_ID = wp.Word\_ID

inner join Part\_of\_speech as pof on wp.Part\_of\_speech\_ID = pof.Part\_of\_speech\_ID

where t.Accepted = 1 and t.Translate\_ID = @word\_translateID and pof.Language\_ID = @lang1



1. Выдача лингвисту предложенных переводов, в соответствии с выбранными им языками

**Запрос**

**Для лингвиста №1**

select wt.Translate\_ID, t.UserID, wp.TextWord from Word\_Translate as wt

left join Translation as t on wt.Translate\_ID = t.Translate\_ID

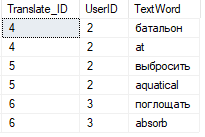
inner join Word\_Phrase as wp on wp.Word\_ID = wt.Word\_ID

inner join Part\_of\_speech as pof on wp.Part\_of\_speech\_ID = pof.Part\_of\_speech\_ID

inner join Linguist\_Language as ll on ll.Language\_ID = pof.Language\_ID

where t.Accepted = 0 and ll.UserID = @userID

order by wt.Translate\_ID



1. Выдача 10 самых оценённых слов из выбранной части речи определенного языка

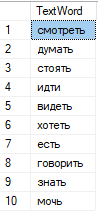
**Функция** из [раздела 5.5.1](#_Запросы_в_терминах)

**Запрос**

**Для русского языка, части речи - глагол**

select top 10 wp.TextWord from Word\_Phrase as wp

where wp.Accepted = 1 and wp.Part\_of\_speech\_ID = @pof order by dbo.word\_rating(wp.Word\_ID) DESC



1. Выдать количество слов в различных языках (По требованию преподавателя)

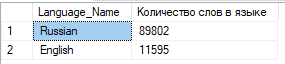
SELECT l.Language\_Name, Count(\*)

FROM Word\_Phrase as wp

inner join Part\_of\_speech as pof on pof.Part\_of\_speech\_ID=wp.Part\_of\_speech\_ID

inner join Language as l on pof.Language\_ID = l.Language\_ID

group by l.Language\_Name



### Оценка размера БД и каждого из файлов

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Отношение | Атрибут | Тип данных | Размер,  байт | Среднее количество | Объем,  байт |
| Language | Language\_ID | integer | 4 | 28 | 952 |
| Language\_Name | Varchar(200) | ~30 |
| Linguist | Work\_experience | datetime | 8 | 28 | 336 |
| UserID | integer | 4 |
| Linguist\_Language | UserID | integer | 4 | 56 | 448 |
| Language\_ID | integer | 4 |
| Part\_of\_speech | Part\_of\_speech\_ID | integer | 4 | 140 | 47320 |
| NamePOF | Varchar(200) | ~30 |
| terminal\_id | Varchar(2000) | ~300 |
| Language\_ID | integer | 4 |
| Rate\_Translate | Rating | integer | 4 | 2000 |  |
| UserID | integer | 4 |
| Translate\_ID | integer | 4 |
| Rate\_Word | Rating | integer | 4 | 2000 | 24000 |
| UserID | integer | 4 |
| Word\_ID | integer | 4 |
| Translation | Translate\_ID | integer | 4 | 28000 | 252000 |
| Accepted | tinyint | 1 |
| UserID | integer | 4 |
| UserTable | UserID | integer | 4 | 100 | 8400 |
| Last\_Name | Varchar(200) | ~10 |
| First\_Name | Varchar(200) | ~10 |
| Middle\_Name | Varchar(200) | ~12 |
| Nickname | Varchar(200) | ~16 |
| LoginUser | Varchar(200) | ~16 |
| PasswordUser | Varchar(200) | ~16 |
| Word\_Phrase | Word\_ID | integer | 4 | 56000 | 13 944 000‬ |
| TextWord | Varchar(2000) | ~15 |
| Transcription | Varchar(2000) | ~20 |
| DescriptionWord | Varchar(2000) | ~200 |
| Part\_of\_speech\_id | integer | 4 |
| UserID | integer | 4 |
| Obsolete | Tinyint | 1 |
| Accepted | Tinyint | 1 |
| Word\_Translate | Word\_ID | integer | 4 | 84000 | 672000 |
| Translate\_ID | integer | 4 |

Средний размер базы 14 973 456 байт, что ~ 14,28 МБайта (при условии, что в среднем 28 языков и в среднем каждый из языков содержит 2000 слов)

# ПРИЛОЖЕНИЕ. ОТЧЕТЫ

Ссылка на репозиторий с ETL-пакетом: <https://github.com/IcePhoenixIce/Integration-Services-Project/tree/master/Integration%20Services%20Project1>

Графики по запросам

1. Вывести 10 оцененных слов в русском языке
2. Вывести количество слов в различных языках