**Описание разработанной схемы лингвистической базы данных**

На основе описанной выше логической структуры была разработана схема лингвистической базы данных. Ее описание по компонентам предоставлено дальше.

*Морфологическая база данных*

Для морфологической базы данных, схема представлена на рисунке 1.

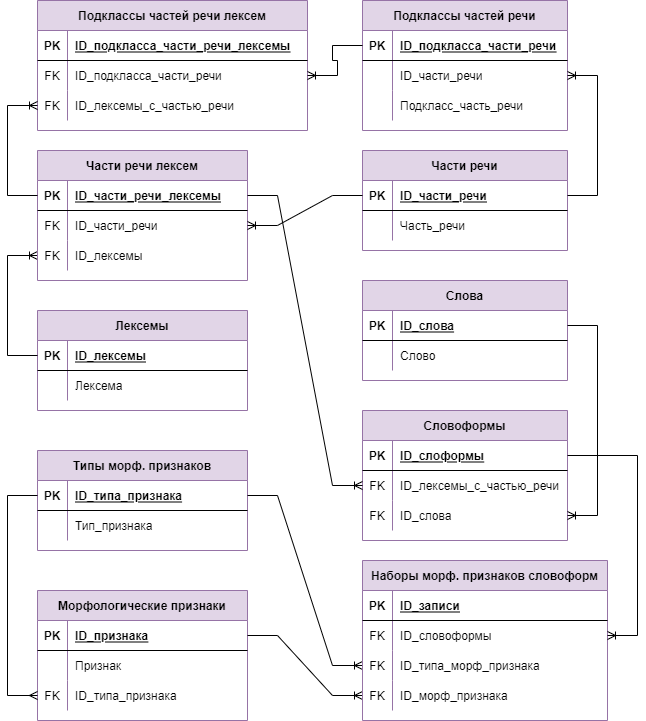


Рисунок 1 – Схема морфологической базы данных

*Описание таблиц МБД*

**Таблица «Слова»**

Содержит слова естественного языка (уникальные)

*Поля таблицы «Слова»:*

* ID\_слова – идентификатор конкретного слова (Тип данных: целое число);
* Слово – конкретное уникальное слово (Тип данных: строка переменного размера).

**Таблица «Части речи»**

Содержит список частей речи.

*Поля таблицы «Части речи»:*

* ID\_части\_речи – идентификатор конкретной части речи (Тип данных: целое число);
* Часть\_речи – название части речи (Тип данных: строка переменного размера).

**Таблица «Подклассы частей речи»**

Содержит список подклассов частей речи.

*Поля таблицы «Части речи»:*

* ID\_подкласса\_части\_речи – идентификатор конкретного подкласса части речи (Тип данных: целое число);
* Подкласс\_часть\_речи – название подкласса части речи (Тип данных: строка переменного размера);
* ID\_части\_речи – идентификатор части речи (Тип данных: целое число).

**Таблица «Лексемы»**

Содержит список лексем (уникальных).

*Поля таблицы «Лексемы»:*

* ID\_лексемы – идентификатор конкретной части речи (Тип данных: целое число);
* Лексема – конкретная уникальная лексема (Тип данных: строка переменного размера).

**Таблица «Части речи лексем»**

Таблица связи лексем с соответствующими им частями речи.

*Поля таблицы «Части речи лексем»:*

* ID\_части\_речи\_лексемы – идентификатор записи о связи лексемы с ее частью речи (Тип данных: целое число);
* ID\_части\_речи – идентификатор конкретной части речи (Тип данных: целое число);
* ID\_лексемы – идентификатор конкретной лексемы (Тип данных: целое число).

**Таблица «Подклассы частей речи лексем»**

Таблица связи лексем с соответствующими им подклассами частямей речи.

*Поля таблицы «Части речи лексем»:*

* ID\_подкласса\_части\_речи\_лексемы – идентификатор записи о связи лексемы с ее поклассом частью речи (Тип данных: целое число);
* ID\_подкласса\_части\_речи – идентификатор подкласса части речи (Тип данных: целое число);
* ID\_лексемы – идентификатор строки таблицы «Части речи лексемы» (Тип данных: целое число).

**Таблица «Словоформы»**

Таблица связи слов с соответствующими им лексемами.

*Поля таблицы «Словоформы»:*

* ID\_словоформы – идентификатор конкретной словоформы (Тип данных: целое число);
* ID\_лексемы\_с\_частью\_речи – идентификатор записи таблицы «Части речи лексем», определяющей лексему с частью речи (Тип данных: целое число);
* ID\_слова – идентификатор слова (Тип данных: целое число).

**Таблица «Типы морфологических признаков»**

Содержит список типов морфологических признаков (падеж, склонение и т.п.).

*Поля таблицы «Типы морфологических признаков»:*

* ID\_типа\_признака – идентификатор конкретной словоформы (Тип данных: целое число);
* Тип\_признака – название типа признака (Тип данных: строка переменного размера).

**Таблица «Морфологические признаки»**

Содержит список возможных значений морфологических признаков (родительный, 1 склонение и т.п.).

*Поля таблицы «Морфологические признаки»:*

* ID\_признака – идентификатор значения признака (Тип данных: целое число);
* Признак – значение признака (Тип данных: строка переменной длины);
* ID\_типа\_признака – идентификатор типа признака (Тип данных: целое число).

**Таблица «Наборы морфологических признаков словоформ»**

Содержит набор морфологических признаков словоформ.

*Поля таблицы «Наборы морфологических признаков словоформ»:*

* ID\_записи – идентификатор записи (Тип данных: целое число);
* ID\_словоформы – идентификатор словоформы (Тип данных: целое число);
* ID\_типа\_морф\_признака – идентификатор типа морфологического признака (Тип данных: целое число);
* ID\_морф\_признака – идентификатор значения морфологического признака (Тип данных: целое число).

*Лексико-семантический словарь*

Схема базы данных для лексико-семантического словаря представлена на рисунке 2.



Рисунок 2 – Схема лексико-семантического словаря

*Описание таблиц лексико-семантического словаря*

Таблица «Подклассы частей речи лексем» является частью МБД и описана в предыдущем разделе.

**Таблица «Типы семантических значений»**

Содержит набор возможных типов семантических значений.

*Поля таблицы «Типы семантических значений»:*

* ID\_типа\_значения – идентификатор типа семантического значения (Тип данных: целое число);
* Тип\_значения – название типа семантического значения (Тип данных: строка переменной длины).

**Таблица «Семантические значения»**

Содержит набор возможных семантических значений разных типов (основное значение, дополнительное значение, тематическая роль и т.п.).

*Поля таблицы «Семантические значения»:*

* ID\_значения – идентификатор семантического значения (Тип данных: целое число);
* ID\_типа\_значения – идентификатор типа семантического значения (Тип данных: целое число).

**Таблица «Основные семантические значения лексем»**

Содержит связь лексем с их основными семантическими значениями.

*Поля таблицы «Основные семантические значения лексем»:*

* ID\_основного\_значения\_лексемы – идентификатор значения конкретной лексемы (Тип данных: целое число);
* ID\_подкласса\_части\_речи\_лексемы – идентификатор конкретной лексемы с определенным подклассом части речи (Тип данных: целое число);
* ID\_основного\_значения – идентификатор основного семантического значения (Тип данных: целое число);
* Комментарий – необязательное поле, в котором может содержаться пример использования данной лексемы с заданным семантическим значением.

**Таблица «Дополнительные семантические значения лексем»**

Содержит связь лексем с их дополнительными семантическими значениями.

*Поля таблицы «Дополнительные семантические значения лексем»:*

* ID\_доп\_значения\_лексемы – идентификатор дополнительного значения конкретной лексемы (Тип данных: целое число);
* ID\_основного\_значения\_лексемы – идентификатор лексемы с определенным основным семантическим значением (Тип данных: целое число);
* ID\_дополнительного\_значения – идентификатор дополнительного семантического значения (Тип данных: целое число)*;*
* Комментарий – необязательное поле, в котором может содержаться пример использования данной лексемы с заданным семантическим значением.

*Словарь глагольно-предложных фреймов*

Схема базы данных для словаря предложных фреймов представлена на рисунке 3.

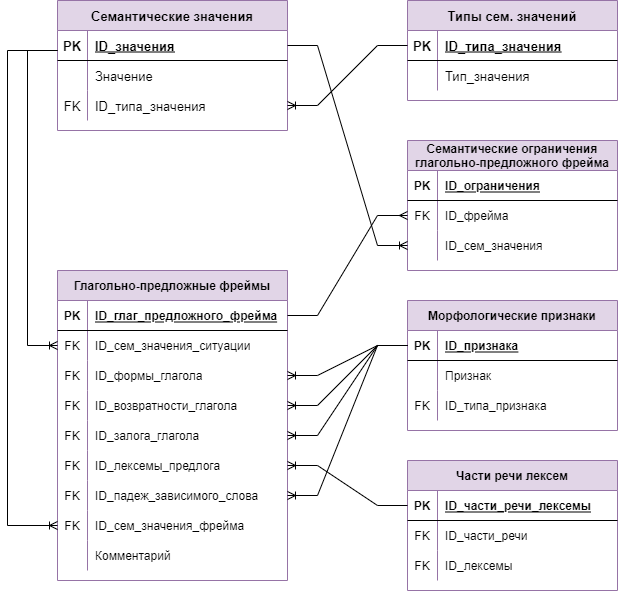


Рисунок 3 – схема словаря глагольно-предложных фреймов

*Описание таблиц словаря глагольно-предложных фреймов*

Таблицы «Части речи лексем» и «Морфологические признаки» описаны в рамках морфологической базы данных. Таблицы «Семантические значения» и «Типы семантических значений» описаны в рамках лексико-семантического словаря.

**Таблица «Глагольно-предложные фреймы»**

Содержит описания глагольно-предложных фреймов.

*Поля таблицы «Глагольно-предложные фреймы»:*

* ID\_глаг\_предложного\_фрейма – идентификатор глагольно-предложного фрейма (Тип данных: целое число);
* ID\_сем\_значения\_ситуации – идентификатор семантического значения ситуации, описываемой глагольной формой (Тип данных: целое число);
* ID\_формы\_глагола – идентификатор признака, определяющего форму глагола (Тип данных: целое число);
* ID\_возвратности\_глагола – идентификатор признака, определяющего возвратность глагола (Тип данных: целое число);
* ID\_залога\_глагола – идентификатор признака, определяющего залог глагола (Тип данных: целое число);
* ID\_лексемы предлога – идентификатор лексемы предлога, связанного с глагольной формой (Тип данных: целое число);
* ID\_падеж\_зависимого\_слова – идентификатор признака, определяющего падеж слова (Тип данных: целое число);
* ID\_сем\_значения\_фрейма – идентификатор семантического значения фрейма (Тип данных: целое число);
* Комментарий – необязательное поле, в котором может содержаться пример использования данного фрейма.

*Словарь предложных фреймов*

Схема базы данных для словаря предложных фреймов представлена на рисунке 4.

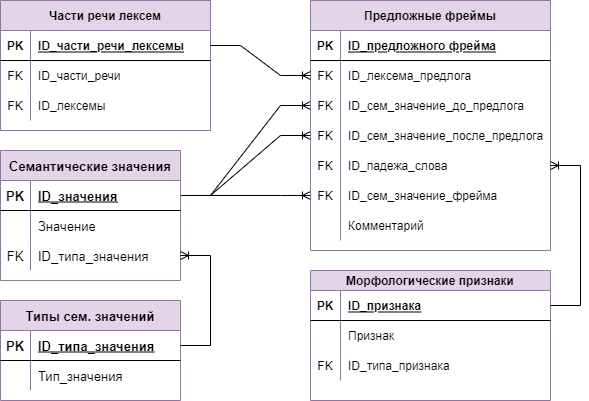


Рисунок 4 – схема словаря предложных фреймов

*Описание таблиц словаря предложных фреймов*

Таблицы «Части речи лексем» и «Морфологические признаки» описаны в рамках морфологической базы данных. Таблицы «Семантические значения» и «Типы семантических значений» описаны в рамках лексико-семантического словаря.

**Таблица «Предложные фреймы»**

Содержит описания предложных фреймов.

*Поля таблицы «Предложные фреймы»:*

* ID\_предложного\_фрейма – идентификатор предложного фрейма (Тип данных: целое число);
* ID\_лексема\_предлога – идентификатор лексемы предлога (Тип данных: целое число);
* ID\_сем\_значения\_до\_предлога – идентификатор дополнительного значения, которое может принимать слово расположенное перед предлогом (Тип данных: целое число);
* ID\_сем\_значения\_после\_предлога – идентификатор дополнительного значения, которое может принимать слово расположенное после предлога (Тип данных: целое число);
* ID\_падеж\_слова – идентификатор падежа, в котором должно находится слово, расположенное после предлога (Тип данных: целое число);
* ID\_сем\_значение\_фрейма – идентификатор семантического значения фрейма (Тип данных: целое число);
* Комментарий – необязательное поле, в котором может содержаться пример использования данного фрейма.

*Компонент разрешения имен*

В следующих разделах будет описана проблема неоднозначности именования параметров в онтологиях, обнаруженная при разработке алгоритма, и данный компонент содержит информацию необходимую для преодоления указанной проблемы при построении запроса. Схема этого компонента представлена на рисунке 5.

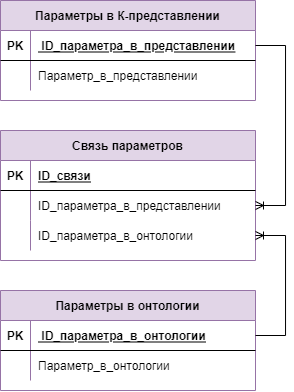


Рисунок 5 – Схема компонента разрешения имен параметров

*Описание таблиц компонента разрешения имен*

**Таблица «Параметры в К-представлении»**

Содержит список параметров, используемых в К-представлении.

*Поля таблицы «Параметры в К-представлении»:*

* ID\_параметра\_в\_представлении – идентификатор параметра, используемого в К-представлении (Тип данных: целое число);
* Параметр\_в\_представлении – параметр, используемый в К-представлении (Тип данных: строка переменной длины).

**Таблица «Параметры в онтологии»**

Содержит список параметров, используемых в онтологии.

*Поля таблицы «Параметры в онтологии»:*

* ID\_параметра\_в\_онтологии – идентификатор параметра, используемого в онтологии (Тип данных: целое число);
* Параметр\_в\_онтологии – параметр, используемый в онтологии (Тип данных: строка переменной длины).

**Таблица «Связи параметров»**

Содержит информации о связи между параметрами К-представления и онтологии.

*Поля таблицы «Связи параметров»:*

* ID\_связи – идентификатор конкретной связи параметров (Тип данных: целое число);
* ID\_параметра\_в\_представлении – идентификатор параметра, используемого в К-представлении (Тип данных: целое число);
* ID\_параметра\_в\_онтологии – идентификатор параметра, используемого в онтологии (Тип данных: целое число).

**Общая схема лингвистической базы данных**

Общая схема всей ЛБД представлена на рисунке 6.

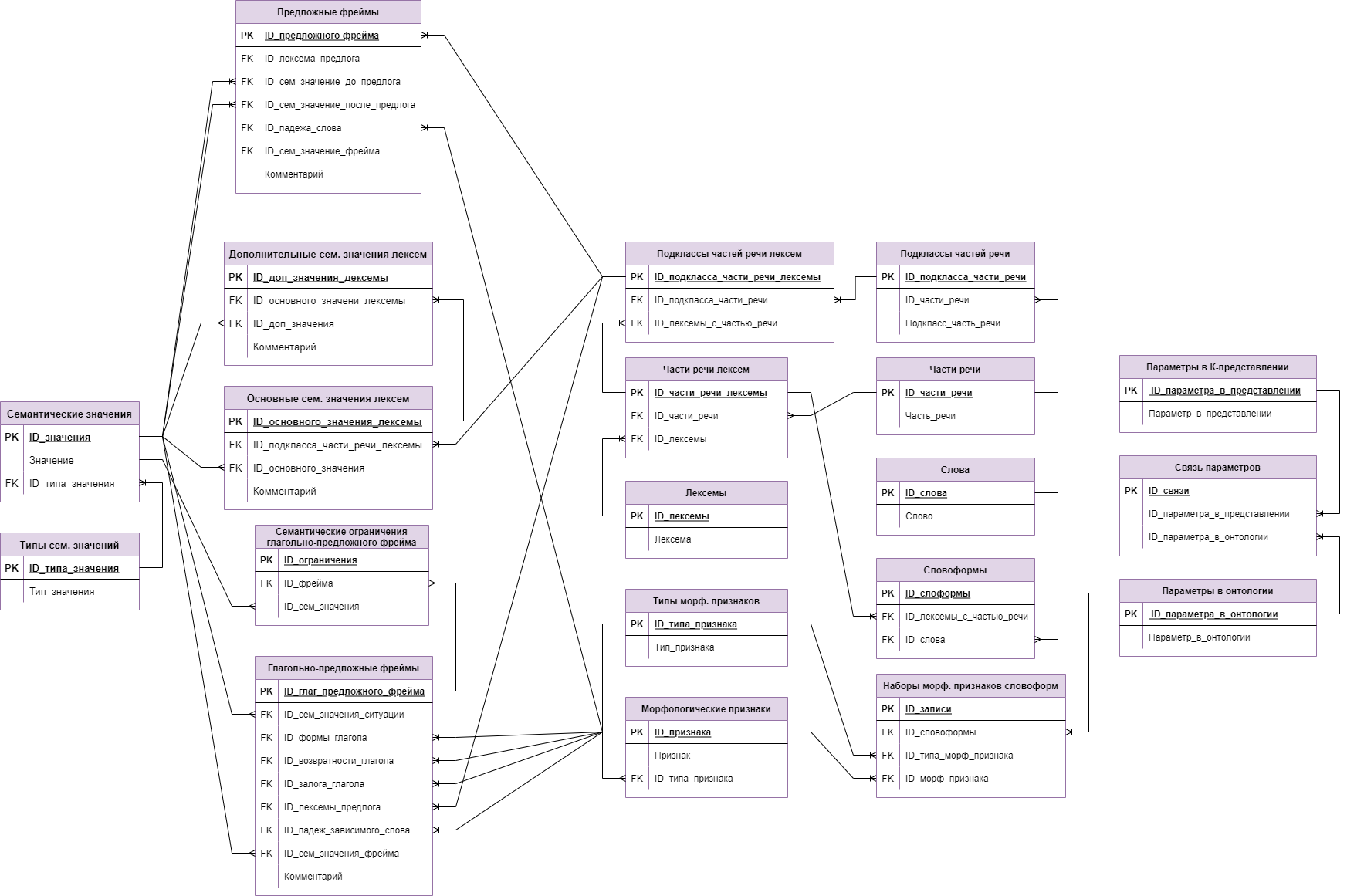


Рисунок 6 – Общая схема лингвистической базы данных