

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ**

**Новосибирский государственный университет (НГУ)  
Факультет информационных технологий**

**Кафедра систем информатики**

**Магистерская диссертация**

Енсегенова Айжан Абильхановна

*Разработка и реализация веб-приложения для  
автоматизированной генерации словоформ географических  
названий*

Научный руководитель  
д.т.н., доцент, с.н.с. ИВТ СО РАН  
Барахнин Владимир Борисович

**Новосибирск 2013**

## СОДЕРЖАНИЕ

1.	ВВЕДЕНИЕ .....	3
2.	ПОСТАНОВКА ЗАДАЧИ .....	5
3.	ОПИСАНИЕ АЛГОРИТМА .....	6
3.1	Основные понятия .....	6
3.2	Процедуры морфологического анализа и синтеза слов.....	6
4.	ОПИСАНИЕ РАБОТЫ ПРОГРАММЫ .....	9
4.1	Реализация. Структура баз данных.....	14
4.2	Примеры работы веб-приложения .....	19
5.	ЗАКЛЮЧЕНИЕ.....	25
6.	ЛИТЕРАТУРА .....	26

## ВВЕДЕНИЕ

В системах орфографического контроля русских текстов, системах автоматического индексирования документов и системах машинного перевода текстов с русского языка на иностранный и с иностранного языка на русский используется принцип построения алгоритма морфологического анализа текстов. Он основан на принципе аналогии, который описывается в работе Белоногова Г. Г. [3].

При автоматической обработке текста возникает проблема «новых» слов. Для синтаксического анализа и синтеза необходимо знать грамматические характеристики слов. Если слова в словаре нет, то морфологический анализ не может быть выполнен, а, следовательно, не могут быть определены грамматические характеристики слова.

Для того чтобы определить грамматические характеристики слов без словаря, Белоногов предложил принцип аналогии. Он основан на том, что существует сильная корреляционная связь между грамматическими характеристиками слов и буквенным составом их концов. Например: организация, приватизация, концентрация имеют ж. р., им. п. и ед. ч.

Свободно распространяемый программный продукт Ispell [1], изначально предназначенный для проверки орфографии на разных языках, в том числе русском [18] (язык проверки определяется словарем, который подключает пользователь). Основной подход, положенный ныне в основу словаря, заключается в использовании нормализованной формы слова и правил словоизменения, отвечающих грамматике русского языка. Все слова были разбиты на флективные классы (типы словоизменения), каждому из которых ставилась в соответствие система окончаний всех словоформ слова-представителя. По этой причине словарь одновременно содержит и важную информацию о морфологии слов, которая необходима для современных русскоязычных поисковых систем. А значит, теоретически, существует алгоритм, который каждому слову, для которого известен флективный класс, сопоставляет всевозможные формы этого слова путем присоединения определенных окончаний к основе.

**Актуальность работы.** Для задач извлечения из текстов географических наименований необходимо иметь все их словоформы. Необходимо пополнение лексического словаря их словоформами. Оно должно проводиться в экспертном режиме, однако работа эксперта по генерации всех словоформ нового слова весьма трудоемка: для существительного с учетом изменения падежа (изменение категории числа для географических названий не происходит) нужно выписать 6 словоформ, для

прилагательного с учетом изменения падежа – 6 словоформы (при этом следует учитывать, что притяжательные прилагательные нередко следует изменять и по родам: *Новосибирская область, Новосибирский район, Новосибирское водохранилище*, т.е. количество словоформ в данном случае достигает 18, а точнее – 24, если еще вспомнить про *Новосибирские острова*). Такой объем механической работы, помимо больших трудозатрат, чреват неизбежным появлением опечаток. Склонение столь большой выборки слов вручную – не эффективно, и не дает гарантий безошибочного ввода данных. А значит, возникает необходимость создать веб-приложение на основе алгоритма Белоногова, позволяющую автоматически генерировать словоформы и разработать интерфейс, удобный и достаточно простой для использования.

**Целью диссертационной работы** являлось создание веб-приложения позволяющего автоматически генерировать словоформы географических названий и разработка удобного и достаточно простого для использования интерфейса. Для достижения цели требовалось решить следующие задачи:

1. Изучить такие процедуры морфологического анализа и синтеза слов, как:
  - a. Морфологическая классификация слов.
  - b. Общий порядок морфологического анализа и синтеза слов.
  - c. Флективный анализ и синтез слов.
  - d. Алгоритмы морфологического анализа и синтеза слов.
2. Освоить технологии Web-программирования.
3. Определить требования к создаваемой программе:
  - a. Программа должна обладать простым и удобным интерфейсом.
  - b. Программа должна добавлять уникальные формы слова в базу данных.
4. Разработать пользовательский интерфейс.
5. Выполнить программную реализацию.
6. Провести оценку полученных результатов.

**Новизна работы** заключается в том, что генерация словоформ географических названий имеет особенности, отсутствующие при генерации словоформ имен нарицательных: географические названия могут быть сложносоставными, например, *Великие Луки или Петропавловск-Камчатский*.

Ниже излагаются содержание и результаты работы.

## ПОСТАНОВКА ЗАДАЧИ

Для автоматизации работы эксперта построено веб-приложение, автоматически генерирующее все словоформы заданного слова (существительного или прилагательного) русского языка [21]. Ограничимся только существительными и прилагательными, поскольку именно эти части речи обычно выступают в качестве новых слов практически в любой предметной области, а глаголы в подавляющем большинстве случаев не являются специфическими для той или иной предметной области и включены в основной словарь Ispell. Очевидно, указанное ограничение вполне допустимо и при решении задачи генерации словоформ географических названий.

В основе работы веб-приложения лежит алгоритм Г.Г.Белоногова [3], использующий разбиение слов языка на флективные классы, т.е. типы словоизменения, каждому из которых ставилась в соответствие система окончаний всех словоформ слова-представителя (основа существительных и прилагательных, как правило, остается неизменной). Всего Г.Г.Белоноговым выделено для существительных 66 флективных классов, для прилагательных –12, каждому из которых поставлен в соответствие полный набор окончаний.

Таким образом, стоит задача дополнить базу данных всевозможными формами слов.

Учитывая вышесказанное, можно сформулировать основные цели работы так:

- Автоматизировать процесс генерации всевозможных словоформ географического названия
- Определить требования к создаваемому веб-приложению
- Разработать интерфейс достаточно удобный для пользователя без специальных знаний

## ОПИСАНИЕ АЛГОРИТМА

### 3.1 Основные понятия

1. Флективный класс – класс слов, имеющих одинаковые признаки склонения. То есть слова, которые в определенной форме имеют одинаковые окончания. Характеризуется набором признаков или словом-представителем.

Например:

- Флективный класс №6.
- Слово-представитель: Томск.
- Окончания: нулевое, -ом.

2. Слово-представитель – слово, принадлежащее определенному флективному классу, обладающее набором признаков, удовлетворяющих флективному классу.

3. Словоформа – определенная форма слова.

Например, словоформы слова Томск:

- Томска
- Томском
- Томске, и др.

4. Типичная форма – форма слова, окончание которой используется для отнесения анализируемого слова к тому или иному флективному классу.

Например, для флективного класса №6 типичными являются формы:

- Томск
- Томском

### 3.2 Процедуры морфологического анализа и синтеза слов

#### **Морфологическая классификация слов.**

Морфологический анализ слов применяется с целью отождествления их различных форм и получения грамматической и семантической информации, необходимой на последующих этапах обработки текстов. Морфологический синтез – с целью получения различных форм слов при декодировании текстовой информации и выдаче ее человеку. Морфологический анализ и синтез могут строиться как на базе словаря основ слов, так и на базе словаря словоформ. В основу построения алгоритмов автоматического морфологического анализа и синтеза положено разбиение всех слов на классы,

определяющие характер изменения буквенного состава форм слов. Эти классы условно названы *морфологическими*.

Морфологические классы слов делятся на два вида:

- 1) Основоизменяющие классы, характеризующие систему изменения основ.
- 2) Флективные классы слов.

*Флективные классы* изменяемых слов выделялись на основе анализа их синтаксической функции и систем падежных, личных и родовых окончаний. Классы неизменяемых слов выделялись только по синтаксическому принципу.

По своей синтаксической функции изменяемые слова объединены в следующие группы:

- 1) Существительные.
- 2) Прилагательные.

Группа «Существительные», в свою очередь, состоит из нескольких подгрупп, выделенных по признакам рода и типу слова (сложносоставное). В каждой группе и подгруппе слова распределены по флективным классам.

Флективный класс может быть охарактеризован либо некоторой системой признаков, либо словом-представителем, которое является носителем этих признаков. Признаками, по которым изменяемое слово может быть отнесено к определенному классу, являются:

- 1) Принадлежность к одной из синтаксических групп (или подгрупп).
- 2) Система окончаний (тип словоизменения).

Определение принадлежности изменяемого слова к синтаксической группе или подгруппе обычно не вызывает затруднений, так как в основу разделения на группы и подгруппы положена традиционная классификация слов.

Для характеристики системы окончаний слова нет необходимости перечислять окончания всех его форм в соответствии с рисунком 1. Обычно достаточно сделать это лишь для нескольких типичных форм. В качестве таких типичных форм для группы «существительные» приняты формы именительного и творительного падежей единственного числа; для группы «прилагательные» - формы именительного падежа единственного числа мужского и женского рода, родительного падежа единственного числа мужского рода и именительного падежа множественного числа.

**Список окончаний слов**

01 — ами	21 — ат	41 — мя	61 — ям
02 — его	22 — ах	42 — ов	62 — ят
03 — еми	23 — ая	43 — ое	63 — ях
04 — ему	24 — ев	44 — ой	64 — яя
05 — емя	25 — ее	45 — ом	65 — † (нуль)
06 — ете	26 — ей	46 — ою	66 — а
07 — ешь	27 — ем	47 — ум	67 — е
10 — ими	30 — ет	50 — ут	70 — и
11 — ите	31 — ех	51 — ух	71 — й
12 — ишь	32 — ею	52 — ую	72 — о
13 — ого	33 — ие	53 — ые	73 — у
14 — ому	34 — ий	54 — ый	74 — ы
15 — умя	35 — им	55 — ым	75 — ь
16 — ыми	36 — ит	56 — ых	76 — ю
17 — ями	37 — их	57 — ют	77 — я
20 — ам	40 — ми	60 — юю	

Рисунок 1. Список окончаний слов

Некоторые классы существительных мужского и женского рода имеют одинаковые окончания во всех формах, принятых в качестве типичных, хотя другие их формы не совпадают.

**Флективный анализ и синтез слов.**

Морфологический анализ и синтез слов производится с помощью словаря основ и ряда вспомогательных таблиц. Морфологический анализ слова начинается с его флективного анализа, с целью правильного выделения его основы.

Флективный анализ слова производится с помощью обращенной морфологической таблицы и списка окончаний. Обращенная морфологическая таблица, в случае со специальными математическими терминами, состоит из двух частей, одна из которых используется для синтеза форм существительных, а вторая - для прилагательных.

Полями таблицы является морфологическая информация, то есть, перечислены признаки всевозможных форм слова. Записи – строки, состоящие из номера флективного класса и номеров окончаний, соответствующих для каждой морфологической информации.



## ОПИСАНИЕ РАБОТЫ ПРОГРАММЫ

В случае добавления в словарь географических названий число возможных флективных классов для существительных значительно уменьшается.

Это происходит, во-первых, за счет классов, относящихся к одушевленным существительным как мужского, так и женского рода (таких классов соответственно 19 и 8). Возможные совпадения названий географических объектов с одушевленными нарицательными существительными (*Орёл, Горняк, Чуваши* и т.п.) не нуждаются в специальном анализе, поскольку такая омонимия выявляется заранее в процессе предварительной работы с тезаурусом при составлении списка географических названий, имеющих «негеографические» омонимы, и «многозначных» названий, а это означает, что образец склонения слова-омонима уже имеется в лексическом словаре. Что же касается совпадения названий географических объектов с одушевленными собственными существительными (русскими фамилиями), то, как известно, соответствующие географические названия склоняются по неодушевленному образцу: с *С.М.Кировым*, но с *городом Кировом*.

Во-вторых, флективные классы для неодушевленных существительных зачастую различаются типом склонения лишь во множественном числе, однако для тех географических названий, которые соответствуют форме единственного числа, нет необходимости генерировать словоформы множественного числа.

Наибольшее количество возможных флективных классов, из которых приходится делать выбор, возникает при генерации словоформ географических названий, изначально стоящих во множественном числе: *Печоры, Спас-Клепики, Выгоничи* и *Ливны* относятся к разным флективным классам. Впрочем, при омонимии географических названий с неодушевленными нарицательными существительными образец склонения слова-омонима также имеется в лексическом словаре, что, в частности, исключает необходимость генерации словоформ географических названий, совпадающих со множественным числом нарицательных неодушевленных нарицательных существительных.

Что же касается прилагательных, входящих в географические названия, то в лексический словарь не входят либо притяжательные прилагательные, относящиеся к географическим названиям (*Болотнинский район*), либо диалектные, простонародные и т.п. слова (*Верхнекокишенгский Погост*), либо прилагательные, выступающие в качестве имен существительных (*Новокручининский*). При этом большинство подобных слов относится к одному флективному классу.

Таким образом, работа с веб-приложением заключается в следующем. Обработывая новое слово, эксперт устанавливает при необходимости его начальную форму и указывает его тип: независимое существительное или прилагательное. При выборе независимого существительного на следующем шаге необходимо указать его род и число. Для прилагательного дополнительные характеристики не указываются.

Для уменьшения размеров надклассов, на которые разбиты флективные классы, применяется модификация алгоритма Г.Г.Белоногова, описанная в работе [21], состоящая в автоматическом анализе окончаний нормализованной словоформы внутри каждого надкласса, что приводит к значительному уменьшению количества элементов, из которых предстоит сделать выбор. Тем самым программа автоматически проводит предварительный анализ окончания слова, отсеивая те классы, к которым данное слово заведомо принадлежать не может. После этого нужно выбрать флективный класс, которому соответствует слово. Для выбора предоставляется таблица возможных флективных классов, которые определяются словом-представителем и его несколькими характерными словоформами. Количество объектов-альтернатив (с учетом указанных выше особенностей склонения географических названий) доведено до рекомендуемого когнитивной психологией (не более 7-9 альтернатив).

Мысленно просклоняв данное слово по указанным формам и сравнив полученные окончания с окончаниями из таблицы, можно однозначно определить его флективный класс. После этого программа генерирует все формы слова, отображая их в виде таблицы, в которой они распределены по падежам и родам (если это прилагательное). Выводится список уникальных словоформ, так как обычно слово может иметь одинаковые окончания в разных формах. На основании этого списка эксперт принимает решение о занесении словоформ в словарь.

### **Решение проблемы омонимии**

Географические названия бывают омонимичны другим словам, являющимися именами как нарицательными: *Орёл*, *Белая* и т.п., так и собственными: *Киров*, *Кострома* и т.п. Кроме того, нередко одно и то же название носят сразу несколько различных географических объектов. Возникает необходимость отсеять из полученного набора слов омонимы географических названий, таковыми не являющиеся, а также установить, к какому конкретно географическому объекту относится «многозначное» название.

Среди имен собственных ведущим по численности разрядом являются топонимы, или географические имена (названия). Топонимы – это собственные имена любых географических объектов: водных (различных водоемов – рек, морей, озер, болот, прудов,

водохранилищ и т.п. – т.н. гидронимы) и земных (объектов рельефа – возвышенностей, падей, равнин, а также островов, мысов и т.п. – т.н. оронимы), как естественных, природных, так и искусственных, созданных руками человека или придуманных им (государств, территорий, населенных пунктов и т.п.). «Географические названия, вошедшие в общую речь, образуют тот ее культурно-исторический слой, который является принадлежностью литературного языка» [14]. Изучаются топонимы в топонимике – разделе ономастики (науки об именах собственных). Специфические черты всех имен собственных, в отличие от имен нарицательных, – это особенности их семантики, непереводаемость, возможность сознательного регулирования со стороны носителей языка и – грамматические особенности. «В своем грамматическом поведении в потоке речи географические названия отличны не только от имен нарицательных, но и, будучи существительными, друг от друга».

Морфологическая специфика имен собственных вообще мало изучена, что определяет необходимость ее дальнейшей разработки. В первую очередь это касается склонения собственных имен, в частности топонимов, ибо, как показывает речевая практика, многие носители языка не знают норм словоизменения некоторых групп топонимов или не подчиняются им. Современная норма словоупотребления географических названий и неустойчива, и неоднородна, она не может быть сведена к одному на все случаи пригодному правилу. СМИ и устный бытовой диалог дают разную картину употребления склоняемых форм топонимов, в то время как «различительные и адресные функции географических названий требуют их правильного написания в склоняемой форме в официальных документах, на картах и в любых других публикациях. В этом практически заинтересованы органы государственного управления, учреждения транспорта и связи, науки, просвещения и культуры, пресса, радио и телевидение».

В данной работе описана современная речевая практика склонения топонимов. Склонение топонимов рассматривается в рамках литературного языка, основным признаком которого является понятие нормы, т.е. «социально одобряемого правила, объективированного реальной речевой практикой, отражающего закономерности языковой системы и подтверждаемого словоупотреблением авторитетных писателей», и ее регламентации, т.е. зафиксированности в учебниках и словарях. Для литературной речи характерна сознательная установка на правильность, ориентация на образцы, стремление строить свою речь по примеру той, которую говорящий принимает за образец. Говоря о языковой норме, обычно указывают на ее устойчивость, стабильность,

общераспространенность и обязательность соблюдения нормативных рекомендаций. Однако кодификация может отставать от современной языковой нормы во времени или расходиться с ней в силу субъективных причин. Так возникает вариантность словоупотребления – «объективное и неизбежное следствие языковой эволюции». Все сказанное в полной мере касается и имен собственных, в том числе топонимов. Отсутствие литературно нормированной лексикографической фиксации словоизменения ономастической лексики сказывается на вариативности ее склонения.

Подавляющее большинство географических названий по существующим нормам литературного языка склоняются. Склонение топонимов в литературном языке ориентировано на склонение апеллятивов (имен нарицательных), однако имеет свою специфику, поскольку конечные элементы многих топонимов не находят себе подобия в обычных словах русского языка. Поэтому склоняются по соответствующим типам русского склонения.

**Склонение географических названий в русском языке:** очень многие собственные имена географических объектов — географические названия (топонимы), прежде всего славянские, в русском языке склоняются по падежам.

Название географического объекта, употреблённое в функции приложения с родовыми географическими терминами *город, село, деревня, хутор, река* и т. п., согласуется с определяемым словом, то есть склоняется, если топоним русский, славянский или давно заимствован и освоен русским языком.

Правильно: *в городе Москве* (не *в городе Москва*), *в городе Санкт-Петербурге*, *у города Ишимбая* (не *у города Ишимбай*), *из города Киева*, *на берегу реки Волги*, *на Зелёном мысе* (не *на мысе Зелёный*), *долина Сухого ручья* (не *долина ручья Сухой*), *на Русском острове* (не *на острове Русский*).

Названия, имеющие форму полного **прилагательного**, как правило, склоняются: *в городе Железнодорожном*, *на Красной площади*, *на Зелёном проспекте* (не *на проспекте Зелёный*), *на Ладожском озере*, *по Белой реке*, *на Лысом холме* (не *на холме Лысый*), *в Баргузинском заповеднике*, *на Гыданском полуострове*, *в Мозамбикском проливе*, *в Силезском воеводстве*.

Русские и другие славянские топонимы на **-ов(о)**, **-ёв(о)**, **-ев(о)**, **-ин(о)**, **-ын(о)** традиционно склоняются: *храм в Осташкове*, *вокзал в Венёве*, *старый город в Люблине*, *телебашня в Останкине*, *дача в Переделкине*, *шоссе к Строгину*,

*строительство в Новокосинé, маршрут из Люблина́, политехнический колледж в Колпине.*

Первая часть **сложносоставных** топонимов должна склоняться как в сочетании с родовым термином, так и без него, если топоним русский или освоен русским языком: *из Камня-Каширского, в Переславле-Залесском, в Могилёве-Подольском, в Ростове-на-Дону; в городе Петропавловске-Камчатском, в городе Ростове-на-Дону.* Бывают, однако, исключения: *в Гусь-Хрустальном.*

Следует сказать о **сдвоенных топонимах**, у которых первая часть названия морфологически имеет средний род: Орехово-Зуево, Ликино-Дулёво, Орехово-Борисово, Выхино-Жулебино, Хорошёво-Мнёвники, Царицыно-Дачное (прежнее название железнодорожной станции), Коньково-Деревлёво и др. Первая часть таких названий — как и другие подобные топонимы — традиционно склоняется: *электричка из Орехова-Зуева, поселились в Орехове-Борисове, подъехал к Царицыну-Дачному.*

#### 4.1 Реализация. Структура баз данных

Для работы веб-приложения требуются следующие баз данных.

1. Class – список флективных классов слов. Флективный класс может быть охарактеризован либо некоторой системой признаков, либо словом-представителем, которое является носителем этих признаков. Здесь содержится информация о каждом флективном классе, а именно:

- Уникальный номер флективного класса
- Слово-представитель
- Окончания типичных форм

I Для существительных:

- Именительный падеж, единственного числа
- Творительный падеж, единственного числа

II Для прилагательных:

- Именительный падеж, мужского рода, единственного числа
- Именительный падеж, женского рода, единственного числа
- Родительный падеж, мужского рода, единственного числа
- Именительный падеж, множественного числа

Эта база данных содержит 78 классов и разделена на смысловые блоки.

В зависимости от выбранных характеристик для введенного слова, отображаются только те классы, которые соответствуют запросам пользователя.

2. Ending – список различных окончаний слов. Для характеристики системы окончаний слова нет необходимости перечислять окончания всех его форм. Обычно достаточно сделать это лишь для нескольких типичных форм “таблица 1”. В ней каждому окончанию присвоен уникальный идентификационный номер, с которым связаны другие таблицы. Эта таблица содержит 78 записей.

Таблица 1.

1	1	ами
2	2	его
3	3	еми

3. Obr-morf – Обращенная морфологическая таблица играет важную роль в процессе генерации словоформ. Именно в ней явно содержится информация о том, какая система окончаний соответствует каждому флективному классу. Обращенная

морфологическая база данных для существительных “таблица 2”. Флективный синтез изменяемых слов производится с помощью словаря основ, обращенной морфологической базы данных и списка окончаний. Обращенная морфологическая база данных состоит из нескольких частей, число которых определяется количеством синтаксических классов изменяемых слов.

Таблица 2.

5	7	6	5	7	7	0	6	1	0	7	3
1	7	6	1	7	7	0	4	1	0	7	3
1	7	6	1	7	0	0	4	1	0	7	3

В первом поле отображены номера флективных классов, остальные поля отображают информацию об окончаниях в различных падежах и числах (для существительного имеется два числа и шесть падежей, то есть 12 полей в таблице), которая берется из таблицы окончаний. При формировании буквенного кода изменяемых слов сначала номер основы заменяется ее буквенным кодом, выбранным из словаря. Затем с помощью обращенной морфологической таблицы и таблицы 1 отыскивается буквенный код окончания и приформировывается к буквенному коду основы слова.

Когда пользователь определяет флективный класс для анализируемого слова, выбираются значения из соответствующей строки. Основе слова прикрепляются все окончания из этой строки, таким образом, создавая все формы слова.

Число строк в этой таблице равно числу флективных классов существительных.

4. Obr-pril – Строение, аналогичное обращенной морфологической таблице для существительных, имеет обращенная морфологическая таблица для прилагательных “таблица 3”. Здесь система окончаний состоит из 24 позиций, так как для прилагательных необходимо хранить окончания шести падежей, трех родов в единственном числе, и шести падежей множественного числа.

Таблица 3.

06	4	3	4	3	5	5	3	4	4	4	3	5	5	3	7	5	0	7
07	4	3	4	3	5	5	3	4	4	4	3	5	5	3	6	5	6	6
10	4	3	4	3	5	5	3	4	4	4	3	5	5	3	7	5	0	7

Таблица содержит 12 строк, и так же связана с таблицей окончаний.

Таблица 4. Структура базы данных **class**.

Название поля	Тип	Описание
Id	Integer	Уникальный идентификатор класса
n	Integer	Номер флективного класса
example	Text	Словопредставитель флективного класса
Im_ed	Text	Окончание в именительном падеже в единственном числе
Tv_ed	Text	Окончание в творительном падеже в единственном числе

Таблица 5. Структура базы данных **ending**.

Название поля	Тип	Описание
Id	Integer	Уникальный идентификатор окончаний
n	Integer	Номер флективного класса
ex	Text	Различные окончания слов

Таблица 6. Структура базы данных **Obr-morf**.

Название поля	Тип	Описание
---------------	-----	----------



Id	Integer	Уникальный идентификатор класса
n	Integer	Номер флективного класса
Im_ed	Integer	Номер окончания в именительном падеже единственного числа
Rod_ed	Integer	Номер окончания в родительном падеже единственного числа
Dat_ed	Integer	Номер окончания в дательном падеже единственного числа
Vin_ed	Integer	Номер окончания в винительном падеже единственного числа
Tv_ed	Integer	Номер окончания в творительном падеже единственного числа
Pr_ed	Integer	Номер окончания в предложном падеже единственного числа

Таблица 7. Структура базы данных **Obr-pril**.

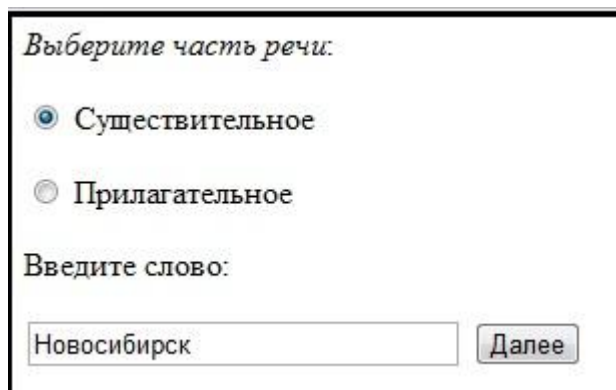
Название поля	Тип	Описание
Id	Integer	Уникальный идентификатор класса
n	Integer	Номер флективного класса
Im_ed	Integer	Номер окончания в именительном падеже единственного числа
Rod_ed	Integer	Номер окончания в родительном падеже единственного числа
Dat_ed	Integer	Номер окончания в

		дательном падеже единственного числа
Vin_ed	Integer	Номер окончания в винительном падеже единственного числа
Tv_ed	Integer	Номер окончания в творительном падеже единственного числа
Pr_ed	Integer	Номер окончания в предложном падеже единственного числа

## 4.2 Примеры работы веб-приложения

Итак, продемонстрируем, как работает веб-приложение.

Пользователю предоставляется возможность ввести географическое название (рис. 1.1, рис.1.2, рис.1.3). На этом же шаге необходимо выбрать часть речи (существительное или прилагательное).



Выберите часть речи:

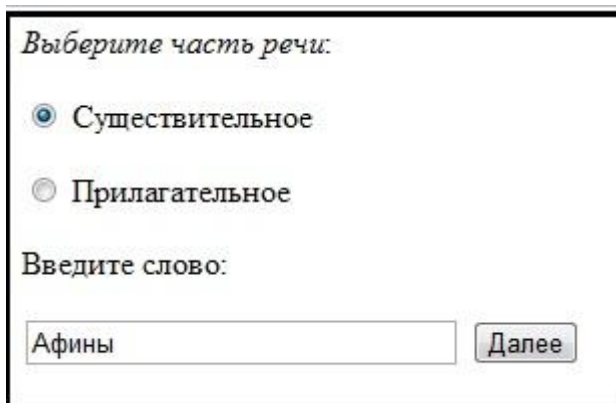
Существительное

Прилагательное

Введите слово:

Новосибирск

Рисунок 1.1. Выбор части речи для Новосибирска



Выберите часть речи:

Существительное

Прилагательное

Введите слово:

Афины

Рисунок 1.2. Выбор части речи для Афины

*Выберите часть речи:*

Существительное

Прилагательное

Введите слово:

Рисунок 1.3. Выбор части речи для Петропавловска-Камчатского

При выборе существительного на следующем шаге необходимо указать род или выбрать к какому типу это слово относится (рис.2.1, рис.2.2, рис.2.3)

Вы ввели слово **Новосибирск**, часть речи которого **существительное**.

*Ваше слово:*

Мужского рода (ед.ч)

Женского рода (ед.ч)

Среднего рода (ед.ч)

Геогр. название во мн.ч

Сложносоставное

Рисунок 2.1. Выбор мужского рода

Вы ввели слово **Афины**, часть речи которого **существительное**.

*Ваше слово:*

- Мужского рода (ед.ч)
- Женского рода (ед.ч)
- Среднего рода (ед.ч)
- Геогр. название во мн.ч
- Сложносоставное

Рисунок 2.2. Выбор географического названия во множественном числе

Вы ввели слово **Петропавловск-Камчатский**, часть речи которого **существительное**.

*Ваше слово:*

- Мужского рода (ед.ч)
- Женского рода (ед.ч)
- Среднего рода (ед.ч)
- Геогр. название во мн.ч
- Сложносоставное

Рисунок 2.3. Выбор сложносоставного

Далее, для существительного нужно выбрать номер флективного класса, к которому соответствует слово. Появляется таблица с номерами флективных классов, которые определяются словом-представителем и его несколькими характерными словоформами, то есть, какие именно окончания имеет данное слово в определенной форме (рис.3.1, рис.3.2). А для сложносоставного географического названия появляется страница с примерными географическими названиями, которые склоняются так же как вводимое сложносоставное географическое название (рис.3.3).

Вы ввели слово **Новосибирск**, часть речи которого **существительное**.

Выберите номер флективного класса, слово-представитель которого склоняется так же как **Новосибирск**:

№ класса	Слово-представитель	Им.п., ед.ч.	Тв.п., ед.ч. Кем? Чем?
<input type="checkbox"/> 1	Курган	+	ом
<input type="checkbox"/> 2	Воронеж	+	ом
<input type="checkbox"/> 6	Томск	+	ом
<input type="checkbox"/> 7	Таганрог	+	ом
<input type="checkbox"/> 10	Миасс	+	ом
<input type="checkbox"/> 11	Череповец	+	ем
<input type="checkbox"/> 14	Чаус	+	ом
<input type="checkbox"/> 15	Владикавказ	+	ом
<input type="checkbox"/> 16	Актаныш	+	ем
<input type="checkbox"/> 17	Лангепас	+	ом

Рисунок 3.1. Выбор флективного класса для Новосибирска

Вы ввели слово **Афины**, часть речи которого **существительное**.

Выберите номер флективного класса, слово-представитель которого склоняется так же как **Афины**:

№ класса	Слово-представитель	Им.п., мн.ч.	Род.п., мн.ч. Кого? Чего?
<input type="checkbox"/> 56	Шахты	+	+
<input type="checkbox"/> 31	Есентуки	+	ов

Рисунок 3.2. Выбор флективного класса для Афины

Вы ввели слово **Петропавловск-Камчатский**, часть речи которого **существительное**.

Выберите географическое название, которое склоняется так же как **Петропавловск-Камчатский**:

- Русское или давно освоенное название: Камень-Каширский
- Первая часть ср.р.: Орехово-Зуево
- Первая часть Спас-, Усть-, Соль-, Алма-: Усть-Каменгорск, Алма-Ата

Рисунок 3.3. Выбор флективного класса для Петропавловска-Камчатского

Просклоняв данное слово по указанным формам, и сравнив полученные окончания с окончаниями из таблицы, можно однозначно определить номер флективного класса. Здесь для удобства выбора предоставляется возможность сортировки по каждому полю таблицы. Кроме того, с учетом окончания анализируемого слова программа самостоятельно отбрасывает ненужные флективные классы, что так же облегчает выбор.

Всевозможные окончания слов хранятся в таблице, которая в свою очередь связана с морфологической таблицей, содержащей информацию о флективном анализе слова. С помощью нее программа генерирует все формы слова, отображая их в виде таблицы, в которой они распределены по падежам, числам и родам, если это прилагательное.

Программа выводит список словоформ, так как обычно слово может иметь одинаковые окончания в разных формах (рис.4.1, рис.4.2, рис.4.3).

Вы ввели флективный класс номер 6

Слово **Новосибирск** имеет следующие *словоформы*:

Основа *Новосибирск*

Падеж	Ед.ч.
Им.п., ед.ч. ( <i>Кто? Что?</i> )	Новосибирск
Род.п., ед.ч. ( <i>Кого? Чего?</i> )	Новосибирска
Дат.п., ед.ч. ( <i>Кому? Чему?</i> )	Новосибирску
Вин.п., ед.ч. ( <i>Кого? Что?</i> )	Новосибирск
Тв.п., ед.ч. ( <i>Кем? Чем?</i> )	Новосибирском
Пр.п., ед.ч. ( <i>О ком? О чем?</i> )	Новосибирске

Рисунок 4.1. Склонение «Новосибирска»

Вы ввели флективный класс номер 56

Слово **Афины** имеет следующие *словоформы*:

Основа *Афины*

Падеж	Мн.ч.
Им.п., ед.ч. ( <i>Кто? Что?</i> )	Афины
Род.п., ед.ч. ( <i>Кого? Чего?</i> )	Афины
Дат.п., ед.ч. ( <i>Кому? Чему?</i> )	Афинам
Вин.п., ед.ч. ( <i>Кого? Что?</i> )	Афины
Тв.п., ед.ч. ( <i>Кем? Чем?</i> )	Афинами
Пр.п., ед.ч. ( <i>О ком? О чем?</i> )	Афинах

Рисунок 4.2. Склонение «Афины»

Слово **Петропавловск-Камчатский** имеет следующие *словоформы*:

Основа *Петропавловск-Камчатский*

Падеж	Ед.ч.
Им.п., ед.ч. ( <i>Кто? Что?</i> )	Петропавловск-Камчатский
Род.п., ед.ч. ( <i>Кого? Чего?</i> )	Петропавловска - Камчатского
Дат.п., ед.ч. ( <i>Кому? Чему?</i> )	Петропавловску - Камчатскому
Вин.п., ед.ч. ( <i>Кого? Что?</i> )	Петропавловск-Камчатский
Тв.п., ед.ч. ( <i>Кем? Чем?</i> )	Петропавловским - Камчатским
Пр.п., ед.ч. ( <i>О ком? О чем?</i> )	Петропавловске - Камчатском

Рисунок 4.3. Склонение «Петропавловска-Камчатского»

При нажатии на кнопку «Добавить в базу» программа вносит в базу все склоненные словоформы.



## ЗАКЛЮЧЕНИЕ

Изучены процедуры морфологического анализа и синтеза слов.

Реализован алгоритм Белоногова и разработан интерфейс, с помощью которого можно достаточно просто и удобно определить флективный класс слова.

Программа генерирует словоформы слова, отбирает уникальные и добавляет их в базу. Программа проверена на выборке географических названий.

Освоены технологии Web-программирования. Определены требования к создаваемому веб-приложению. Создан удобный простой интерфейс веб-приложения, где генерируются географические названия.

Итак, задачу можно считать технически решенной.

## ЛИТЕРАТУРА

1. *IsPELL* – Spell checker. – <http://directory.fsf.org/ispell.html>
2. Барандеев, А.В. **Склонение однословных топонимов на -а(я)** // Русский язык в школе, 1989, №5.
3. Белоногов, Г.Г. **Автоматизация процессов накопления, поиска и обобщения информации**, Издательство «Наука», Москва, 1979.
4. *Библиотека* морфологического анализа *phpMorphy*. – <http://phpmorphy.sourceforge.net>
5. Богатырев, В. И. **Автоматизированные информационные системы**, Издательство «Советское радио», Москва, 1973.
6. Граудина, Л.К. **Грамматическая правильность русской речи**. 2-е изд., Москва, 2001.
7. Граудина, Л.К. **Норма употребления топонимов с финалями -ов(о),-ев(о),-ин(о),-ын(о)** // Ономастика и норма. — М.: Наука, 1976.
8. Граудина, Л.К. **Современная норма склонения топонимов** (в сочетании с географическим термином) // Ономастика и грамматика. – М.: Наука, 1981.
9. Губарь, Н.Т. **Морфологический анализ слов на основе словаря словоформ**. – В кн.: Научно-техническая информация, М.: ВИНТИ, 1975, серия 2, №9.
10. Каневский, Е.А. **Некоторые вопросы пополнения морфологического словаря терминами предметной области** // Тр. междунар. семинара Диалог'2001 по компьютерной лингвистике и ее приложениям. – Аксаково, 2001. – Т.2. – С. 156-160.
11. Когаловский, М.Р. **Технология баз данных на персональных ЭВМ**. – М.: Финансы и статистика, 1992.
12. Котов, Р. Г. **Автоматизированные информационно-поисковые системы**, Издательство «Советское радио», Москва, 1968.
13. Лабунько, О.М. **Склонение географических названий в современном русском литературном языке**. Москва, 1964.
14. Левашов, Е.А. **Географические названия**. Словарь-справочник. СПб. 2000.
15. Мартин, Дж. **Организация баз данных в вычислительных системах**. – М.: Мир, 1978.
16. Официальный сайт. Руководство MySQL (<http://dev.mysql.com/doc/>)
17. Официальный сайт. Руководство PHP (<http://www.php.net/manual/en/>)

18. *Словарь* русского языка для Ispell. –  
<http://semiconductors.phys.msu.su/~swan/orthography.html>
19. Толковый словарь русского языка: в 4 т. / под ред. Д.Н. Ушакова. – М.: Сов. энцикл.; ОГИЗ; Гос. изд-во иностр. и нац. словарей, 1935-1940.
20. Ульман, Дж. **Основы систем баз данных** / пер. с англ. – М.: Финансы и статистика, 1983.
21. Шокин, Ю.И. **Проблемы поиска информации**. Новосибирск: Наука, 2010.