

**«Разработка и создание базы данных  
фонетического разбора слов русского  
языка»**

*Садуахас Нургуль*  
*гр. 7203*

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

---

В современном мире филология все более тесно взаимодействует с информатикой. Такое сближение было обусловлено двумя вещами: появлением мировой сети INTERNET и электронной полиграфии. Соответственно нам нужны методы и модели, позволяющие работать с новыми аналогами книг. В перспективе такие модели откроют широкие возможности для филологов. Автоматизация процесса обработки стихотворных текстов сделает возможным изучение больших поэтических и применение различных методов интеллектуального анализа данных.

**Целью работы** является разработка структуры базы данных фонетического разбора словоформ слов русского языка и создание алгоритма ее наполнения.

---

Достижение поставленной цели осуществляется путем решения следующих **задач**:

- ✗ исследование стилистических закономерностей поэтических материалов, в частности, исследование закономерностей образования стихотворных размеров;
- ✗ разработать и реализовать алгоритм обработки электронного словаря, извлекающий следующую информацию: орфографическая запись слова, ударение в слове, деление слова на слоги, фонетическая транскрипция слова и характеристика всех его звуков;
- ✗ разработать структуру базы данных фонетического разбора слов русского языка;
- ✗ для каждого извлеченного слова построить все его словоформы.

Для такого анализа используются фонетические словари. Наиболее полным из известных нам сетевых фонетических словарей открытого доступа – «Словарь полного фонетического разбора».

Главная Даля Ожегова Шведовой Ожегова Ушакова Ефремовой Синонимов Брокгауза и Ефрона СТС Иностранных слов Имена БЮС Анаграммы Опечатки Однокоренные слова Фонетический разбор Переводчик Тестирование Афоризмы

### СЛОВАРИ ОНЛАЙН

Искать

[А](#) - [\[ Б \]](#) - [В](#) - [Г](#) - [Д](#) - [Е](#) - [Ё](#) - [Ж](#) - [З](#) - [И](#) - [Й](#) - [К](#) - [Л](#) - [М](#) - [Н](#) - [О](#) - [П](#) - [Р](#) - [С](#) - [Т](#) - [У](#) - [Ф](#) - [Х](#) - [Ц](#) - [Ч](#) - [Ш](#) - [Щ](#) - [Ы](#) - [Э](#) - [Ю](#) - [Я](#) - [Б](#) - [Багр](#) - [Байс](#) - [Бала](#) - [Балл](#) - [Банд](#) - [Бара](#) - [\[ Баро \]](#) - [Баск](#) - [Баты](#) - [Бега](#) - [Беза](#) - [Безг](#) - [Безл](#) - [Безо](#) - [Безы](#) - [Белё](#) - [Бело](#) - [Бель](#) - [Берг](#) - [Берн](#) - [Бесп](#) - [Бесп](#) - [Бесс](#) - [Бесш](#) - [Библ](#) - [Биме](#) - [Биоп](#) - [Бисе](#) - [Благ](#) - [Блаж](#) - [Близ](#) - [Блок](#) - [Бога](#) - [Бого](#) - [Бойк](#) - [Боло](#) - [Бомб](#) - [Боро](#) - [Босн](#) - [Брад](#) - [Брат](#) - [Брет](#) - [Брок](#) - [Брош](#) - [Брюс](#) - [Будн](#) - [Буко](#) - [Бума](#) - [Бурл](#) - [Буго](#) - [Были](#) - [Бюре](#) -

<a href="#">Барограмма</a>	<a href="#">Баррикадный</a>	<a href="#">Барчонок</a>
<a href="#">Барограф</a>	<a href="#">Баррикады</a>	<a href="#">Барчук</a>
<a href="#">Барокамера</a>	<a href="#">Барриль</a>	<a href="#">Барщина</a>
<a href="#">Барокко</a>	<a href="#">Барристер</a>	<a href="#">Барщинник</a>
<a href="#">Бароклинность</a>	<a href="#">Барс</a>	<a href="#">Барщинный</a>
<a href="#">Бароклинный</a>	<a href="#">Барселонец</a>	<a href="#">Барщинский</a>
<a href="#">Барометр</a>	<a href="#">Барселонка</a>	<a href="#">Барыга</a>
<a href="#">Барометрический</a>	<a href="#">Барселонский</a>	<a href="#">Барынька</a>
<a href="#">Барон</a>	<a href="#">Барселонцы</a>	<a href="#">Барыня</a>
<a href="#">Барон Мюнхаузен</a>	<a href="#">Барсёнок</a>	<a href="#">Барыш</a>

На первом шаге извлекаются ссылки на буквы алфавита, затем ссылки на подразделы, соответствующие выбранной букве. Данные заносятся в базу данных «dictionary» в таблицы **letters** и **subletters**.

# Фонетический разбор слов с транскрипцией

А - Б - В - Г - Д - Е - Ё - Ж - З - И - Й - К - Л - М - Н - О - П - Р - С - Т - У - Ф - Х - Ц - Ч - Ш -  
Щ - Ы - Э - Ю - Я

[А] - Абри - Авар - Авиа - Авто - Авто - Авто - Агио - Агро - Аден - Адье - Азоф - Аква -  
Акро - Акум - Алда - Алле - Алты - Альт - Амби - Аmmo - Анаб - Анат - Англ - Анид - Анон -  
Анти - Анти - Анти - Анты - Апол - Апро - Арге - Арис - Арон - Арха - Архи - Аспе - Астм -  
Атак - Атте - Ауто - Ахва - Аэро -

А

А Джорно

А Капелла

А Не То

А Также

А Темпо

А То

А-а-а

А-конт

А-ля

А-ля Фуршет

Аароновец

Аароновский

Аароновцы

Аббревиация

Аболиционизм

Абброморфема

Аболиционист

Абвер

Аболиционистский

Абверовец

Аболиция

Абверовский

Абонемент

Абдикация

Абонементный

Абдомен

Абонемтодержатель

Абдоминальный

Абонент

Абдуктор

Абонентка

Абдукторный

Абонентный

Абдукция

Абонентский

Аберрационный

Абонирование

Аберрация

Абонированный

Абзац

Абонировать

Далее, с помощью регулярных выражений из выбранного подраздела извлекаются все слова вместе с их фонетическим разбором. Результаты заносятся в БД и выводятся на экран.

# Фонетический разбор слов с транскрипцией

[Вернуться к алфавиту](#)

Орфографическая запись слова	Ударение в слове	Деление слова на слоги (перенос слова)	Фонетическая транскрипция слова	Характеристика всех звуков	Словоформы
кинокамера	кинок'amera	ки-но-ка-ме-ра	[к'инак'ам'ира]	к [к'] - согласный, мягкий, глухой, парный и [и] - гласный, безударный н [н] - согласный, твердый, звонкий, непарный, сонорный о [а] - гласный, безударный к [к] - согласный, твердый, глухой, парный а [а] - гласный, ударный м [м'] - согласный, мягкий, звонкий, непарный, сонорный е [и] - гласный, безударный р [р] - согласный, твердый, звонкий, непарный, сонорный а [а] - гласный, безударный	кинокамера кинокамеры кинокамере кинокамеру кинокамерой кинокамеройо кинокамер кинокамерам кинокамерами кинокамерах
кинокартина	кинокарт'ина	ки-но-кар-ти-на	[к'инакарт'ина]	к [к'] - согласный, мягкий, глухой, парный и [и] - гласный, безударный н [н] - согласный, твердый, звонкий, непарный, сонорный о [а] - гласный, безударный к [к] - согласный, твердый, глухой, парный а [а] - гласный, безударный р [р] - согласный, твердый, звонкий, непарный, сонорный т [т'] - согласный, мягкий, глухой, парный и [и] - гласный, ударный н [н] - согласный, твердый, звонкий, непарный, сонорный а [а] - гласный, безударный	кинокартина кинокартины кинокартине кинокартину кинокартиной кинокартинойо кинокартин кинокартинам кинокартинами кинокартинах

## СТРУКТУРА ТАБЛИЦЫ БД:

Поля	Расшифровка	Тип
id	номер	INT(11)
spelling	орфографическая запись	VARCHAR(100)
accent	ударение в слове	VARCHAR(100)
syllables	деление на слоги	VARCHAR(100)
transcription	транскрипция	VARCHAR(100)
sounds	характеристика звуков	VARCHAR(3000)
wordforms	словоформы	VARCHAR(3000)

Однако использование «Словаря полного фонетического разбора» для анализа фонетически характеристик стиха осложняется тем, что в нем приведены только начальные формы слов, поэтому необходима генерация фонетической записи словоформ .

**Вавилонская Башня**

[Switch to English](#)

Проект этимологической базы данных



Начните [отсюда](#)

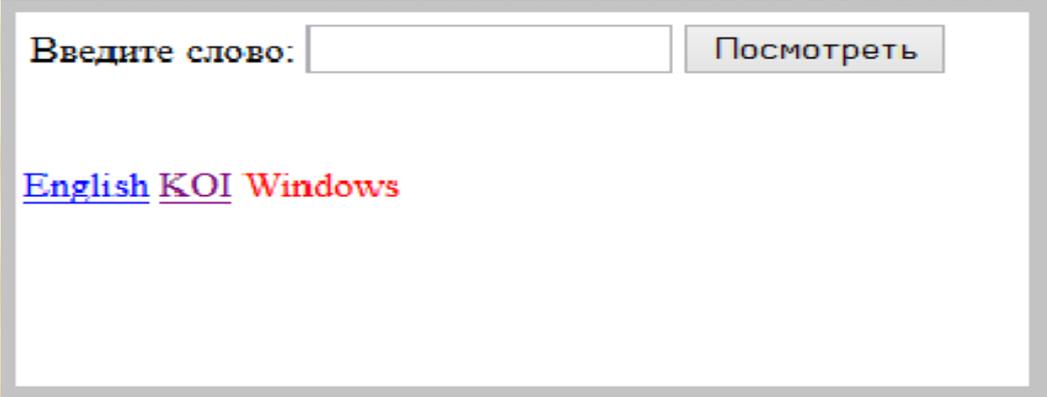
Русские словари и морфология

Copyright 1998-2005  
[Сергей Старостин](#)

"Вавилонская Башня" – международный Интернет – проект, посвященный сравнительно-историческому языкознанию.

Программой анализа выдаются следующие сведения:

- исходная словоформа (по А. А. Зализняку);
  - словарная информация, то есть морфологический индекс русского слова и имеющиеся комментарии из грамматического словаря А. А. Зализняка;
  - перевод, то есть набор словарных статей из словаря Мюллера, в которых содержится соответствующее русское слово, с готовыми отсылками на соответствующие словарные статьи;
  - морфологическая характеристика введенного русского слова.
- В случае многозначности введенной формы выводятся все варианты анализа.



Введите слово:

[English](#) [KOI](#) [Windows](#)

1. Исходная форма: атлас

Словарная информация: м 1a (географич.)

Перевод: [atlas I](#); [atlas II](#); [roadbook](#); [satin](#);

Морфологическая характеристика: Ns ,Asi

2. Исходная форма: атлас

Словарная информация: м 1a, P2 (ткань)

Морфологическая характеристика: Ns ,Asi

1 вариант:

	Ед. число	Множ. число
<b>Именительный</b>	а'глас	а'гласы
<b>Родительный</b>	а'гласа	а'гласов
<b>Дательный</b>	а'гласу	а'гласам
<b>Винительный неод.</b>	а'глас	а'гласы
<b>Творительный</b>	а'гласом	а'гласами
<b>Предложный</b>	а'гласе	а'гласах

2 вариант:

	Ед. число	Множ. число
<b>Именительный</b>	атла'с	атла'сы
<b>Родительный</b>	атла'са	атла'сов
<b>Родительный 2</b>	атла'су	
<b>Дательный</b>	атла'су	атла'сам
<b>Винительный неод.</b>	атла'с	атла'сы
<b>Творительный</b>	атла'сом	атла'сами
<b>Предложный</b>	атла'се	атла'сах

Пример: морфологический анализ слово «атлас»

# ЗАКЛЮЧЕНИЕ

---

- Изучена предметная область фонетического разбора слов русского языка;
- разработана структура базы данных фонетического разбора словоформ слов русского языка;
- создан алгоритм наполнения базы;
- разработан и реализован алгоритм обработки электронного словаря, извлекающий следующую информацию: орфографическая запись слова, ударение в слове, деление слова на слоги, фонетическая транскрипция слова и характеристика всех его звуков;
- для каждого извлеченного слова в словаре построены все его словоформы;
- создан понятный интерфейс для внесения корректировок в полученные данные.

---

Спасибо за внимание

## ЭТИ ТАБЛИЦЫ ИМЕЮТ ОДИНАКОВУЮ СТРУКТУРУ.

		Тип
id	номер	INT(11)
letter	буква / подраздел	VARCHAR(5)
refer	ссылка на раздел, относящийся к этой букве / ссылка на подраздел	VARCHAR(50)