
Research Data

05 Jun 2023

Experimental multi-dimensional scaling of web-scraping results from the A.A Zalizniak Grammatical Dictionary and the Russian National Corpus. Creating a corpus fragment of all possible word-forms of modified Russian sound verbs using web-scraping methodology. Compilation of a summary table for the present tense, future tense, imperative, imperfective and perfective gerund forms.

Irina Ivliyeva

Missouri University of Science and Technology, ivliyeva@mst.edu

Perry Koob

*Missouri S&T Information Technology, koobp@mst.edu*Follow this and additional works at: https://scholarsmine.mst.edu/research_data Part of the [Russian Linguistics Commons](#)

Recommended Citation

Ivliyeva, Irina and Koob, Perry, "Experimental multi-dimensional scaling of web-scraping results from the A.A Zalizniak Grammatical Dictionary and the Russian National Corpus. Creating a corpus fragment of all possible word-forms of modified Russian sound verbs using web-scraping methodology. Compilation of a summary table for the present tense, future tense, imperative, imperfective and perfective gerund forms." (2023). *Research Data*. 11.

https://scholarsmine.mst.edu/research_data/11

This Data is brought to you for free and open access by Scholars' Mine. It has been accepted for inclusion in Research Data by an authorized administrator of Scholars' Mine. This work is protected by U. S. Copyright Law. Unauthorized use including reproduction for redistribution requires the permission of the copyright holder. For more information, please contact scholarsmine@mst.edu.

Experimental multi-dimensional scaling of web-scraping results from the A.A Zalizniak Grammatical Dictionary and the Russian National Corpus. Creating a corpus fragment of all possible word-forms of modified Russian sound verbs using web-scraping methodology. Compilation of a summary table for the present tense, future tense, imperative, imperfective and perfective gerund forms.

Abstract

The emergence and development of electronic versions of dictionaries and corpus databases allows the researcher to finally do what was technically impossible on paper: to collect, compile and analyze the entire index of all possible verbal forms of different ranges and scales. This project attempts to improve the method of web extraction in relation to the source material (lexical-semantic group of Russian sound verbs, semantically modified at the word-forming level) and summarize the search results as an interactive summary table. A novel, four-position system of numbering the verbal forms have been introduced and a subsequent experimental multi-dimensional scaling of results successfully carried out. The output takes into account not only all documented (submitted) modifications of sound verbs from the [A.A Zaliznijak Grammatical Dictionary](#) and the [Russian National Corpus](#), but also reveals lacunae, dublets, and indicates new (potential) units. The results of the study may be useful for the development of various web applications for the search, collection, and visualization of linguistic material. Possibilities of combinatorial optimization in the processing of open and closed linguistic databases can be particularly important when extracting information from various digital lexicographic sources (across a single or multiple languages), from national linguistic corpora, as well as from digital text collections.

Появление и развитие электронных версий словарей русского языка и корпусных баз данных позволяют исследователю сделать то, что прежде было технически неосуществимо при работе с бумажными словарями: собрать, скомпилировать (соединить результаты) и проанализировать обобщающий индекс всех возможных глагольных форм в различных диапазонах и масштабах.

В данном проекте предпринята попытка не только усовершенствовать методику веб-извлечения применительно к исходному материалу (лексико-семантической группе русских звуковых глаголов, модифицированных на словообразовательном уровне), но и обобщить результаты поиска.

Разработана и введена новая четырехпозиционная система нумерации исходных форм глагола и их словоформ. Успешно проведено системное извлечение и фиксирование результатов поиска в виде многомерной сводной интерактивной таблицы.

Выходные данные документируют не только все словоформы звуковых глаголов, представленные в Грамматическом словаре А.А. Зализняка и Национальном корпусе русского языка, но и выявляют существующие дублеты, внутриязыковые лакуны, указывают на новые (потенциальные) единицы.

Результаты исследования могут быть полезны в разработке различных веб-приложений для поиска, сбора и визуализации лингвистического материала. Возможности комбинаторной оптимизации открытых и закрытых баз данных могут быть особенно значимы при извлечении информации из цифровых лексикографических источников (на одном или нескольких языках), из национальных языковых корпусов, а также из электронных текстовых коллекций.

Disciplines
Russian Linguistics

Ivliyeva, Irina V., Koob, Perry. **Experimental multi-dimensional scaling of web-scraping results from the A.A Zalizniak Grammatical Dictionary and the Russian National Corpus. Creating a corpus fragment of all possible word-forms of modified Russian sound verbs using web-scraping methodology. Compilation of a summary table for the present tense, future tense, imperative, imperfective and perfective gerund forms.** April 2022 – May 2023. Working theses. June 2023.

The emergence and development of electronic versions of dictionaries and corpus databases allows the researcher to finally do what was technically impossible on paper: to collect, compile and analyze the entire index of all possible verbal forms of different ranges and scales. This project attempts to improve the method of web extraction in relation to the source material (lexical-semantic group of Russian sound verbs, semantically modified at the word-forming level) and summarize the search results as an interactive summary table. A novel, four-position system of numbering the verbal forms have been introduced and a subsequent experimental multi-dimensional scaling of results successfully carried out. The output takes into account not only all documented (submitted) modifications of sound verbs from the [A.A Zaliznijak Grammatical Dictionary](#) and the [Russian National Corpus](#), but also reveals lacunae, dublets, and indicates new (potential) units. The results of the study may be useful for the development of various web applications for the search, collection, and visualization of linguistic material. Possibilities of combinatorial optimization in the processing of open and closed linguistic databases can be particularly important when extracting information from various digital lexicographic sources (across a single or multiple languages), from national linguistic corpora, as well as from digital text collections.

Ивлиева, И.В., Куб, Перри. **Экспериментальное многомерное шкалирование результатов веб-извлечения из Грамматического словаря А.А. Зализняка и базы данных Национального корпуса русского языка (НКРЯ). Создание фрагмента корпуса словоформ модифицированных глаголов звучания русского языка методом веб-извлечения. Составление сводной таблицы форм настоящего и будущего времени, повелительного наклонения, деепричастий несовершенного и совершенного вида.** Апрель 2022 – Май 2023. Рабочие тезисы. 5 июня 2023.

Появление и развитие электронных версий словарей русского языка и корпусных баз данных позволяют исследователю сделать то, что прежде было технически неосуществимо при работе с бумажными словарями: собрать, скомпилировать (соединить результаты) и проанализировать обобщающий индекс всех возможных глагольных форм в различных диапазонах и масштабах.

В данном проекте предпринята попытка не только усовершенствовать методику веб-извлечения применительно к исходному материалу (лексико-семантической группе русских звуковых глаголов, модифицированных на словообразовательном уровне), но и обобщить результаты поиска.

Разработана и введена новая четырехпозиционная система нумерации исходных форм глагола и их словоформ. Успешно проведено системное извлечение и фиксирование результатов поиска в виде многомерной сводной интерактивной таблицы.

Выходные данные документируют не только все словоформы звуковых глаголов, представленные в Грамматическом словаре А.А. Зализняка и Национальном корпусе русского языка, но и выявляют существующие дублиеты, внутриязыковые лакуны, указывают на новые (потенциальные) единицы.

Результаты исследования могут быть полезны в разработке различных веб-приложений для поиска, сбора и визуализации лингвистического материала. Возможности комбинаторной оптимизации открытых и закрытых баз данных могут быть особенно значимы при извлечении информации из цифровых лексикографических источников (на одном или нескольких языках), из национальных языковых корпусов, а также из электронных текстовых коллекций.

ПОСТАНОВКА ПРОБЛЕМЫ.

Система глагольных форм в русском языке имеет **матричную структуру** (существуют как полные парадигмы глаголов, так и дефектные - типа «побежу», «пропылесосу»). Известно, что современные словари не приводят **все** возможные формы слов. Воспроизвести в бумажном издании формы слов даже одной части речи - технически неисполнимая задача. Однако с появлением электронных версий словарей впервые возникает шанс её решить. Для чего необходимы таблицы глагольных форм? Прежде всего, чтобы заполнить «матрицы» (таблицы) каждого глагола всеми имеющимися в русском языке актуальными словоформами, которые фиксируются в лексикографических ресурсах. Далее, предполагается выявить существующие дублиеты, внутриязыковые лакуны, а также указывать на новые (потенциальные) единицы.

ХРОНОЛОГИЯ ПРОЕКТА.

С 2019 года на базе методологии веб-извлечения из электронных словарей (Грамматический словарь А.А. Зализняка и Национальный корпус русского языка) велась работа по составлению сводных таблиц, включающих глагольные формы всех времен (наст., буд., прош.) и наклонений (изъявительное, повелительное), учитывая дифференциацию по родам (м., ж., ср.), числам (ед., мн.) и лицам (1-е лицо, 2-е лицо, 3-е лицо). Подробнее смотрите библиографию в конце данной работы.

В период с мая 2021 по апрель 2022 года проводились отладка программы для веб-извлечения, оптимизация представления результатов поиска, расширение диапазона и синхронизация результатов поиска

одновременно в словаре Зализняка и НКРЯ. Подробнее об этом см.: Ivliyeva, Irina V., Koob, Perry. “Creating a corpus of all possible word forms of modified Russian sound verbs using web-scraping methodology. Compilation and adjustment of summary tables for the future tense, imperative, and gerund forms. Forms of “dual action”. Experimental multi-dimensional scaling of web-scraping results.” (2022). [Electronic resource]. Research Data. March 2022. – Access mode: https://scholarsmine.mst.edu/research_data/9/

ДОСТУП К ИСТОЧНИКАМ.

Результаты, полученные в период с мая 2021 по март 2022 [Ивлиева, Куб. 2022] года не являлись точными, полными или окончательными. Данные результаты можно назвать *промежуточными*, так как получены они были в процессе пробного обращения к источникам (Зализняк и НКРЯ) 21 февраля 2022 года, после очередной ступени отладки поисковой компьютерной программы.

Попытка получения точных, окончательных результатов в полном объёме 24 февраля 2022 года не принесла успеха: поисковая компьютерная программа *не получила доступа* к основному источнику - [Грамматическому словарю А.А. Зализняка](#). Авторы проекта регулярно (21-ого числа ежемесячно) проверяли доступ к данному источнику и документировали наличие или отсутствие доступа к нему, например:

Морфологический словарь русского языка - О морфологическом словаре morfologija.ru	21-Apr-22	9:30 CDT
Please Wait... Cloudflare morfologija.ru	21-Apr-22	
Please Wait... Cloudflare morfologija.ru	21-Apr-22	
Морфологический словарь русского языка - О морфологическом словаре morfologija.ru	21-Mar-22	8:42 CDT
Please Wait... Cloudflare morfologija.ru	21-Mar-22	
Please Wait... Cloudflare morfologija.ru	21-Mar-22	
Морфологический словарь русского языка - О морфологическом словаре morfologija.ru	21-Feb-22	16:07 CST
https://morfologija.ru/	21-May-22	7:51 CDT

ЗАДАЧА I.

В период с июня 2022 года по апрель 2023 года исследовательская группа не только отслеживала наличие/отсутствие доступа к основному источнику - [Грамматическому словарю А.А. Зализняка](#). Параллельно проводилась дальнейшая отладка компьютерного алгоритма и микро-тестирование эффективности посковой компьютерной программы с целью решения некоторых методических трудностей, которые были выявлены на предыдущем этапе исследования. Так, в процессе заполнения индивидуальных таблиц по каждому глаголу, маркирования элементов и последующего анализа их состава, была установлена необходимость введения унифицированной **четырёхступенчатой нумерации** элементов в [СВОДНОЙ ТАБЛИЦЕ](#) (колонка А).

Кроме того, были выявлены некоторые особенности технического характера представления результатов поиска, которые связаны непосредственно с семантикой глагольных модификаций, например:

1. Совпадение форм будущего времени и императива глаголов звучания, или **форм совместного действия**;

2. Омонимия форм модификаций у глаголов близкой семантики (при отсутствии указания ударения в исходном слове в источнике поиска). См. [Приложение 4](#). Ivliyeva, Koob. Appendix 4. Double_imperfective_gerund-2023-05-18

18 мая 2023 года была предпринята **успешная попытка** внедрить окончательно отлаженную версию программы веб-извлечения и получить систематические результаты, которые отражены в данной работе в виде обобщающей **сводной таблице** словоформ глаголов звучания, модифицированных на словообразовательном уровне: формы настоящего времени, будущего времени, повелительного наклонения, и их деепричастные формы несовершенного и совершенного вида.

Пример 1. Страница глагола АХАТЬ в электронном словаре А.А. Зализняка <http://morfologija.ru/>

Введите слово:

Словоформы слова "ахать"

ахать → ахать – инфинитив, действ. зал.

Часть речи: инфинитив - ахать

Часть речи: глагол

	Настоящее время	Прошедшее время	Будущее время	Повелительное наклонение
Я	ахаю	ахал; ахала; ахало		
Ты	ахаешь	ахал; ахала; ахало		ахай
Он/она	ахает	ахал; ахала; ахало		
Мы	ахаем	ахали		
Вы	ахаете	ахали		ахайте
Они	ахают	ахали		

Часть речи: деепричастие

	Несовершенный вид	Совершенный вид
ахая		ахав; ахавши

СВОДНАЯ ТАБЛИЦА. Условные обозначения и соответствия.**1) Колонки.**

Колонка	Тип данных	Title
A	Порядковый номер слова или словоформы	Number
B	Исходное слово - центр модификационного гнезда и словообразовательного гнезда (инфинитив)	Root
C	Модифицированный дериват (инфинитив)	Word
D	Личные местоимения	Personal Pronouns
E	Настоящее время (словоформы)	Present tense
F	Будущее время (словоформы)	Future tense
G	Повелительное наклонение (словоформы)	Imperative
H	Деепричастие, несовершенный вид (словоформы)	Imperferctive Gerund
I	Деепричастие, совершенный вид (словоформы)	Perferctive Gerund

2) Система нумерации по позициям в колонке А.

В колонке А имеются четырёхступенчатые/четырёхпозиционные номера.

A	B	C	D	E	F	G	H	I
number	root	word	pronoun	present	future	imperative	Imp. gerund	Perf. gerund
1.0.0.0	ахать	ахать	--	--	--	--	ахая	--

Позиция 1: основное слово (исходный немодифицированный глагол, колонка В).

A	B
1.0.0.0	ахать
2.0.0.0	бабахать

Позиция 2: модифицированный дериват (модифицированный глагол, колонка С).

А	В	С
1.1.0.1	ахать	ахнуть
1.2.0.0	ахать	<u>взахать</u>
1.3.0.1	ахать	<u>заахать</u>

Позиция 3: словоизменительная парадигма (личные местоимения, колонка D и словоформы в колонках E, F, G).

А	В	С	Д	Е	Ф
1.1.1.0	ахать	ахнуть	Я	--	ахну
1.1.2.0	ахать	ахнуть	Ты	--	ахнешь
1.1.3.0	ахать	ахнуть	Он/она	--	ахнет
1.1.4.1	ахать	ахнуть	Мы	--	ахнем
1.1.4.2	ахать	ахнуть	Мы	--	ахнем
1.1.5.0	ахать	ахнуть	Вы	--	ахнете
1.1.6.0	ахать	ахнуть	Они	--	ахнут

Позиция 4:

Немодифицированный глагол, деепричастие несовершенного вида, колонка Н.

А	В	С	Д	Е	Ф	Г	Н	І
1.0.0.0	ахать	ахать	--	--	--	--	ахая	--

Модифицированный глагол, деепричастия совершенного вида, колонка І).

А	В	С	Д	Е	Ф	Г	Н	І
1.3.0.1	ахать	заахать	--		--	--	--	заахав
1.3.0.2	ахать	заахать	--		--	--	--	заахавши

3) **Двойной прочерк** (--) в графе показывает, что в электронном словаре А.А. Зализняка <http://morphologija.ru/> данное слово (или его словоформа) не обнаружены.

ВЫБОРОЧНАЯ ПРОВЕРКА (вручную) РЕЗУЛЬТАТОВ ВЕБ-ИЗВЛЕЧЕНИЯ: выделение цветом.

А) Немодифицированный глагол, деепричастие несовершенного вида, результат поиска - колонка Н.

A	B	C	D	E	F	G	H	I
1.0.0.0	ахать	ахать	--	--	--	--	ахая	--

Б) Модифицированный глагол, деепричастия совершенного вида, результат поиска - колонка I.

A	B	C	D	E	F	G	H	I
1.1.0.1	ахать	ахнуть	--	--	--	--	--	ахнув
1.1.0.2	ахать	ахнуть	--	--	--	--	--	ахнувши

В) Немодифицированный глагол, деепричастие совершенного вида, результат поиска - колонка I (аномалия).

A	B	C	D	E	F	G	H	I
1.0.0.1	ахать	ахать	--	--	--	--	--	ахав
1.0.0.2	ахать	ахать	--	--	--	--	--	ахавши

См. [Приложение 3](#). Ivliyeva, Koob. Ivliyeva, Koob. Appendix 3. Double_perfective_gerund-2023-05-18

Г) Словоизменительная парадигма (личные местоимения, колонка D; результаты поиска (словоформы) в колонках E, F, G).

A	B	C	D	E	F	G	H	I
3.4.1.0	барабанить	пробарабанить	Я	--	пробарабаню	--	--	--

3.4.2.0	барабанить	пробарабанить	Ты	--	пробарабанишь	пробарабань	--	--
3.4.3.0	барабанить	пробарабанить	Он/она	--	пробарабанит	--	--	--
3.4.4.1	барабанить	пробарабанить	Мы	--	пробарабаним	пробарабаним	--	--
3.4.4.2	барабанить	пробарабанить	Мы	--	пробарабаним	пробарабаните	--	--
3.4.5.0	барабанить	пробарабанить	Вы	--	пробарабаните	пробарабаньте	--	--
3.4.6.0	барабанить	пробарабанить	Они	--	пробарабанят	--	--	--

Красным цветом выделены **формы совместного действия**.

См. [Приложение 2](#). Ivliyeva, Koob. Appendix 2. Double_pronoun_dual_action -2023-05-18

Д) Отсутствие цвета, не найдено в источнике (--).

A	B	C	D	E	F	G	H	I
1.2.0.0	ахать	взахать	--	--	--	--	--	--

См. [Приложение 1](#). Ivliyeva, Koob. Appendix 1. Words not_in_morfologija-2023-05-18

Пример 2. Фрагмент **СВОДНОЙ ТАБЛИЦЫ**.

A	B	C	D	E	F	G	H	I
number	root	word	pronoun	present	future	imperative	Imp. gerund	Perf. gerund
1.0.0.0	ахать	ахать	--	--	--	--	ахая	--
1.0.0.1	ахать	ахать	--	--	--	--	--	ахав
1.0.0.2	ахать	ахать	--	--	--	--	--	ахавши
1.1.0.1	ахать	ахнуть	--	--	--	--	--	ахнув
1.1.0.2	ахать	ахнуть	--	--	--	--	--	ахнувши
1.1.1.0	ахать	ахнуть	Я	--	ахну	--	--	--
1.1.2.0	ахать	ахнуть	Ты	--	ахнешь	ахни	--	--
1.1.3.0	ахать	ахнуть	Он/она	--	ахнет	--	--	--
1.1.4.1	ахать	ахнуть	Мы	--	ахнем	ахнем	--	--

1.1.4.2	ахать	ахнуть	Мы	--	ахнем	ахнемте	--	--
1.1.5.0	ахать	ахнуть	Вы	--	ахнете	ахните	--	--
1.1.6.0	ахать	ахнуть	Они	--	ахнут	--	--	--
1.2.0.0	ахать	взахать	--	--	--	--	--	--
1.3.0.1	ахать	заахать	--	--	--	--	--	заахав
1.3.0.2	ахать	заахать	--	--	--	--	--	заахавши
1.3.1.0	ахать	заахать	Я	--	заахаю	--	--	--
1.3.2.0	ахать	заахать	Ты	--	заахаешь	заахай	--	--
1.3.3.0	ахать	заахать	Он/она	--	заахает	--	--	--
1.3.4.1	ахать	заахать	Мы	--	заахаем	заахаем	--	--
1.3.4.2	ахать	заахать	Мы	--	заахаем	заахаемте	--	--
1.3.5.0	ахать	заахать	Вы	--	заахаете	заахайте	--	--
1.3.6.0	ахать	заахать	Они	--	заахают	--	--	--
1.4.0.1	ахать	поахать	--	--	--	--	--	поахав
1.4.0.2	ахать	поахать	--	--	--	--	--	поахавши
1.4.1.0	ахать	поахать	Я	--	поахаю	--	--	--
1.4.2.0	ахать	поахать	Ты	--	поахаешь	поахай	--	--
1.4.3.0	ахать	поахать	Он/она	--	поахает	--	--	--
1.4.4.1	ахать	поахать	Мы	--	поахаем	поахаем	--	--
1.4.4.2	ахать	поахать	Мы	--	поахаем	поахаемте	--	--
1.4.5.0	ахать	поахать	Вы	--	поахаете	поахайте	--	--
1.4.6.0	ахать	поахать	Они	--	поахают	--	--	--
1.5.0.0	ахать	поахивать	--	--	--	--	--	--
1.6.0.1	ахать	проахать	--	--	--	--	--	проахав
1.6.0.2	ахать	проахать	--	--	--	--	--	проахавши
1.6.1.0	ахать	проахать	Я	--	проахаю	--	--	--
1.6.2.0	ахать	проахать	Ты	--	проахаешь	проахай	--	--
1.6.3.0	ахать	проахать	Он/она	--	проахает	--	--	--
1.6.4.1	ахать	проахать	Мы	--	проахаем	проахаем	--	--
1.6.4.2	ахать	проахать	Мы	--	проахаем	проахаемте	--	--
1.6.5.0	ахать	проахать	Вы	--	проахаете	проахайте	--	--

1.6.6.0	ахать	проахать	Они	--	проахают	--	--	--
1.7.0.0	ахать	разахаться	--	--	--	--	--	разахавшись
1.7.1.0	ахать	разахаться	Я	--	разахуюсь	--	--	--
1.7.2.0	ахать	разахаться	Ты	--	разахнешься	разахайся	--	--
1.7.3.0	ахать	разахаться	Он/она	--	разахается	--	--	--
1.7.4.1	ахать	разахаться	Мы	--	разахаемся	разахаемся	--	--
1.7.4.2	ахать	разахаться	Мы	--	разахаемся	разахаетесь	--	--
1.7.5.0	ахать	разахаться	Вы	--	разахаетесь	разахайтесь	--	--
1.7.6.0	ахать	разахаться	Они	--	разахаются	--	--	--
2.0.0.0	бабахать	бабахать	--	--	--	--	бабахая	--
2.0.0.1	бабахать	бабахать	--	--	--	--	--	бабахав
2.0.0.2	бабахать	бабахать	--	--	--	--	--	бабахавши
2.1.0.1	бабахать	бабахнуть	--	--	--	--	--	бабахнув
2.1.0.2	бабахать	бабахнуть	--	--	--	--	--	бабахнувши
2.1.1.0	бабахать	бабахнуть	Я	--	бабахну	--	--	--
2.1.2.0	бабахать	бабахнуть	Ты	--	бабахнешь	бабахни	--	--
2.1.3.0	бабахать	бабахнуть	Он/она	--	бабахнет	--	--	--
2.1.4.1	бабахать	бабахнуть	Мы	--	бабахнем	бабахнем	--	--
2.1.4.2	бабахать	бабахнуть	Мы	--	бабахнем	бабахнемте	--	--
2.1.5.0	бабахать	бабахнуть	Вы	--	бабахнете	бабахните	--	--
2.1.6.0	бабахать	бабахнуть	Они	--	бабахнут	--	--	--
2.2.0.0	бабахать	забабахать	--	--	--	--	--	--
3.0.0.0	барабанить	барабанить	--	--	--	--	барабаня	--
3.0.0.1	барабанить	барабанить	--	--	--	--	--	барабанив
3.0.0.2	барабанить	барабанить	--	--	--	--	--	барабанивши
3.1.0.1	барабанить	забарабанить	--	--	--	--	--	забарабанив
3.1.0.2	барабанить	забарабанить	--	--	--	--	--	забарабанивши
3.1.1.0	барабанить	забарабанить	Я	--	забарабаню	--	--	--
3.1.2.0	барабанить	забарабанить	Ты	--	забарабанишь	забарабань	--	--
3.1.3.0	барабанить	забарабанить	Он/она	--	забарабанит	--	--	--
3.1.4.1	барабанить	забарабанить	Мы	--	забарабаним	забарабаним	--	--

3.1.4.2	барабанить	забарабанить	Мы	--	забарабаним	забарабанимте	--	--
3.1.5.0	барабанить	забарабанить	Вы	--	забарабаните	забарабаньте	--	--
3.1.6.0	барабанить	забарабанить	Они	--	забарабанят	--	--	--
3.2.0.1	барабанить	отбарабанить	--	--	--	--	--	отбарабанив
3.2.0.2	барабанить	отбарабанить	--	--	--	--	--	отбарабанивши
3.2.0.3	барабанить	отбарабанить	--	--	--	--	--	отбарабаня
3.2.1.0	барабанить	отбарабанить	Я	--	отбарабаню	--	--	--
3.2.2.0	барабанить	отбарабанить	Ты	--	отбарабанишь	отбарабань	--	--
3.2.3.0	барабанить	отбарабанить	Он/она	--	отбарабанит	--	--	--
3.2.4.1	барабанить	отбарабанить	Мы	--	отбарабаним	отбарабаним	--	--
3.2.4.2	барабанить	отбарабанить	Мы	--	отбарабаним	отбарабанимте	--	--
3.2.5.0	барабанить	отбарабанить	Вы	--	отбарабаните	отбарабаньте	--	--
3.2.6.0	барабанить	отбарабанить	Они	--	отбарабанят	--	--	--
3.3.0.1	барабанить	побарабанить	--	--	--	--	--	побарабанив
3.3.0.2	барабанить	побарабанить	--	--	--	--	--	побарабанивши
3.3.1.0	барабанить	побарабанить	Я	--	побарабаню	--	--	--
3.3.2.0	барабанить	побарабанить	Ты	--	побарабанишь	побарабань	--	--
3.3.3.0	барабанить	побарабанить	Он/она	--	побарабанит	--	--	--
3.3.4.1	барабанить	побарабанить	Мы	--	побарабаним	побарабаним	--	--
3.3.4.2	барабанить	побарабанить	Мы	--	побарабаним	побарабанимте	--	--
3.3.5.0	барабанить	побарабанить	Вы	--	побарабаните	побарабаньте	--	--
3.3.6.0	барабанить	побарабанить	Они	--	побарабанят	--	--	--
3.4.0.1	барабанить	пробарабанить	--	--	--	--	--	пробарабанив
3.4.0.2	барабанить	пробарабанить	--	--	--	--	--	пробарабанивши
3.4.0.3	барабанить	пробарабанить	--	--	--	--	--	пробарабаня
3.4.1.0	барабанить	пробарабанить	Я	--	пробарабаню	--	--	--
3.4.2.0	барабанить	пробарабанить	Ты	--	пробарабанишь	пробарабань	--	--
3.4.3.0	барабанить	пробарабанить	Он/она	--	пробарабанит	--	--	--
3.4.4.1	барабанить	пробарабанить	Мы	--	пробарабаним	пробарабаним	--	--
3.4.4.2	барабанить	пробарабанить	Мы	--	пробарабаним	пробарабанимте	--	--
3.4.5.0	барабанить	пробарабанить	Вы	--	пробарабаните	пробарабаньте	--	--

3.4.6.0	барабанить	пробарабанить	Они	--	пробарабанят	--	--	--
4.0.0.0	басить	басить	--	--	--	--	бася	--

ЗАДАЧА II.

В период с июня 2022 года по март 2023 года была также успешно выполнена задача создания алгоритма поиска (выборки) каждой словоформы звуковых глаголов, зарегистрированной словаре Зализняка, и одновременного соединения результатов в **СВОДНОЙ ТАБЛИЦЕ** с результатами поиска в базе данных Национального корпуса русского языка. Также были отшлифованы принципы реализации интерактивной ссылки (многомерное шкалирование результатов веб-извлечения) для каждой словоформы в сводной таблице .

Таким образом, в **СВОДНОЙ ТАБЛИЦЕ** каждая словоформа *будущего и настоящего времени, повелительного наклонения*, а также *деепричастные* формы глаголов звучания сопровождаются **электронной сноской** на результаты поиска в НКРЯ. Исследователь нажимает на словоформу мышкой и попадает на соответствующую страницу поиска в **Национальном корпусе русского языка**. Все элементы, представленные в колонках C, E, F, D, H, I, снабжены электронными отсылками к Русскому Национальному Корпусу <https://ruscorpora.ru/>.

Пример 2. Колонка F, будущее время, словоформа **АХНУ**.

A	B	C	D	E	F	G
Number	Root	Word	Pronoun	Present	Future	Imperative
1.1.1.0	ахать	ахнуть	Я	--	АХНУ	--
1.1.2.0	ахать	ахнуть	Ты	--	ахнешь	ахни

Данная словоформа снабжена отсылкой к Национальному Корпусу русского языка <https://ruscorpora.ru/>:

ахну =HYPERLINK("https://ruscorpora.ru/explore?req=ахну","ахну")

При нажатии на слоформу **АХНУ** производится отсылка к следующим результатам поиска в НКРЯ:

The screenshot shows a web browser window with the URL <https://ruscorpora.ru/explore?req=ахну>. The page header includes the logo for the National Russian Language Corpus (НКРЯ) and navigation links for 'Меню' and 'ENG'. The main content area displays search statistics: '131 488 текстов • 374 449 975 слов' and '46 корпусов • 3 вида поиска'. The search query 'ахну' is entered in a search box. Below the search box, there are two main search options: 'Поиск точных форм в выбранном корпусе' (20 texts, 22 examples) and 'Лексико-грамматический поиск в корпусе' (1 text, 1 example). Both options have a 'Уточнить запрос' button. At the bottom, there is a 'Совет дня' section. The Windows taskbar at the bottom shows the date and time as 12:17 PM on 5/22/2023.

Обзор возможностей НКРЯ

https://ruscorpora.ru/explore?req=ахну

НАЦИОНАЛЬНЫЙ КОРПУС русского языка

Меню ENG

Корпус *Основной* • 131 488 текстов • 374 449 975 слов
омонимия снята, синтаксически размечен

Запрос **Вернуться к поиску** • 46 корпусов • 3 вида поиска "ахну"

Поиск точных форм в выбранном корпусе ?
• 20 текстов • 22 примера
Показать все примеры
Уточнить запрос

Лексико-грамматический поиск в корпусе ?
• 1 текст • 1 пример
Показать все примеры
Уточнить запрос

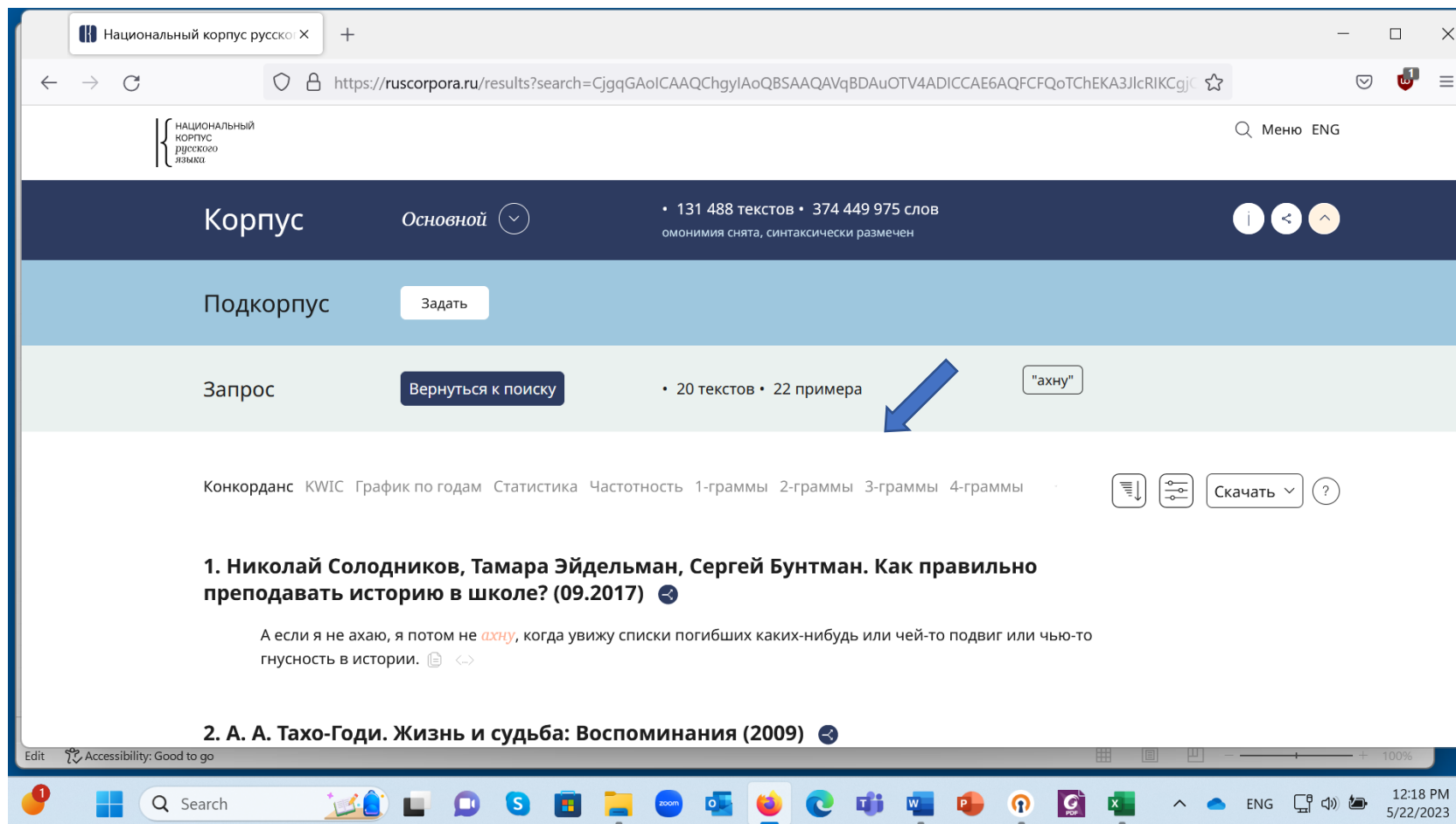
Совет дня

Edit Accessibility: Good to go

Search

12:17 PM 5/22/2023

При нажатии «Уточнить запрос» производится отсылка к количественным и качественным результатам поиска:



The screenshot shows a web browser window displaying the National Russian Corpus (НКРЯ) search results page. The browser's address bar shows the URL: <https://ruscorpora.ru/results?search=CjgqGAoICAAQChgylAoQBSAAQAVqBDAuOTV4ADICCAE6AQFCFQoTChEKA3JlcRIKGjC>. The page header includes the logo of the National Russian Corpus and the text "Корпус Основной" with statistics: "131 488 текстов • 374 449 975 слов". Below the header, there are sections for "Подкорпус" (Subcorpus) with a "Задать" (Set) button, and "Запрос" (Query) with a "Вернуться к поиску" (Return to search) button and the query "ахну". A blue arrow points to the "ахну" query. Below the query, there are navigation options: "Конкорданс", "KWIC", "График по годам", "Статистика", "Частотность", "1-граммы", "2-граммы", "3-граммы", "4-граммы", and a "Скачать" (Download) button. The search results are listed as follows:

- 1. Николай Солодников, Тамара Эйдельман, Сергей Бунтман. Как правильно преподавать историю в школе? (09.2017)**
А если я не ахаю, я потом не **ахну**, когда увижу списки погибших каких-нибудь или чей-то подвиг или чью-то гнусность в истории.
- 2. А. А. Тахо-Годи. Жизнь и судьба: Воспоминания (2009)**

Поскольку НКРЯ является закрытой системой, изменений статистических результатов поиска не предполагается. Трудностей доступа при помощи компьютерной программы (веб-извлечения) к Национальному корпусу русского языка не обнаружено.

ПРИЛОЖЕНИЯ

[Приложение 1](#). Ivliyeva, Koob. Appendix 1. Words not_in_morfologija-2023-05-18

[Приложение 2](#). Ivliyeva, Koob. Appendix 2. Double_pronoun_dual_action -2023-05-18

[Приложение 3](#). Ivliyeva, Koob. Ivliyeva, Koob. Appendix 3. Double_perfective_gerund-2023-05-18

[Приложение 4](#). Ivliyeva, Koob. Appendix 4. Double_imperfective_gerund-2023-05-18

Примечание. В приложениях нет электронных ссылок на НКРЯ. Обращайтесь к **СВОДНОЙ ТАБЛИЦЕ**.

ИССЛЕДОВАТЕЛЬСКАЯ ГРУППА, КОНТАКТЫ.

За подробностями о компьютерной программе (web-scraping algorithm) и по вопросам методологического характера просьба обращаться непосредственно к научно-исследовательской группе:

Ирина Ивлиева, ivliyeva@mst.edu, профессор кафедры искусств, иностранных языков и философии

Перри Куб, koobp@mst.edu, эксперт-администратор баз данных

Последние публикации авторов проекта по данной тематике (2019-2022)

Ivliyeva, Irina V., Koob, Perry. “Creating a corpus of all possible word forms of modified Russian sound verbs using web-scraping methodology. Compilation and adjustment of summary tables for the future tense, imperative, and gerund forms. Forms of “dual action”.

Experimental multi-dimensional scaling of web-scraping results.” (2022). [Electronic resource]. Research Data. March 2022. – Access mode: https://scholarsmine.mst.edu/research_data/9/

Ивлиева И.В. Развитие терминологии словообразовательного синтеза. Особенности терминосистемы (на материале глаголов звучания русского языка)// Интерактивная наука. – 2022. – № 6 (71). – С. 86 -94. – ISSN 2414-9411. DOI 10.21661/r-557017. [Electronic resource]. - Access mode: https://interactive-science.media/ru/article/557017/discussion_platform

Ивлиева И.В. Потенциальные единицы глагольного синтеза: свойства и возможности (на материале русских глаголов звучания) // Интерактивная наука. – 2021. – № 5 (60). – С. 11-19. – ISSN 2414-9411. [Electronic resource]. - Access mode: https://interactive-science.media/ru/article/554295/discussion_platform

Ivliyeva, I., Koob, Perry. “Method of web-extraction (web scraping) of Russian verb paradigms from electronic dictionaries and databases. Matrix organization of lacunae, their codification and classification (on the material of the verbs of sound)” (2021). [Electronic resource]. Research Data. 7. – Access mode: https://scholarsmine.mst.edu/research_data/7

Ivliyeva, Irina V., Koob, Perry. “Lacunae Matrices, Web Scraping. Theses.” (2021). Russian Linguistics Research. 1. [Electronic resource]. – Access mode: https://scholarsmine.mst.edu/russian_linguistics_research/1

Koob, Perry B., and Ivliyeva, Irina V. "Russian National Corpus Web Scraping Project 2019-2020." [Electronic resource]. February 1, 2020. – Access mode: https://scholarsmine.mst.edu/artlan_phil_facwork/158/

Ивлиева И.В. Особенности семантики лакун-модификаций. К идее создания словаря лакун (на материале глаголов звучания) // Интерактивная наука. – 2020. – № 6 (52). – С. 57-70. – ISSN 2414-9411. [Electronic resource]. - Access mode: https://interactive-science.media/ru/article/551671/discussion_platform

Ивлиева И.В. Роль имплицитных X-компонентов значения в синтезе русских глагольных модификаций (на примере глаголов механической фонации) // Интерактивная наука. 2019. – № 6 (40). С. 106-113. – ISSN 2414-9411. Access mode: https://interactive-science.media/en/article/497213/discussion_platform

ЭЛЕКТРОННЫЕ ЛЕКСИКОГРАФИЧЕСКИЕ РЕСУРСЫ

Грамматический словарь А.А. Зализняка. The grammatical dictionary of the Russian language by A.A. Zalyzniak, <https://morfologija.ru/>. Accessed May 18, 2023.

Национальный корпус русского языка. The Russian National Corpus, <https://ruscorpora.ru/>. Accessed May 18, 2023.