

ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА

руководство пользователя

v3.4

Мансур Сайхунов

Тавзих Ибрагимов

Jorma Luutonen

Содержание

[Что такое Письменный корпус татарского языка?](#)

[Зачем нужен корпус языка?](#)

[Как будет развиваться Письменный корпус татарского языка?](#)

[Системные требования](#)

[Как произвести поиск в Письменном корпусе татарского языка?](#)

[Как вернуться на страницу поиска?](#)

[Как получить прямую ссылку на определенную страницу с результатами поиска?](#)

[У Вас нет татарской раскладки клавиатуры?](#)

[Как найти в Корпусе словоформу и ее контекстную статистику?](#)

[Как найти в Корпусе отдельную лексему?](#)

[Как найти в Корпусе все лексемы, начинающиеся с определенного префикса?](#)

[Как найти в Корпусе все лексемы, заканчивающиеся определенным постфиксом?](#)

[Как найти в Корпусе словосочетание?](#)

[Система сложного морфологического поиска:](#)

[Что представляет из себя данная система?](#)

[Как найти в Корпусе комбинации с указанием словоформ \(словосочетания\)?](#)

[Как найти в Корпусе комбинации с указанием леммы?](#)

[Как найти в Корпусе комбинации с указанием морфологических \(грамматических\) тегов?](#)

[Как найти в Корпусе комбинации с указанием начала, середины и/или конца слова?](#)

[Как найти в Корпусе комбинации с указанием группы параметров для одного слова?](#)

[Ввод поисковых параметров в графическом режиме](#)

[Как прослушать найденные в Корпусе предложения?](#)

[Дополнительные статистические материалы!](#)

[Публикации](#)

[Описание](#)

[Использование и цитирование](#)

[О проекте](#)

[Благодарности](#)

[Контакты для Ваших предложений](#)

[На других языках](#)

Что такое Письменный корпус татарского языка?

Корпус представляет современный письменный татарский язык в электронной форме. Объем корпуса в настоящее время составляет более 116 млн. слов, число различных словоформ 1 млн. 500 тысяч единиц. При создании Корпуса использованы материалы, находящиеся в открытом доступе на различных WEB-ресурсах. Более половины текстов ($\approx 60\%$ от общего объема Корпуса) представляют публицистический стиль татарского языка. Остальные стили письменного литературного татарского языка содержатся в Корпусе примерно в следующем соотношении:

- стиль художественной прозы $\approx 35\%$,
- стиль научной литературы гуманитарного профиля $\approx 4\%$,
- стиль деловых бумаг $\approx 1\%$.

В качестве основных источников лингвистического материала были использованы тексты, представленные на наиболее распространенных WEB-сайтах, таких, например, как “Татар – информ”, “Ислам – татарлар Һәм мөселманнар” (Ислам – татары и мусульмане), “Татар электрон китапханәсе” (Татарская электронная библиотека), “Дуслык” (Дружба), “Азатлык” (Свобода), “Яңа гасыр” (Новый век), “Татар әдипләре” (Татарские писатели) и др.

Основное предназначение Письменного корпуса татарского языка – исследование лексико-семантической системы татарского языка в рамках статистической лексикологии и когнитивной лингвистики.

Корпус может использоваться в трех режимах:

- статистическом (поиск заданного слова, выделение его правого и левого контекстов, определение частоты употребления словоформ),
- лемматизации (благодаря морфологической разметке возможно выделение словоформ, относящихся к заданной лемме, и характерных для них различных грамматических категорий),
- поиска словоформ по шаблону (по начальной, средней и/или конечной части слова).

Кроме того, в поисковой системе корпуса предусмотрено озвучение предложений, предъявляемых в качестве примеров употребления заданного слова.

При необходимости корпус-менеджер может расширить правый и левый контексты заданного слова и тем самым увеличить точность в определении значения заданного слова или словосочетания.

Поисковая система Письменного корпуса позволяет проводить следующие операции:

- организовать поиск нужных слов, выявить частоту их употребления, находить примеры, подтверждающие употребление данного слова или данной формы в языке;

- определить, какие слова могут следовать впереди и за заданным словом (левый и правый контексты заданного слова), а также употребительность контекстно-связанных слов в сочетании с заданным словом.

Зачем нужен корпус языка?

При отсутствии стилистической дифференцированности (точнее в условиях отсутствия четких границ между указанными выше стилями) и наличии большого объёма, можно сказать, что Письменный корпус адекватно отражает современное состояние татарского языка. В многочисленных текстах, написанных в комфортных условиях и наиболее подготовленными людьми более, чем в устной форме языка раскрывается миропонимание не только авторов, но и читателей, а, следовательно, всей этнической общности и особенности самой эпохи, как памятника данного периода исторического развития общества.

Письменный корпус татарского языка будет представлять интерес для всех последующих поколений.

Как будет развиваться Письменный корпус татарского языка?

Один из видных представителей лингвистов XX века Л.Блумфилд свою книгу «Язык» заканчивает словами «И хотя сейчас это только мечта, можно надеяться, что в недалеком будущем изучение языка поможет нам понимать поступки людей и управлять этими поступками». Реализация этих замыслов потребует построения различного рода подкорпусов, электронных словарей, как обычных, так и словарей-тезаурусов. Появится необходимость расшифровки того, как языковая общность накапливает свой жизненный опыт, познает и преобразовывает мир, в котором она живет. Очевидно, что указанные потребности будут определять и развитие корпусной лингвистики.

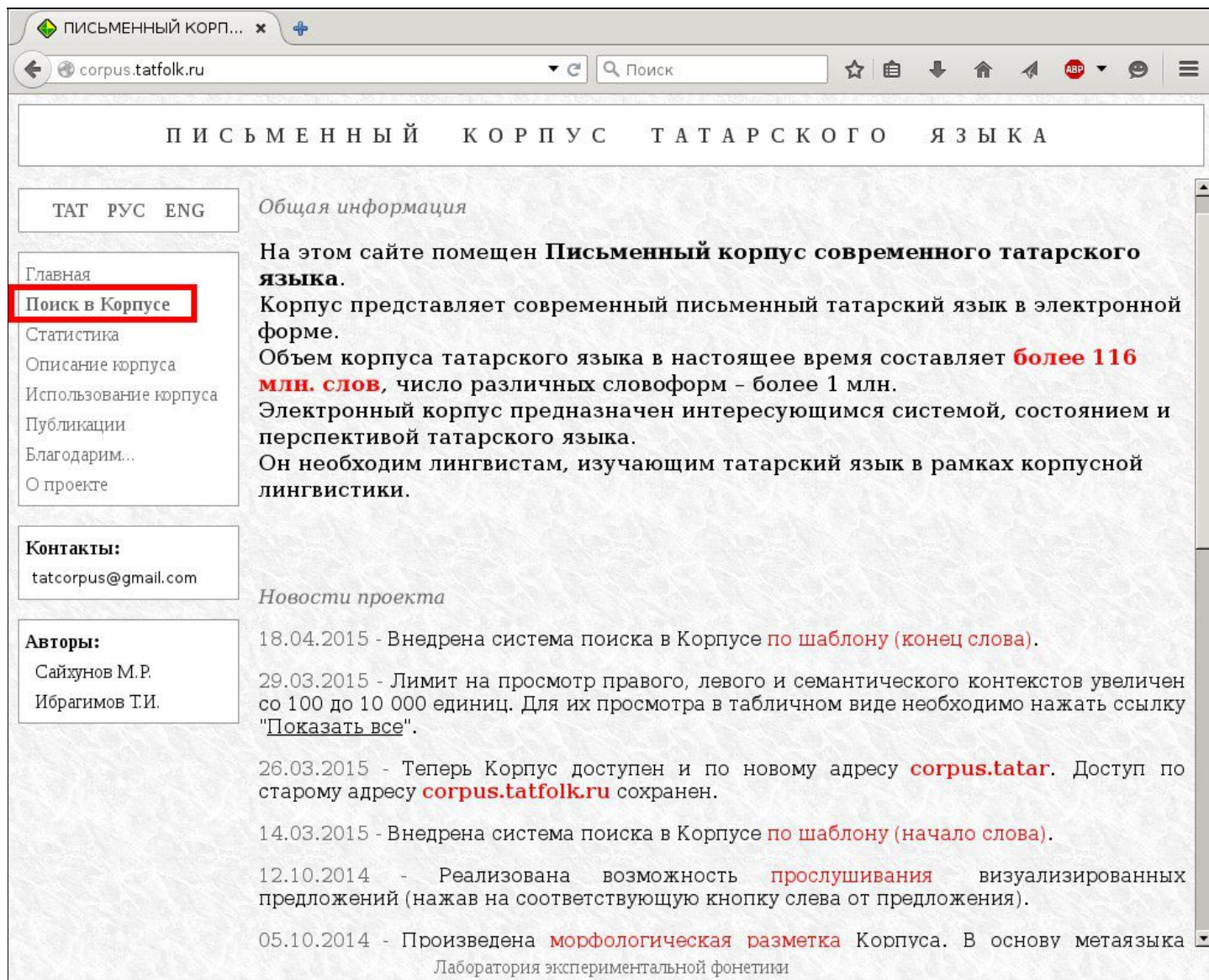
Системные требования

Для корректной работы с Корпусом рекомендуется:

1. Использовать актуальные версии таких браузеров, как:
 - Mozilla Firefox: <https://www.mozilla.org/en-US/firefox/all/#ru>
 - Google Chrome: <https://www.google.ru/intl/ru/chrome/browser/desktop/index.html>
2. Включить в браузере поддержку JavaScript.
3. Если имеются проблемы с воспроизведением предложений:
 - попробуйте использовать другой браузер,
 - установите актуальную версию Adobe Flash Player: <https://get.adobe.com/flashplayer/>

Как произвести поиск в Письменном корпусе татарского языка?

На главной странице сайта corpus.tatar (или corpus.tatfolk.ru) в меню слева найдите ссылку «Поиск в корпусе» и щелкните на ней мышкой:



The screenshot shows a web browser window with the address bar containing "corpus.tatfolk.ru". The page title is "ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА". On the left side, there is a navigation menu with the following items: Главная, Поиск в Корпусе (highlighted with a red box), Статистика, Описание корпуса, Использование корпуса, Публикации, Благодарим..., and О проекте. Below the menu, there are sections for "Контакты:" (tatcorpus@gmail.com) and "Авторы:" (Сайхунов М.Р., Ибрагимов Т.И.). The main content area is titled "Общая информация" and contains the following text: "На этом сайте помещен Письменный корпус современного татарского языка. Корпус представляет современный письменный татарский язык в электронной форме. Объем корпуса татарского языка в настоящее время составляет **более 116 млн. слов**, число различных словоформ - более 1 млн. Электронный корпус предназначен интересующимся системой, состоянием и перспективой татарского языка. Он необходим лингвистам, изучающим татарский язык в рамках корпусной лингвистики." Below this, there is a section titled "Новости проекта" with several entries: "18.04.2015 - Внедрена система поиска в Корпусе по шаблону (конец слова).", "29.03.2015 - Лимит на просмотр правого, левого и семантического контекстов увеличен со 100 до 10 000 единиц. Для их просмотра в табличном виде необходимо нажать ссылку 'Показать все'.", "26.03.2015 - Теперь Корпус доступен и по новому адресу **corpus.tatar**. Доступ по старому адресу **corpus.tatfolk.ru** сохранен.", "14.03.2015 - Внедрена система поиска в Корпусе по шаблону (начало слова).", "12.10.2014 - Реализована возможность прослушивания визуализированных предложений (нажав на соответствующую кнопку слева от предложения).", "05.10.2014 - Произведена морфологическая разметка Корпуса. В основу метаязыка Лаборатория экспериментальной фонетики".

После этого откроется страница «Поиск в Письменном корпусе татарского языка»:

ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА

ТАТ РУС ENG на главную

Главная
Поиск в Корпусе
Статистика
Описание корпуса
Использование корпуса
Публикации
Благодарим...
О проекте

Контакты:
tatcorpus@gmail.com

Авторы:
Сайхунов М.Р.
Ибрагимов Т.И.

Поиск в Письменном корпусе татарского языка

Выберите тип поиска и введите необходимое слово:

Поиск коллокаций в контекстном (статистическом) корпусе по словоформе (например, **китапны, авылларга, килмәдегез**)

Поиск в морфологически размеченном корпусе по лемме (например, **китап, авыл, кил**)

Поиск по шаблону (указывается начало или конец слова длиной 3 и более символов, например, **китап*, авыл*, *гез, *ләргә**)

Лаборатория экспериментальной фонетики

Как вернуться на страницу поиска?

После выполнения поискового запроса, для возвращения на страницу задания параметров поиска с сохранением ранее введенных данных можно использовать следующие ссылки:

The screenshot shows a web browser window with the URL `corpus.tatfolk.ru`. The page title is "ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА". There are navigation tabs for "ТАТ", "РУС", and "ENG". A search bar contains the text "[Вернуться на страницу поиска] [Перейти к предложениям]". The main content area displays the search results for the word "кутан".

ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА

ТАТ РУС ENG [Вернуться на страницу поиска] [Перейти к предложениям]

Главная
Поиск в Корпусе
Статистика
Описание корпуса
Использование корпуса
Публикации
Благодарим...
О проекте

Контакты:
tatcorpus@gmail.com

Авторы:
Сайхунов М.Р.
Ибрагимов Т.И.

СЛОВО: *кутан*

КОЛИЧЕСТВО СОВПАДЕНИЙ: 31866

ПРАВЫЙ КОНТЕКСТ: нәшриятында (1409; 19717.9), нәшрияты (1026; 14194.5), булып (871; 3345.21), авторы (869; 8785.8), итеп (787; 3081.54), " (628; 620.15), та (555; 2390.1), белән (476; 140.15), дөнья (407; 2037.03), » (392; 435.78), укырга (382; 2465.02), нәшриятының (379; 5576.17), укучылар (305; 1878.74), укып (304; 1943.77), - (303; 43.75), чыгару (262; 1630.57), күргәзмәсе (254; 2204.72), уку (252; 1002.89), нәшер (237; 1964.75), бүләк (204; 972.04), бастырып (196; 1450.99), турында (184; 205.99), бастыру (184; 1679.97), басылып (173; 1192.09), чыгарган (163; 1126.32), рәвешендә (160; 1110.37), чыгарырга (160; 1137.4), басу (159; 1292.96), укучы (157; 831.22), укый (150; 766.25), нәшриятына (149; 2191.66), алып (144; 133.16), " (143; 12.47), язган (142; 690.19), язып (142; 709.43), сөючеләр (134; 1248.08), — (132; 20.56), язарга (131; 853.71), укымый (127; 1443.4), чыкты (126; 399.29), кибете (126; 1205.3), нәшрияты (125; 2116.87), бар (123; 59.78), чыгара (117; 659.77), чыгарды (113; 743.19), яза (112; 496.18), язучу (105; 623.5), укуга (101; 951.18), язарлык (100; 1228.2), сәүдәсе (99; 1153.05), кибетләрендә (99; 1098.99), сатып (98; 307.17), - (97; 4.65), күргәзмәләре (96; 1068.96), кибетендә (94; 1000.95), тәкъдим (92; 169.96), кибетләре (91; 973.71), буларак (81; 153.79), кибетенә (80; 861.09), сүзе (74; 339.54), чыгарып (74; 347.04), фонды (70; 280.96), кына (69; 99.87), сату (65; 288.18), киштәсе (65; 920.46), килеп (63; 36.21), укыган (63; 286.33), өстендә (61; 220.12), авторлары (56; 416.63), киштәләрендә (56; 616.84), кебек (55; 14.85), киштәләре (55; 669.43), тәкъдир (53; 303.83), аша (50; 64.75), теле (50; 89.08), тотып (50; 137.34), кибетләренә (50; 582.79), алырга (49; 86.31), нәшрияте (49; 637.86), чыккан (48; 66.4), укучыларны (48; 277.7), тарату (48; 315.49), алам (46; 201.67), укыйм (46; 313.4), күргәзмә-ярминкәсе (45; 516.95), нәшр (43; 445.79), укуны (42; 332.42), яки (41; 34.96), укучыларга (40; 194.98), нәшриятлары (40; 490.78), укучыларнын (39; 169.41), язды (38; 206.85).

Лаборатория экспериментальной фонетики

Также, имеется ссылка и в самом низу страницы:

ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА

(источник: "Белем.ру" порталы (web-сайт)) [Найти текст](#)

TAT РУС ENG

Главная
Поиск в Корпусе
Статистика
Описание корпуса
Использование корпуса
Публикации
Благодарим...
О проекте

Контакты:
tatcorpus@gmail.com

Авторы:
Сайхунов М.Р.
Ибрагимов Т.И.

▶ Шулай ук телевизордан яңгыраган куркыныч хәбәрләргә йөрәккә якин алмаска, үзгә берәр мавыктыргыч шөгьль табарга, кызыклы **kitap** укырга, күбрәк жиләк-жимеш һәм шоколад ашарга кушалар.
(источник: "Татарстан яшьләре" газетасы (web-сайт)) [Найти текст](#)

▶ Күрәсең, безнең халыкка **kitap** байлыгы түгел, ә башка байлык кыйммәтрәк, чөнки бик кыйммәт, шундый затлы китаплар ун сум, шуның кадәр төшерделәр **kitap** бәясен, китапның бәясе төшәргә тиеш түгел, чөнки ул бик күп кешенең хезмәте.
(источник: "Татар-информ" МА Татарстан Республикасы мәгълумат агентлыгы (web-сайт)) [Найти текст](#)

▶ **Kitap** бизәү өлкәсендә «Жайдак» үзен тиз таныта.
(источник: Максим Глухов. Эзләнүләр эжимеше)

▶ Башка кунакларны көткәндә, **kitap** сүзләрен сөйләләр иде.
(источник: "Азатлык Радиосы" (web-сайт)) [Найти текст](#)

▶ 1) Дүрт йөзгә якин фотосурәт төшергән фотографларның яшен һәм елъязманың Калининградта чыгарылуын да исәпкә алсаң, **kitap** шактый гына бәялегә охшап тора.
(источник: Гульшат Галиуллина. Татар теле: лексикология (таблицалар, анализ үрнәкләре, күнегүләр, сүзлекчә). КАЗАН - 2007)

< [Предыдущая](#) | [В начало](#) | [Следующая](#) >

[[Наверх](#)] [[Вернуться на страницу поиска](#)]

Лаборатория экспериментальной фонетики

Как получить прямую ссылку на определенную страницу с результатами поиска?

Пожалуйста, не пытайтесь в качестве ссылки на результаты поиска использовать данные из адресной строки браузера, потому что сайт Корпуса имеет фреймовую структуру и там отображается лишь ссылка на главную страницу. Для получения прямой ссылки имеется кнопка «**Прямая ссылка**» в самом верху страницы с результатами поиска:

ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА

TAT РУС ENG [Вернуться на страницу поиска] [Перейти к предложениям] **[Прямая ссылка]** на главную

СЛОВО: **kutan** ▶

КОЛИЧЕСТВО СОВПАДЕНИЙ: 31866

ПРАВЫЙ КОНТЕКСТ: нәшриятында (1409; 19717.9), нәшрияты (1026; 14194.5), булып (871; 3345.21), авторы (869; 8785.8), итеп (787; 3081.54), " (628; 620.15), та (555; 2390.1), белән (476; 140.15), дөнья (407; 2037.03), » (392; 435.78), укырга (382; 2465.02), нәшриятының (379; 5576.17), укучылар (305; 1878.74), укып (304; 1943.77), - (303; 43.75), чыгару (262; 1630.57), күргәзмәсе (254; 2204.72), уку (252; 1002.89), нәшер (237; 1964.75), бүләк (204; 972.04), бастырып (196; 1450.99), турында (184; 205.99), бастыру (184; 1679.97), басылып (173; 1192.09), чыгарган (163; 1126.32), рәвешендә (160; 1110.37), чыгарырга (160; 1137.4), басу (159; 1292.96), укучы (157; 831.22), укый (150; 766.25), нәшриятына (149; 2191.66), алып (144; 133.16), " (143; 12.47), язган (142; 690.19), язып (142; 709.43), сөючеләр (134; 1248.08), — (132; 20.56), язарга (131; 853.71), укымый (127; 1443.4), чыкты (126; 399.29), кибете (126; 1205.3), нәшрияты (125; 2116.87), бар (123; 59.78), чыгара (117; 659.77), чыгарды (113; 743.19), яза (112; 496.18), язу (105; 623.5), укуга (101; 951.18), язарлык (100; 1228.2), сәүдәсе (99; 1153.05), кибетләрендә (99; 1098.99), сатып (98; 307.17), - (97; 4.65), күргәзмәләре (96; 1068.96), кибетендә (94; 1000.95), тәкъдим (92; 169.96), кибетләре (91; 973.71), буларак (81; 153.79), кибетенә (80; 861.09), сүзе (74; 339.54), чыгарып (74; 347.04), фонды (70; 280.96), кына (69; 99.87), сату (65; 288.18), киштәсе (65; 920.46), килеп (63; 36.21), укыган (63; 286.33), өстендә (61; 220.12), авторлары (56; 416.63), киштәләрендә (56; 616.84), кебек (55; 14.85), киштәләре (55; 669.43), тәкъдир (53; 303.83), аша (50; 64.75), теле (50; 89.08), тотып (50; 137.34), кибетләренә (50; 582.79), алырга (49; 86.31), нәшрияте (49; 637.86), чыккан (48; 66.4), укучыларны (48; 277.7), тарату (48; 315.49), алам (46; 201.67), укыйм (46; 313.4), күргәзмә-ярминкәсе (45; 516.95), нәшр (43; 445.79), укуны (42; 332.42), яки (41; 34.96), укучыларга (40; 194.98), нәшриятлары (40; 490.78), укучыларның (39; 169.41), язды (38; 206.85), дөньясына (38; 214.69), нәшриятендә (38; 487.07), сөюче (38; 309.94), битләре (38; 318.18), киштәсендә (37; 467.97), битләренә (36; 322.77), бастыра (36; 229.48), палатасы (35; 151.32) [Показать все](#)

ЛЕВЫЙ КОНТЕКСТ: %^% (2870; 265.87), татарстан (2781; 16768.4), , (2522; 144.94), бу (1203; 2840.53), аерым (771; 4948.61), " (709; 742.16), бер (680; 1378.6), _NUMBER_ (660; 66.09), дигән (659; 2501.77), тупынла (502; 1398.4) « (365; 325.43) алегә (362; 926.18) яна (359; 865.88) исемле (350; 2094.36)

Лаборатория экспериментальной фонетики

В результате откроется специальное окно, где вы сможете увидеть и скопировать необходимую ссылку:

ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА

ТАТ РУС ENG [Вернуться на страницу поиска] [Перейти к предложениям] [Прямая ссылка] на главную

СЛОВО: **куман** ▶

КОЛИЧЕСТВО СОВПАДАНИЙ: 31866

ПРАВЫЙ

авторы (89), дөнья (40), укучылар (40), күргәзмәсе (40), бастырып чыгарган укучы (157), язган (142; 690.19), язып (142; 709.43), сөючеләр (134; 1248.08), — (132; 20.56), язарга (131; 853.71), укымый (127; 1443.4), чыкты (126; 399.29), кибете (126; 1205.3), нәшрият (125; 2116.87), бар (123; 59.78), чыгара (117; 659.77), чыгарды (113; 743.19), яза (112; 496.18), язу (105; 623.5), укуга (101; 951.18), язарлык (100; 1228.2), сәүдәсе (99; 1153.05), кибетләрендә (99; 1098.99), сатып (98; 307.17), - (97; 4.65), күргәзмәләре (96; 1068.96), кибетендә (94; 1000.95), төкъдим (92; 169.96), кибетләре (91; 973.71), буларак (81; 153.79), кибетенә (80; 861.09), сүзе (74; 339.54), чыгарып (74; 347.04), фонды (70; 280.96), кына (69; 99.87), сату (65; 288.18), киштәсе (65; 920.46), килеп (63; 36.21), укыган (63; 286.33), өстендә (61; 220.12), авторлары (56; 416.63), киштәләрендә (56; 616.84), кебек (55; 14.85), киштәләре (55; 669.43), төкъдир (53; 303.83), аша (50; 64.75), теле (50; 89.08), тотып (50; 137.34), кибетләренә (50; 582.79), алырга (49; 86.31), нәшрияте (49; 637.86), чыккан (48; 66.4), укучыларны (48; 277.7), тарату (48; 315.49), алам (46; 201.67), укыйм (46; 313.4), күргәзмә-ярминкәсе (45; 516.95), нәшр (43; 445.79), укуну (42; 332.42), яки (41; 34.96), укучыларга (40; 194.98), нәшриятлары (40; 490.78), укучыларның (39; 169.41), язды (38; 206.85), дөньясына (38; 214.69), нәшриятендә (38; 487.07), сөюче (38; 309.94), битләре (38; 318.18), киштәсендә (37; 467.97), битләренә (36; 322.77), бастыра (36; 229.48), палатасы (35; 151.32) Показать все

ЛЕВЫЙ КОНТЕКСТ: % ^% (2870; 265.87), татарстан (2781; 16768.4), , (2522; 144.94), бу (1203; 2840.53), аерым (771; 4948.61), " (709; 742.16), бер (680; 1378.6), _NUMBER_ (660; 66.09), дигән (659; 2501.77), түрүндә (509; 1398.4), « (365; 325.43), алегә (362; 926.18), яна (359; 865.88), исемле (350; 2094.36)

Лаборатория экспериментальной фонетики

The page at http://search.corpus.tatar:8083 says:

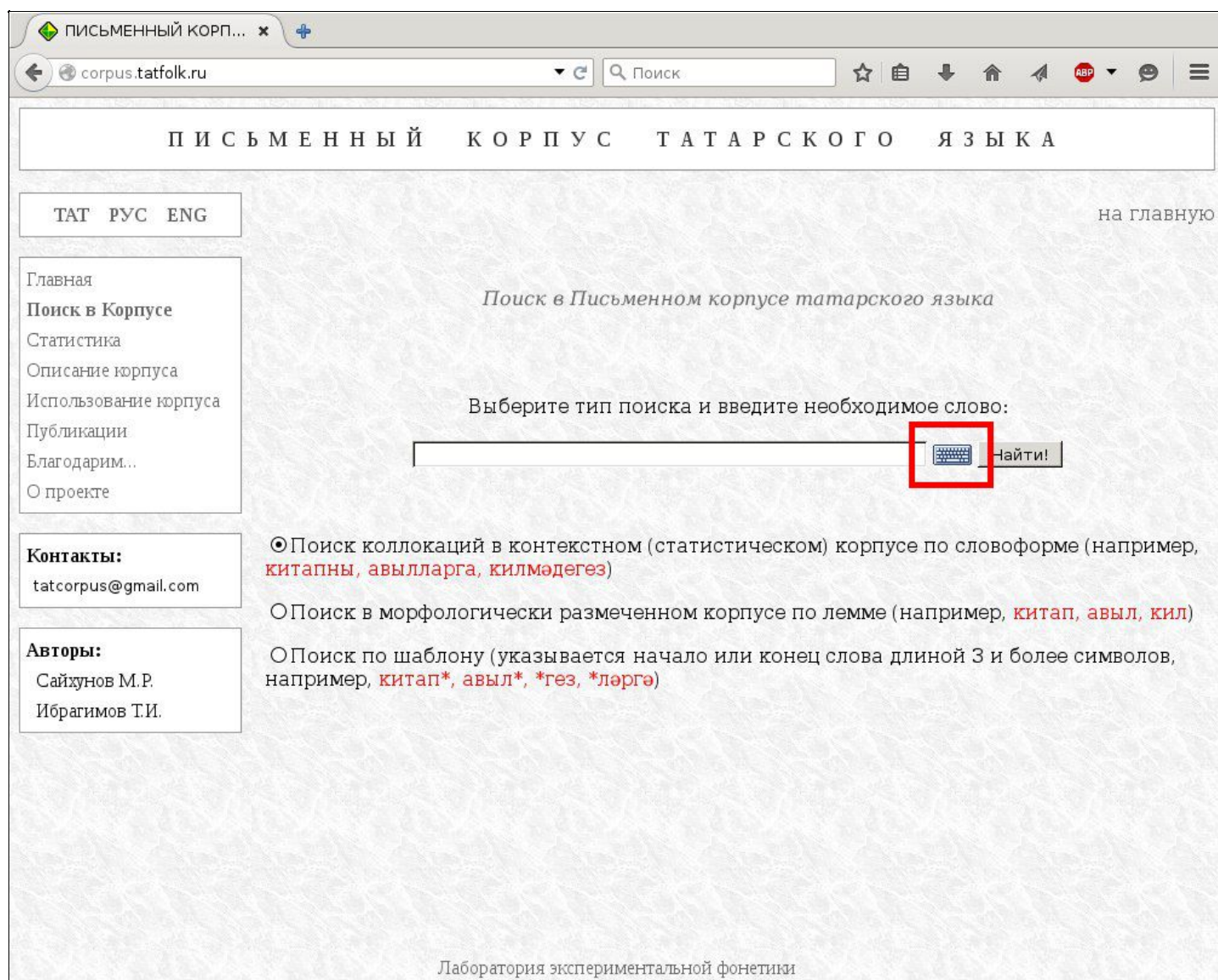
Прямая ссылка

</corpus/search/find.php?keyword=%D0%BA%D0%B8%D1%82%D0%B0%D>

Cancel OK

У Вас нет татарской раскладки клавиатуры?

Если у Вас по какой-то причине нет татарской раскладки клавиатуры, то в Корпусе на странице поиска Вы можете набирать нужное слово прямо с экрана. Для этого щелкните мышкой на специальном значке, расположенном между текстовым полем (где Вы печатаете слово) и кнопкой «**Найти!**»:



ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА

TAT РУС ENG на главную


Главная
Поиск в Корпусе
Статистика
Описание корпуса
Использование корпуса
Публикации
Благодарим...
О проекте

Контакты:
tatcorpus@gmail.com

Авторы:
Сайхунов М.Р.
Ибрагимов Т.И.

Поиск в Письменном корпусе татарского языка

Выберите тип поиска и введите необходимое слово:



Поиск коллокаций в контекстном (статистическом) корпусе по словоформе (например, **китапны, авылларга, килмәдегез**)

Поиск в морфологически размеченном корпусе по лемме (например, **китап, авыл, кил**)

Поиск по шаблону (указывается начало или конец слова длиной 3 и более символов, например, **китап*, авыл*, *гез, *ләргә**)

Лаборатория экспериментальной фонетики

После этого перед Вами появится **виртуальная клавиатура**, на которой Вы можете, щелкая мышкой на соответствующих буквах или же просто нажимая кнопки на своей клавиатуре, набрать нужное татарское слово:

The screenshot shows a web browser window with the URL `corpus.tatfolk.ru`. The page title is "ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА". On the left, there is a navigation menu with options: Главная, Поиск в Корпусе, Статистика, Описание корпуса, Использование корпуса, Публикации, Благодарим..., and О проекте. Below the menu are sections for "Контакты:" (tatcorpus@gmail.com) and "Авторы:" (Сайхунов М.Р., Ибрагимов Т.И.). The main content area features a search interface with the heading "Поиск в Письменном корпусе татарского языка" and the instruction "Выберите тип поиска и введите необходимое слово:". There are three radio button options for search types: "Поиск коллокаций в контекстном (статистическом) корпусе по словоформе (например, **китапны, авылларга, килмәдегез**)", "Поиск в морфологически размеченном корпусе по лемме (например, **китап, авыл, кил**)", and "Поиск...". A search input field and a "Найти!" button are also present. Overlaid on the bottom of the browser window is a "Virtual Keyboard - Mozilla Firefox" window. This virtual keyboard is configured for the "Tatar" language and includes keys for Cyrillic characters (ё, һ, 1, 2, 3, 4, 5, 6, 7, 8, 9, 0, -, =, \, Bk.sp), Latin characters (й, ц, е, у, к, е, н, г, ш, щ, ә, з, х, ь, ү), and other Tatar-specific characters (ф, ы, в, а, п, р, о, л, д, ң, э, я, ч, с, м, и, т, ь, ж, <, >, ю, ., '). The keyboard also features standard function keys like Tab, CapsLock, Shift, Ctrl, and Alt. The version number "VirtualKeyboard 3.7.2.797" is visible at the bottom left of the keyboard window.

Как найти в Корпусе словоформу и ее контекстную статистику?

В окне поиска в текстовом поле наберите нужную словоформу (например, **китапның**), убедитесь, что чуть ниже «галочкой» отмечен пункт «Поиск коллокаций в контекстном (статистическом) корпусе по словоформе...» и щелкните на кнопке «Найти!»:

ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА

TAT РУС ENG на главную

Главная
Поиск в Корпусе
Статистика
Описание корпуса
Использование корпуса
Публикации
Благодарим...
О проекте

Выберите тип поиска и введите необходимое слово:

Поиск коллокаций в контекстном (статистическом) корпусе по словоформе (например, **китапны, авылларга, килмәдегез**)

Поиск в морфологически размеченном корпусе по лемме (например, **китап, авыл, кил**)

Поиск по шаблону (указывается начало или конец слова длиной 3 и более символов, например, **китап*, авыл*, *гез, *ләргә**)

Контакты:
tatcorpus@gmail.com

Авторы:
Сайхунов М.Р.
Ибрагимов Т.И.

Лаборатория экспериментальной фонетики

В отдельном окне откроются результаты поиска. Здесь в самом начале указывается сама искомая словоформа и количество ее совпадений в Корпусе. Далее следуют три раздела:

- **Правый контекст** - список слов, отсортированных по частотности в порядке убывания, представленных в Корпусе в позиции справа от искомого слова. Т.е. это те слова, которые находятся в предложениях после того слова, которое ищем. В скобках рядом с каждым словом даны два числа:

1. Число, выделенное **оранжевым** цветом, обозначает количество повторений в Корпусе данного сочетания слов (искомое **китапның** + текущее слово из правого контекста, например, **эчтәлеген**). В данном случае, получается, что словосочетание «**китапның эчтәлеген**» в Корпусе встречается 26 раз.

2. Число, выделенное **зеленым** цветом, обозначает так называемый **показатель логарифмического правдоподобия** для текущего словосочетания в Корпусе.

Подробнее об этом можно почитать в личном блоге Теда Даннинга по адресу:

<http://tdunning.blogspot.ru/2008/03/surprise-and-coincidence.html>

ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА

ТАТ РУС ENG

СЛОВО: **китапның**
КОЛИЧЕСТВО СОВПАДЕНИЙ: 3120

ПРАВЫЙ КОНТЕКСТ: авторы (146; 1625.81), _NUMBER_ (144; 134.06), беренче (119; 579.32), икенче (86; 475.46), исеме (73; 567.74), төп (56; 282.71), исем (55; 453.97), бер (52; 84.03), соңгы (50; 232.22), эчтәлегә (44; 476.6), ахырында (42; 342.94), « (39; 39.47), һәр (39; 152.04), кереш (37; 400.51) “ (30; 3.95), иң (29; 59.37), авторлары (28; 299.1), тиражы (26; 323.12) **эчтәлеген (26; 285.34)**, мөхәррире (26; 226.15), исемен (22; 139.32), эчтәлегенә (21; 262.48), төзүчесе (20; 259.23), баш (17; 47.7), кадрен (16; 134.95), титул (16; 222.34), фәнни (15; 71.71), язылу (14; 117.73), әһәмияте (14; 113.48), дөньяга (13; 62.35), тышлыгы (13; 190.91), тышлыгында (13; 192.15), яна (13; 10.95), тагын (13; 19.43), электрон (12; 62.8), кыйммәте (12; 122.25), редакторы (12; 109.51), ахырына (12; 77.81), рус (12; 40.54), тулыландырылган (11; 145.04), басылып (11; 66.24), үз (11; 4.91), эченә (11; 58.35), башында (11; 48.32), дөнья (11; 28.38), нинди (11; 21.16), берничә (11; 26.21), зурлыгы (11; 110.07), дөвамы (10; 89.46), азагында (10; 69.35), ике (10; 7.43), эчендә (9; 23.93), аерым (9; 21.22), беренчесе (9; 61.76), бәясә (9; 45.17), күп (9; 4.68), тарихы (9; 38.27), үзен (9; 30.2), тышлыгын (9; 143.19), исеменә (9; 52.86), өченчә (9; 31.36), дәрәжәсен (8; 48.51), рәссамы (8; 66.54), үзенә (8; 23.84), чыгуы (8; 57.27), тышлыгына (7; 98.96), идея (7; 47.72), буеннан-буена (7; 78.26), илаһи (7; 49.64), тулы (7; 19.42), тышына (7; 80.96), максаты (7; 28.99), кулъязмасы (7; 74.51), тәржемәсен (7; 82.86), концепциясе (7; 61.05), дүртенчә (7; 35.75), ролен (6; 34.36), алгы (6; 40.17), берсен (6; 33.5), тәүге (6; 28.69), битләрен (6; 57.76), бәясен (6; 41.32), нигезенә (6; 48.04), урыны (6; 21.71), җаваплы (6; 23.14), нигезен (6; 43.45), авырлыгы (6; 44.28), барлык (6; 7.04), калган (6; 9.22), киңәйтелгән (5; 36.03), битләре (5; 44.65), тәржемәсе (5; 42.59), архив (5; 34.78), озакламый (5; 31.25), тышында (5; 56.4), баштагы (5; 41.35), кырыена (5; 45.43), россиягә (5; 34.23), эчтәлегендә (5; 61.1), рецензенты (5; 86) [Показать все](#)

ЛЕВЫЙ КОНТЕКСТ: %^% (903; 1453.07), бу (305; 1270.4), , (224; 4.69),

Лаборатория экспериментальной фонетики

- **Левый контекст** - то же самое, что и в предыдущем пункте, только подразумевается левое слово, т.е. которое в предложениях идет перед искомым словом!
- **Семантический контекст** - список слов, отсортированных по частотности в порядке убывания, представленных в Корпусе в одном предложении с искомым словом. Т.е. это те слова, которые находятся в том же предложении, где и искомое слово. В этом случае данное и искомое слово могут как стоять непосредственно рядом, так и быть разделены другими словами. Этот раздел представляет различные семантические связи, которые обусловлены нахождением слов в контексте одного предложения.

ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА

TAT РУС ENG

Главная
 Поиск в Корпусе
 Статистика
 Описание корпуса
 Использование корпуса
 Публикации
 Благодарим...
 О проекте

Контакты:
 tatcorpus@gmail.com

Авторы:
 Сайхунов М.Р.
 Ибрагимов Т.И.

ЛЕВЫЙ КОНТЕКСТ: %^% (903; 1453.07), бу (305; 1270.4), , (224; 4.69), изге (150; 1333.51), әлеге (119; 576.55), ул (69; 103.51), дигән (47; 150.31), исемле (30; 171.36), яңа (27; 52.82), һәр (24; 71.81), шул (24; 33.4), бер (22; 9.61), чыккан (19; 72.8), ук (18; 19.23), булачак (17; 52.69), турындагы (16; 65.46), ә (16; 8.43), аталган (12; 62.09), шушы (12; 21.82), беренче (12; 10.21), икенче (12; 22.34), торган (11; 8.26), әһле (11; 128.83), берничә (11; 26.21), әмма (11; 15.16), язылган (10; 37.13), кулъязма (10; 93.08), күргән (9; 39.25), багышланган (9; 26.72), автор (9; 46.24), басма (9; 60.61), өч (9; 17.26), борынгы (8; 31.92), ителгән (7; 16.58), мондый (7; 9.98), тик (7; 7.75), чөнки (7; 8.42), миңа (7; 12.19), исемнәре (6; 33.04), басылган (6; 32.22), тарихи (6; 17.88), алган (6; 7.74), алынган (6; 14.72), кадерле (6; 27.01), шигърьләр (5; 20.66), торучы (5; 20.74), томлы (5; 51.84), кызыл (5; 13.67), киләчәктә (5; 13.1), текст (5; 34.75), томлык (4; 34.99), биографик-публицистик (4; 86.32), исемләнгән (4; 17.56), фәнни (4; 9.3), хәтта (4; 4.64), исеме (4; 8.84), андый (4; 9.43), кызыксыну (3; 11.72), кыскача (3; 16), битлек (3; 22.31), бүленгән (3; 15.88), тормышында (3; 12.86), мәшһүр (3; 12.91), тышлы (3; 27.7), уникаль (3; 15.58), калын (3; 15.57), ошбу (3; 16.82), данә (3; 16.46), ушбу (3; 29.99), аңлатмалар (3; 24.08), электрон (3; 7.87), күңелле (3; 6.62), яза (3; 6.16), * (3; 5.43), әлеге (3; 4.41), жиде (3; 8.44), һәрбер (3; 8.87), теләгән (3; 9.18), гадәти (3; 10.6), моны (3; 4.01), әсәрләрдә (3; 23.57), ләп (3; 21.02), чыгарылган (3; 11.69), кирәкле (3; 7.54), битле (2; 14.27), мөхәммәтшиннар (2; 27.08), шәжәрәсе (2; 16.98), көшифәгә (2; 34.78), битенә (2; 12.49), жәмәгатьчеләгендә (2; 20.1), исемдәге (2; 12.57), көйле (2; 15.02), авторлар (2; 11.45), габидулла (2; 24.94), жавапны (2; 11.57), хикмәтле (2; 13), бәлкәм (2; 10.89), бар—алары (2; 43.16), әзерләнүче (2; 17.86), фикерләре (2; 12.84) [Показать все](#)

СЕМАНТИЧЕСКИЙ КОНТЕКСТ: . (2934; 244.58), , (1741; 50.28), һәм (713; 187.99), _NUMBER_ (586; 39.54), бу (536; 338.75), да (431; 23.99), белән (417; 9.5), « (378; 334.19), » (353; 313.95), - (336; 24.63), дә (330; 17.62), " (318; 24.2), " (294; 22.12), ул (294; 24.14), дип (281; 39.63), : (260; 73.1), бер (254; 65.54), (228; 142.83),) (223; 128.39), беренче (215; 249.58), турында (209; 179.09), татар (201; 116.14), авторы (201; 1294.23), әлеге (190; 179.29), дигән (187; 193.42), анын (171; 33.32), изге (164; 623.39), - (149; 19.47).

Лаборатория экспериментальной фонетики

Чуть ниже следующим разделом являются «Примеры»:

ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА

TAT РУС ENG

Главная
Поиск в Корпусе
Статистика
Описание корпуса
Использование корпуса
Публикации
Благодарим...
О проекте

Контакты:
tatcorpus@gmail.com

Авторы:
Сайхунов М.Р.
Ибрагимов Т.И.

ПРИМЕРЫ:

▶ Элеге уңайдан Бөтендөнъя татар конгрессы Башкарма комитеты рәисе урынбасары Ренат Вәлиуллин **китапның** тиздән сатуга чыгачагын, шулай ук республикадан читтә дә таратылачагын белдерде.
(источник: "Татар-информ" МА Татарстан Республикасы мәгълүмат агентлыгы (web-сайт)) [Найми текст](#)

▶ Шуңа күрә дә китабымның изге максаты - жыйналган мәгълүматларны, тарихи чыганаclar белән тулыландырып, халкымның үзенә кайтару». 3 мең данә тираж белән нәшер ителгән **китапның** тышлыгын рәссам Рәисә Сәйфуллина бизәгән.
(источник: "Татар-информ" МА Татарстан Республикасы мәгълүмат агентлыгы (web-сайт)) [Найми текст](#)

▶ Дүрт тематик кисәккә бүленгән **китапның** эпиграфлары ук игътибарны туплый, кызыксынуны көчәйтә.
(источник: РИФӘ РАХМАН. Татар телендә 100 сочинение)

▶ **Китапның** икенче басмасы алдагы басмада жибәрелгән күп кенә житешсезлекләрдән, авыр сүзләрдән азат ителгән.
(источник: "Фән һәм тел" журналы (№27))

▶ Профессор Г. Ф. Саттаров "Казанчы, Сабанчы, Уракчы, Урманчы, Ындырчы, Буранчы, Туйчы, Ямгурчы һ. б. — борынгы төрки татар кешесе исемнәре" дип яза алда телгә алынган хезмәтендә (85нче бит.) "Башкортстанда феодализм һәм капитализм тарихыннан" дигән **китапның** (Уфа, 1971 ел) 267нче битендә "Слобоцкие татаровя без хлебного жалованья Иткул Бехтемиров, Сабанча Чилпанов..." дигән юлларны укыйбыз.
(источник: "Кызыл таң" газетасы (web-сайт)) [Найми текст](#)

Лаборатория экспериментальной фонетики

Здесь приводится список из примерно 50 предложений, в которых употребляется искомая словоформа. Для удобства данная словоформа выделяется красным цветом.

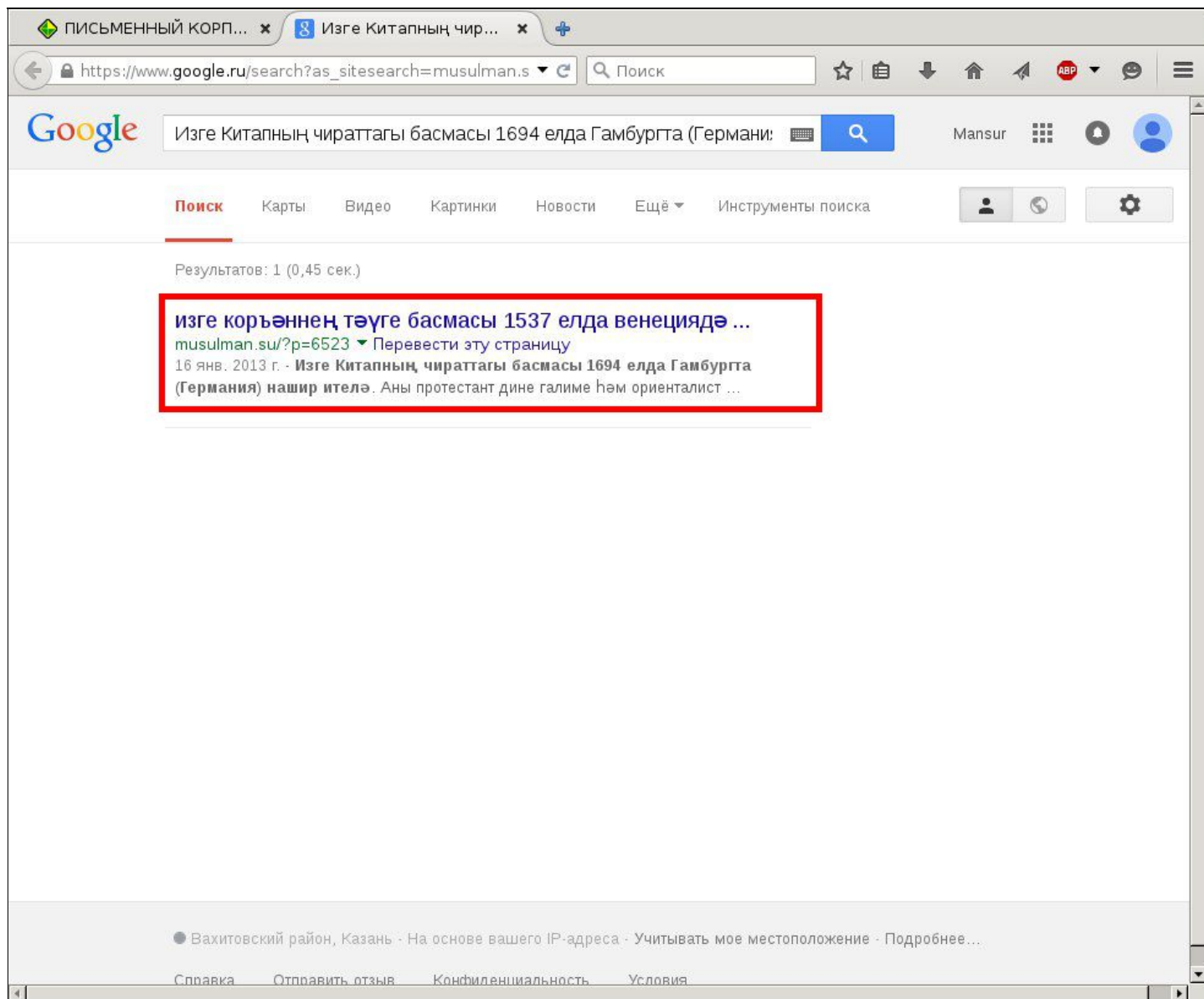
После каждого предложения мелким шрифтом указывается источник, откуда данное предложение было взято.

Если источником является какой-либо веб-ресурс (все помеченные как «**web-сайт**»), то приводится ссылка «**Найти текст**».

The screenshot shows a web browser window with the URL 'corpus.tatfolk.ru'. The page title is 'ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА'. On the left, there is a navigation menu with options like 'Главная', 'Поиск в Корпусе', and 'Статистика'. The main content area displays search results for the word 'китапның'. The second result is highlighted with a red box: 'Изге **Китапның** чираттагы басмасы 1694 елда Гамбургта (Германия) нашир ителә. (источник: "Ислам - Татарлар һәм мөселманнар" (web-сайт)) [Найти текст](#)'. Other results include 'Хәлбуки, табыг иттерү ноктасыннан **китапның** озын булуы матлуб түгелдер.', 'Өйгә эш катлаулашты: әкиятне укып, үзеңә ошаган фрагментын сүрәткә төшерергә, һәм **китапның** ни хақында булуын бер жөмлә белән генә язып куярга кирәк.', 'Китапның әһәмиятен якташыбыз, тарих фәннәре кандидаты, Казан федераль (педагогика) университеты профессоры, ТР Мәгълүматлаштыру академиясенең действительный әгъзасы Иршат Гафаров һәм чуваш төбәкләрен өйрәнү союзының мактау-лы рәисе Виталий Станьял югары бәяләделәр.', and '"Сагыну" дигән **китапның** кыйбласы менә шул тарафта.'

Функция поиска оригинала текста не представлена для художественной литературы из-за ограничений, связанных с авторским правом.

Нажав на данную ссылку, пользователь будет перенаправлен на сайт поисковой системы **Google**:



И, в большинстве случаев, первая же ссылка будет указывать на страницу с текстом, содержащим необходимое предложение:

The screenshot shows a web browser window with the address bar containing 'musulman.su/?p=6523'. The main content area displays the title 'ИЗГЕ КОРЪЭННЕҢ ТӘҮГЕ БАСМАСЫ 1537 ЕЛДА ВЕНЕЦИЯДӘ ДӨНЬЯ КҮРҮЕ ТУРЫНДА ИШЕТКӘН ИДЕГЕЗМЕ?' and a search result snippet: 'Изге Китапның чираттагы басмасы 1694 елда Гамбургта (Германия) нашир ителә. Аны протестант дине галиме һәм ориенталист Абрахам Хинкельман әзерли.' This snippet is highlighted in blue. The browser's search bar at the bottom shows '(Германия) нашир ителә.' and '1-е из 1 совпадения'. The right sidebar contains sections for 'Эзләү', 'Радио', and 'Видео', with a video player showing a man speaking at a podium.

Однако предложение здесь цветом не выделяется. Пользователь может его найти, воспользовавшись функцией поиска веб-браузера (нажав одновременно на клавиатуре **Ctrl** и **F**) или просто просмотрев текст.

В нижней части списка предложений имеются ссылки для перехода на другие страницы, т.е. пользователь, кликнув мышкой на ссылку «Следующая» перейдет для просмотра следующих 50 предложений и т.д., пока не будут показаны все предложения с употреблением искомого слова.

The screenshot shows a web browser window with the address bar displaying 'corpus.tatfolk.ru'. The page title is 'ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА'. The main content area displays a list of search results for the word 'Китапның'. Each result is preceded by a play button icon and includes a short text snippet, a source citation, and a link to the full text ('Найти текст').

Navigation links are located at the bottom of the page, enclosed in a red box: '< Предыдущая | В начало | Следующая >'. Below these links, the text 'Лаборатория экспериментальной фонетики' is visible.

ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА

TAT РУС ENG

Главная
Поиск в Корпусе
Статистика
Описание корпуса
Использование корпуса
Публикации
Благодарим...
О проекте

Контакты:
tatcorpus@gmail.com

Авторы:
Сайхунов М.Р.
Ибрагимов Т.И.

▶ **Китапның** авторы — алдынгы фикерле гарәп (Мисыр) язучысы (чыгышы белән көрд) Касыйм Әмин (1865—1909).
(источник: "Габдулла Тукайга багышланган Интернет-портал" (web-caйт)) [Найти текст](#)

▶ **Һәм шушы китапның** дөнья күрүендә Сезнең ярдәм итүегезне сорыйбыз.
(источник: "Туганайлар" газетасы (web-caйт)) [Найти текст](#)

▶ **Китапның** беренче өлеше ел фасыллары белән бәйлә горөф-гадәтләрне үз эченә алса, икенче өлешендә гаилә традицияләре каләмгә алынган.
(источник: "Белем.ру" порталы (web-caйт)) [Найти текст](#)

▶ **Китапның** беренче кисәгендә 30-35 документ файдаланылган.
(источник: РИЗАЭДДИН ФӘХРЕДДИН: МИРАСЫ ҺӘМ ХӘЗЕРГЕ ЗАМАН: Мөкаләләр әҗьентыгы. Казан, 16 ноябрь 1999 ел. - Казань, 2003.)

▶ **Китапның** кулъязма вариантларында һәм беренче (1847) басмасында Т. Ялчыголның үзе иҗат иткән байтак шигъри юллар бар.
(источник: "Татар әдипләре" порталы (web-caйт)) [Найти текст](#)

▶ **Өч китапның** да үзәген кеше гомере, күңелдәге кичерешләр һәм дөнья турында уйланулар тәшкит иткәнлектән алар бер-берсен тулыландыралар.
(источник: "Татар әдипләре" порталы (web-caйт)) [Найти текст](#)

▶ **Китапның** төп өлешен архив документлары, ягъни ревизия материаллары алып тора да.
(источник: "Азатлык Радиосы" (web-caйт)) [Найти текст](#)

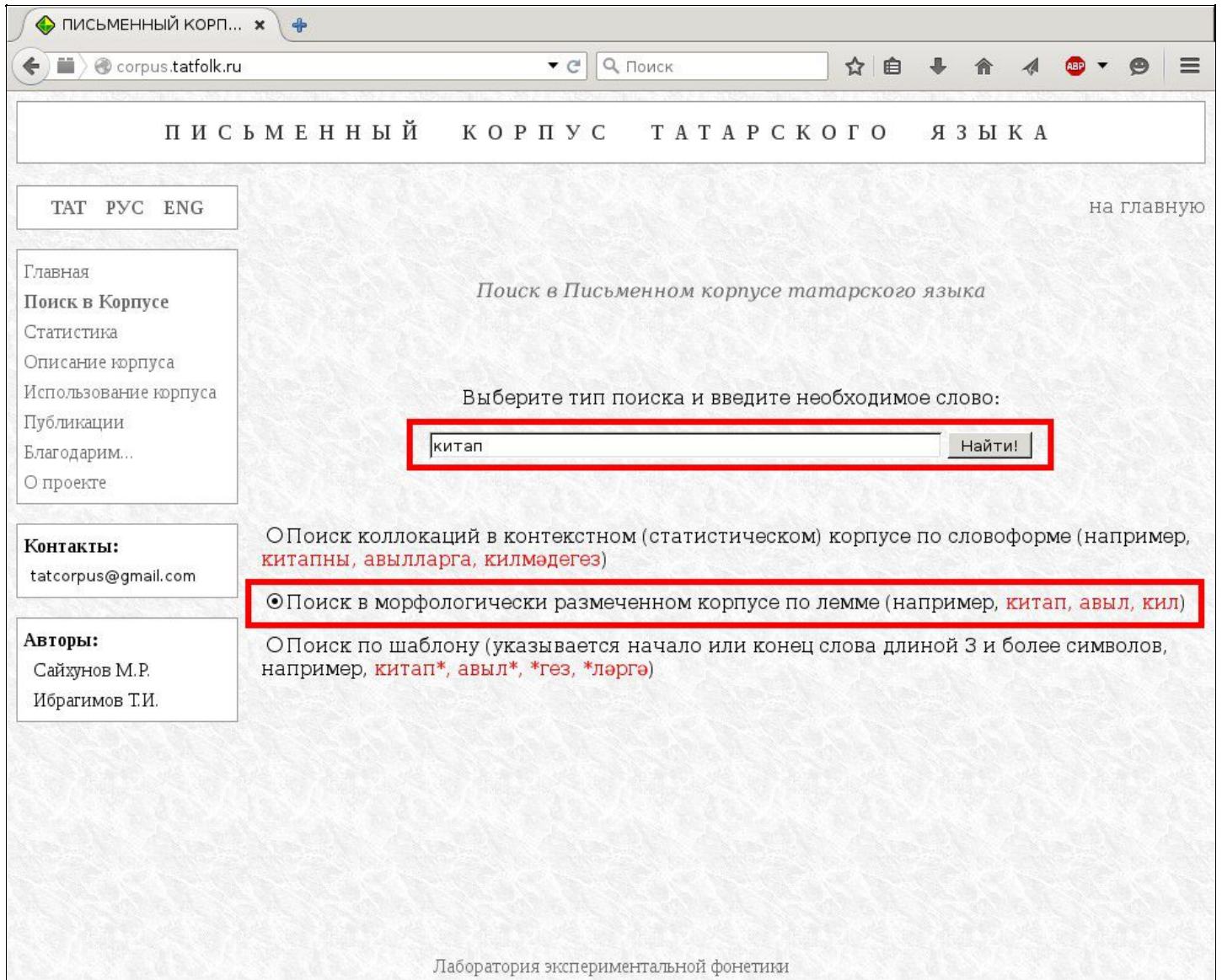
< Предыдущая | В начало | Следующая >

Лаборатория экспериментальной фонетики

Как найти в Корпусе отдельную лексему?

Благодаря наличию морфологической разметки Корпуса, является возможным поиск различных словоформ указанной лексемы.

Для этого на странице Поиска в текстовом поле необходимо набрать лемму (например, *тормыш, тутыр, матур*) и отметить «галочкой» пункт «Поиск в морфологически размеченном корпусе по лемме...».



П И С Ъ М Е Н Н Ы Й К О Р П У С Т А Т А Р С К О Г О Я З Ы К А

TAT RUS ENG на главную

Главная
Поиск в Корпусе
Статистика
Описание корпуса
Использование корпуса
Публикации
Благодарим...
О проекте

Поиск в Письменном корпусе татарского языка

Выберите тип поиска и введите необходимое слово:

Поиск коллокаций в контекстном (статистическом) корпусе по словоформе (например, *китапны, авылларга, килмәдегез*)

Поиск в морфологически размеченном корпусе по лемме (например, *китап, авыл, кил*)

Поиск по шаблону (указывается начало или конец слова длиной 3 и более символов, например, *китап*, авыл*, *гез, *ләргә*)

Контакты:
tatcorpus@gmail.com

Авторы:
Сайхунов М.Р.
Ибрагимов Т.И.

Лаборатория экспериментальной фонетики

После нажатия кнопки «Найти!» появится страница со списком предложений, в которых искомая лемма представлена в различных грамматических формах.

The screenshot shows a web browser window with the URL corpus.tatfolk.ru. The page title is "ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА". The search results are for the lemma "китап".

Navigation: TAT РУС ENG

ЛЕММА: **китап**

Left sidebar menu:

- Главная
- Поиск в Корпусе
- Статистика
- Описание корпуса
- Использование корпуса
- Публикации
- Благодарим...
- О проекте

Contact: tatcorpus@gmail.com

Authors: Сайхунов М.Р., Ибрагимов Т.И.

Search results:

- ▶ Фатирыннан чыгып киткәндә, шагыйрь миңа “Таңнар, кичеләр” дигән **китабын** бүләк итте.
(источник: Бари ИСЛАМ. БИК ГАДИ ДӘ, СЕРЛЕ ДӘ СИН, ТОРМЫШ)
- ▶ Вахтаны кабул итеп алганнан соң, өстәл артына утырып йоклап китмәс өчен бер детектив **китап** укый башлады.
(источник: Бари ИСЛАМ. БИК ГАДИ ДӘ, СЕРЛЕ ДӘ СИН, ТОРМЫШ)
- ▶ Болар хакында 1916 елда Уфада чыккан “Башкиры” дигән **китапта** урыс этнографы яза.
(источник: Илдус Хужжин. Топонимнар)
- ▶ Казан галимнарыннан Д.Рамазанованың "Формирование татарских говоров юго-западной Башкирии" дигән **китабында** да күп "башкорт" авылларының казан татарларынан формалашканы күренә.
(источник: Илдус Хужжин. Топонимнар)
- ▶ "... А.Әсфәндияровның “История Башкирских сел Пермской и Свердловской областей” дигән **китабы** буенча, безнең ата-бабаларыбыз Башкорт жиреннән чыкканнар, чөнки Әртә районы жирләре элек Башкортостанга кергән булган.
(источник: Илдус Хужжин. Топонимнар)
- ▶ Бу авыл халкы (алар таптәрләр дип язылганнар) 1765 елда Сыскы вулысы башкортларынан жир алганнар дип языла “Западные башкиры” дигән **китапта**.
(источник: Илдус Хужжин. Топонимнар)

Лаборатория экспериментальной фонетики

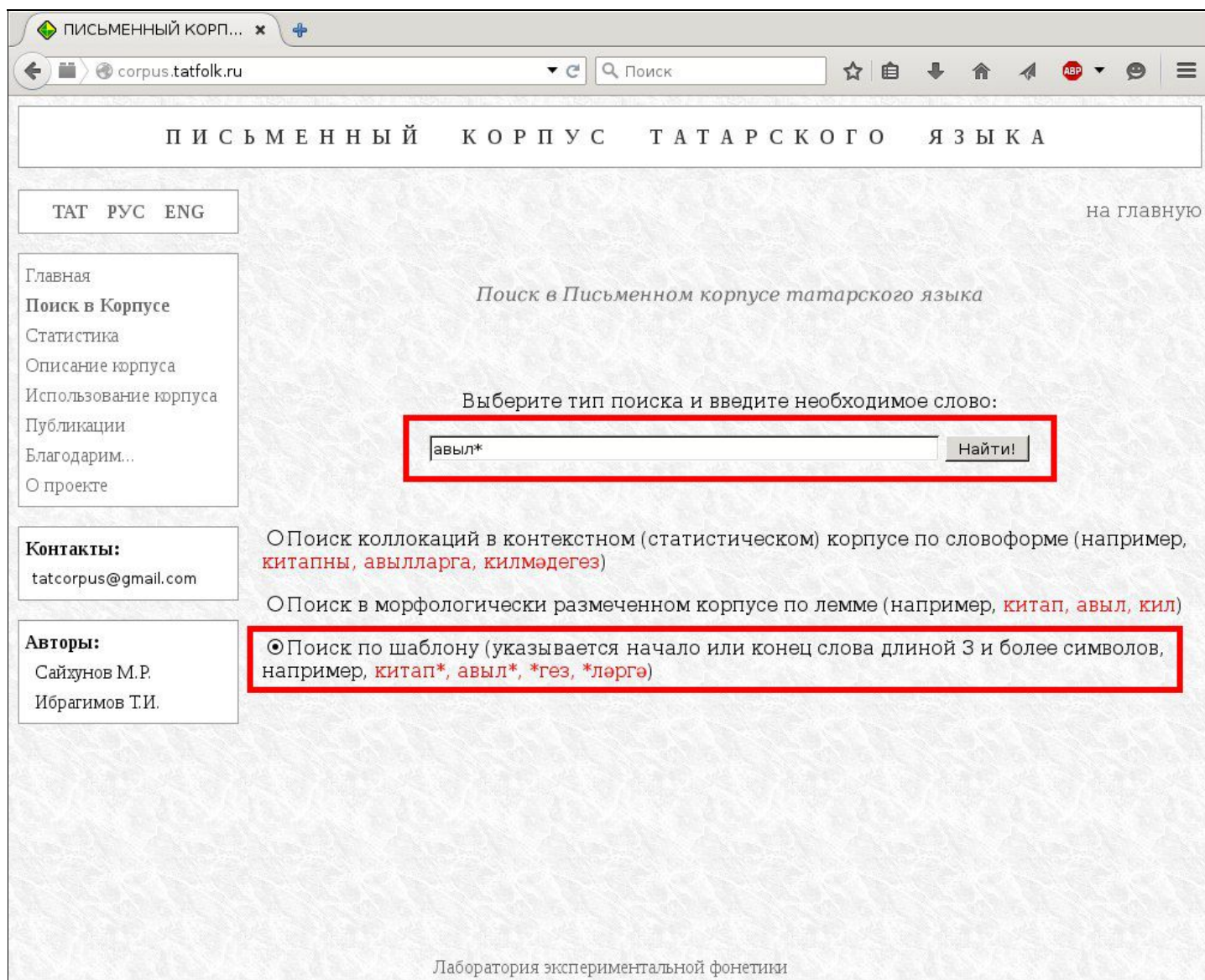
Указание источников и возможность просмотра полных текстов является идентичным описанному в предыдущем разделе.

Как найти в Корпусе все лексемы, начинающиеся с определенного префикса?

В Корпусе имеется возможность поиска лексем, имеющих одинаковую начальную часть.

Для этого на странице Поиска в текстовом поле указывается префикс и «галочкой» отмечается пункт «Поиск по шаблону...». Необходимо помнить, что длина префикса может составлять три и более символов.

Префикс должен иметь вид «авыл*», т.е. символ «звездочка» в конце является обязательным.



ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА

TAT РУС ENG на главную

Главная
Поиск в Корпусе
Статистика
Описание корпуса
Использование корпуса
Публикации
Благодарим...
О проекте

Поиск в Письменном корпусе татарского языка

Выберите тип поиска и введите необходимое слово:

○ Поиск коллокаций в контекстном (статистическом) корпусе по словоформе (например, **китапны, авылларга, килмәдегез**)

○ Поиск в морфологически размеченном корпусе по лемме (например, **китап, авыл, кил**)

◎ Поиск по шаблону (указывается начало или конец слова длиной 3 и более символов, например, **китап*, авыл*, *гез, *ләргә**)

Контакты:
tatcorpus@gmail.com

Авторы:
Сайхунов М.Р.
Ибрагимов Т.И.

Лаборатория экспериментальной фонетики

Вот пример того, что можно получить в результате такого запроса:

ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА

TAT РУС ENG СЛОВО: **авыл***

Главная
Поиск в Корпусе
Статистика
Описание корпуса
Использование корпуса
Публикации
Благодарим...
О проекте

Контакты:
tatcorpus@gmail.com

Авторы:
Сайхунов М.Р.
Ибрагимов Т.И.

▶ Шушыны күрә торып **авыл** жирлеге рәисе берни эшләми, ә **авыл** тәрбияләгән Русия мөфтиләр шурасы рәисе Равил Гайнетдин килеп бер хәлебезне дә белгәне юк, ди **авыл** халкы.
(источник: "Ислам - Татарлар һәм мөселманнар" (web-сайт)) [Найми текст](#)

▶ Алар район үзәгендәге кибетләргә рейдлар ясылар, кайчакта **авыл** кибетләрендә дә булалар.
(источник: "Авыл таңнары" газетасы (web-сайт)) [Найми текст](#)

▶ Бу вакыйгадан соң буага «Актаныш» буасы исемен кушалар, һәм шуннан соң бу исем елгага да күчә, елга исемнән **авыл** да «Актаныш:» исемен йөртә.
(источник: "Татар әкиятләре" (web-сайт)) [Найми текст](#)

▶ Метрәй **авыл** жирлеге үз эченә биш **авыл**: Метрәй (административ үзәк), Салауз Мухан, Ольгино, Иске Чакмак һәм Үрнәк **авыл**ларын ала.
(источник: "Мөслим районының официал мәгълүмәти сайты" (web-сайт)) [Найми текст](#)

▶ Әмма **авыл** хужалыгы эшқуарлары бер-берсеннән аерым хезмәт куя.
(источник: "Татар-информ" МА Татарстан Республикасы мәгълүмәт агентлығы (web-сайт)) [Найми текст](#)

▶ **Авыл** жирлекләре урамнарындагы электр энергиясенә кулланылышы хақында да ризасызлыктар бик күп.
(источник: "Туган як" газетасы (web-сайт)) [Найми текст](#)

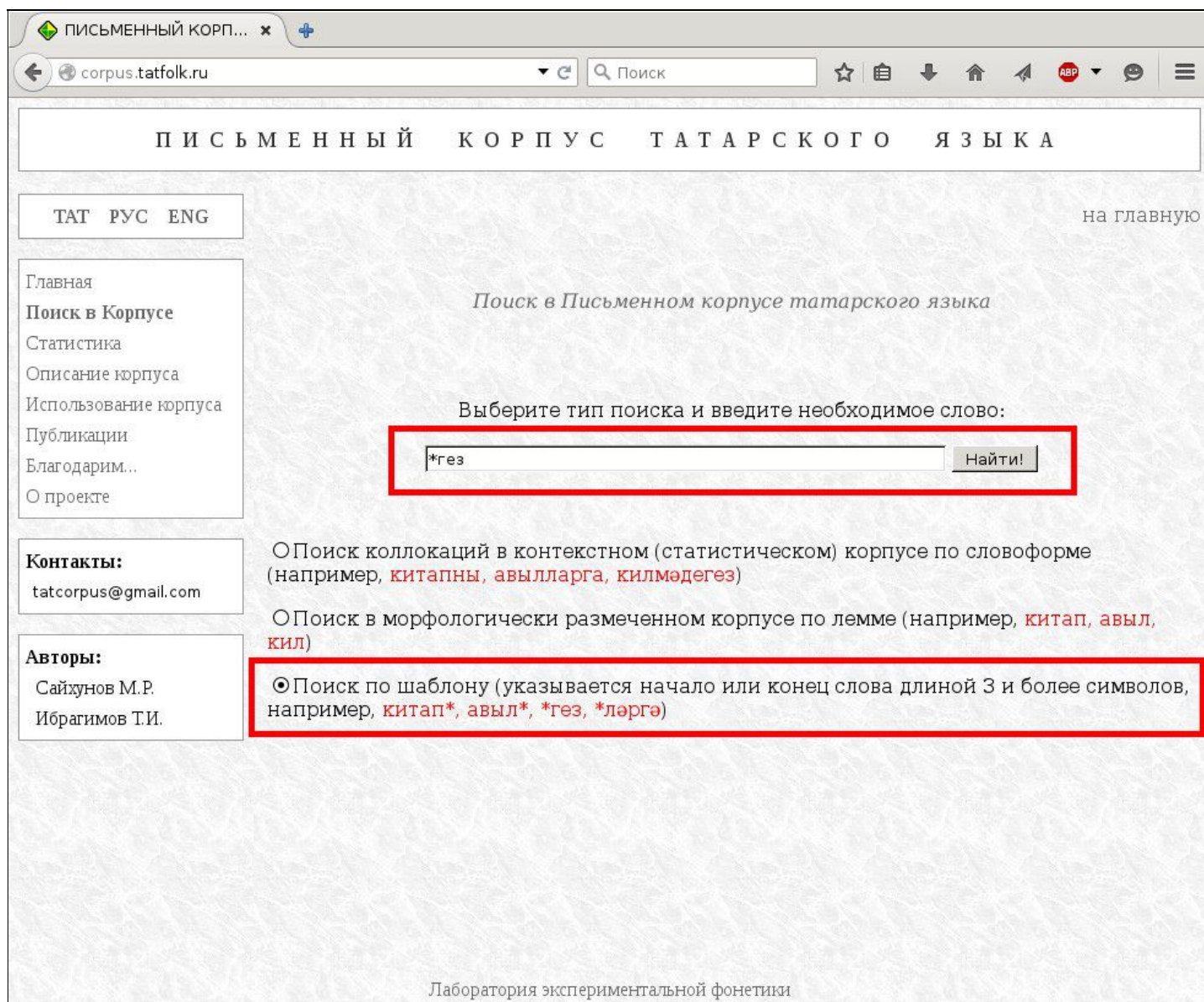
Лаборатория экспериментальной фонетики

Как найти в Корпусе все лексемы, заканчивающиеся определенным постфиксом?

В Корпусе имеется возможность поиска лексем, имеющих одинаковую конечную часть.

Для этого на странице Поиска в текстовом поле указывается постфикс и «галочкой» отмечается пункт «Поиск по шаблону...». Необходимо помнить, что длина постфикса может составлять три и более символов.

Поисковый запрос должен иметь вид «*гез», т.е. символ «звездочка» в начале является обязательным.



ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА

TAT РУС ENG [на главную](#)

Главная
Поиск в Корпусе
Статистика
Описание корпуса
Использование корпуса
Публикации
Благодарим...
О проекте

Поиск в Письменном корпусе татарского языка

Выберите тип поиска и введите необходимое слово:

Поиск коллокаций в контекстном (статистическом) корпусе по словоформе (например, **китапны, авылларга, килмәдегез**)

Поиск в морфологически размеченном корпусе по лемме (например, **китап, авыл, кил**)

Поиск по шаблону (указывается начало или конец слова длиной 3 и более символов, например, **китап*, авыл*, *гез, *ләргә**)

Контакты:
tatcorpus@gmail.com

Авторы:
Сайхунов М.Р.
Ибрагимов Т.И.

Лаборатория экспериментальной фонетики

Вот пример того, что можно получить в результате такого запроса:

ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА

TAT РУС ENG СЛОВО: *гез

Главная
Поиск в Корпусе
Статистика
Описание корпуса
Использование корпуса
Публикации
Благодарим...
О проекте

Контакты:
tatcorpus@gmail.com

Авторы:
Сайхунов М.Р.
Ибрагимов Т.И.

▶ Бу ханнар арасындагы көрәш Кече Мөхәммәтнең жиңүе белән төгәлләнә һәм ул Олы Урдага нигез салучы булып санала.
(источник: "Татарча текстлар, программалар" (web-сайт)) [Найти текст](#)

▶ Узган атнада бәйсез табиблар уздырган экспертиза, журналистны хастаханәдә тоту өчен нигез юк, дип белдергән иде.
(источник: "Азатлык Радиосы" (web-сайт)) [Найти текст](#)

▶ Чагыштырмача күптән түгел генә нигез салынган мәчетне төзүне тиз тоттылар.
(источник: "Татар-информ" МА Татарстан Республикасы мәгълумат агентлыгы (web-сайт)) [Найти текст](#)

▶ Европа өлгесендәге татар театрына нигез салучы бит син.
(источник: Низами Р.М. Без яшибез... - Казан: Татарстан китап нәширияты, 2011.)

▶ Бүген биредә "Казан заманча төргәкләү заводы"ның нигез ташлары салынды.
(источник: "Татар-информ" МА Татарстан Республикасы мәгълумат агентлыгы (web-сайт)) [Найти текст](#)

▶ Безгә крестьян ихтилаллары дип аңлатылган ул вакыттагы Степан Разинның бөтенроссия гражданныр сугышын, чынбарлыкта исә Алтын урданың законлы чын варислары булган Олы урда - Сарай татарлары, (казаклары дип тә әйтергә була) белән Мәскәү арасында Алтын урда империясе тәхетке өчен барган иң канкөешлы, иң аяусыз, иң компромиссыз,

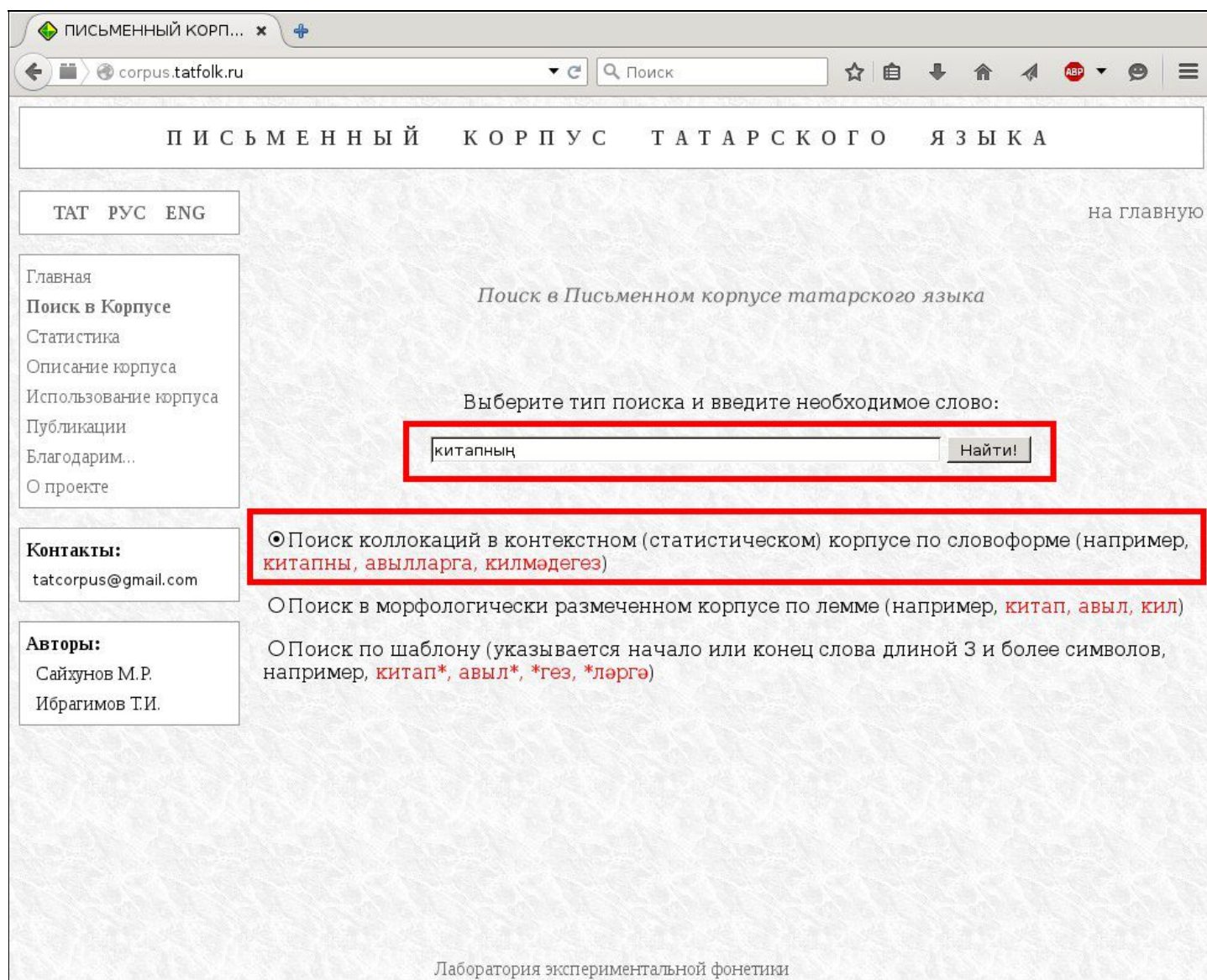
Лаборатория экспериментальной фонетики

Как найти в Корпусе словосочетание?

Поиск словосочетания в Корпусе имеет свою специфику. Так как основной особенностью Письменного корпуса татарского языка является его контекстно-статистическая направленность, то и словосочетания представлены в этом разделе.

Допустим, нам необходимо найти предложения с сочетанием «*китапның эчтәлеген*». Для достижения этой цели имеется три способа:

- 1) **Через правый контекст.** На странице поиска в текстовом поле пишем то слово, которое стоит слева, т.е. «*китапның*»! Отмечаем «галочкой» пункт «**Поиск коллокаций в контекстном (статистическом) корпусе по словоформе...**»



The screenshot shows the website 'ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА' (Written Corpus of the Tatar Language). The search interface includes a search bar with the word 'китапның' entered and a 'Найти!' button. Below the search bar, there are three radio button options for search types. The first option, 'Поиск коллокаций в контекстном (статистическом) корпусе по словоформе (например, китапны, авылларга, килмәдегез)', is selected and highlighted with a red box. The other two options are 'Поиск в морфологически размеченном корпусе по лемме (например, китап, авыл, кил)' and 'Поиск по шаблону (указывается начало или конец слова длиной 3 и более символов, например, китап*, авыл*, *гез, *ләргә)'. The website also features a navigation menu on the left and contact information at the bottom.

В открывшемся окне в разделе «Правый контекст» находим слово «эчтәлеген»:

ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА

TAT РУС ENG

СЛОВО: *китапның*
КОЛИЧЕСТВО СОВПАДЕНИЙ: 3120

ПРАВЫЙ КОНТЕКСТ: авторы (146; 1625.81), _NUMBER_ (144; 134.06), беренче (119; 579.32), икенче (86; 475.46), исеме (73; 567.74), төп (56; 282.71), исем (55; 453.97), бер (52; 84.03), соңгы (50; 232.22), эчтәлегә (44; 476.6), ахырында (42; 342.94), « (39; 39.47), һәр (39; 152.04), көреш (37; 400.51), “ (30; 3.95), иң (29; 59.37), авторлары (28; 299.1), тиражы (26; 323.12) **эчтәлеген (26; 285.34)**, мөхәррире (26; 226.15), исемен (22; 139.32), эчтәлегенә (21; 262.48), төзүчесе (20; 259.23), баш (17; 47.7), кадрен (16; 134.95), титул (16; 222.34), фәнни (15; 71.71), язылу (14; 117.73), әһәмияте (14; 113.48), дөньяга (13; 62.35), тышлыгы (13; 190.91), тышлыгында (13; 192.15), яңа (13; 10.95), тагын (13; 19.43), электрон (12; 62.8), кыйммәте (12; 122.25), редакторы (12; 109.51), ахырына (12; 77.81), рус (12; 40.54), туыландырылган (11; 145.04), басылып (11; 66.24), үз (11; 4.91), эченә (11; 58.35), башында (11; 48.32), дөнья (11; 28.38), нинди (11; 21.16), берничә (11; 26.21), зурлыгы (11; 110.07), дәвам (10; 89.46), азагында (10; 69.35), ике (10; 7.43), эчендә (9; 23.93), аерым (9; 21.22), беренчесе (9; 61.76), бәясе (9; 45.17), күп (9; 4.68), тарихы (9; 38.27), үзен (9; 30.2), тышлыгын (9; 143.19), исеменә (9; 52.86), өченче (9; 31.36), дәрәжәсен (8; 48.51), рәссамы (8; 66.54), үзенә (8; 23.84), чыгуы (8; 57.27), тышлыгына (7; 98.96), идея (7; 47.72), буеннан-буена (7; 78.26), илаһи (7; 49.64), тулы (7; 19.42), тышына (7; 80.96), максаты (7; 28.99), кулъязмасы (7; 74.51), тәржемәсен (7; 82.86), концепциясе (7; 61.05), дүртенче (7; 35.75), ролен (6; 34.36), алгы (6; 40.17), берсен (6; 33.5), тәүге (6; 28.69), битләрен (6; 57.76), бәясен (6; 41.32), нигезенә (6; 48.04), урыны (6; 21.71), җаваплы (6; 23.14), нигезен (6; 43.45), авырлыгы (6; 44.28), барлык (6; 7.04), калган (6; 9.22), киңәйтелгән (5; 36.03), битләре (5; 44.65), тәржемәсе (5; 42.59), архив (5; 34.78), озакламый (5; 31.25), тышында (5; 56.4), баштагы (5; 41.35), кырыена (5; 45.43), россиягә (5; 34.23), эчтәлегендә (5; 61.1), рецензенты (5; 86) [Показать все](#)

ЛЕВЫЙ КОНТЕКСТ: %^% (903; 1453.07), бу (305; 1270.4), , (224; 4.69)

Лаборатория экспериментальной фонетики

После щелчка мышью на данном слове, оно выделится красным цветом:

ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА

TAT РУС ENG

СЛОВО: **китапның**

КОЛИЧЕСТВО СОВПАДЕНИЙ: 3120

ПРАВЫЙ КОНТЕКСТ: авторы (146; 1625.81), _NUMBER_ (144; 134.06), беренче (119; 579.32), икенче (86; 475.46), исеме (73; 567.74), төп (56; 282.71), исем (55; 453.97), бер (52; 84.03), соңгы (50; 232.22), эчтөлөгө (44; 476.6), ахырында (42; 342.94), « (39; 39.47), һәр (39; 152.04), кереш (37; 400.51), " (30; 3.95), иң (29; 59.37), авторлары (28; 299.1), тиражы (26; 323.12), **эчтөлөгөн** (26; 285.34), мөхәррире (26; 226.15), исемән (22; 139.32), эчтөлөгәнә (21; 262.48), төзүчесе (20; 259.23), баш (17; 47.7), кадерен (16; 134.95), титул (16; 222.34), фәнни (15; 71.71), язылу (14; 117.73), әһәмияте (14; 113.48), дөньяга (13; 62.35), тышлыгы (13; 190.91), тышлыгында (13; 192.15), яңа (13; 10.95), тагын (13; 19.43), электрон (12; 62.8), кыйммәте (12; 122.25), редакторы (12; 109.51), ахырына (12; 77.81), рус (12; 40.54), туыландырылган (11; 145.04), басылып (11; 66.24), үз (11; 4.91), эченә (11; 58.35), башында (11; 48.32), дөнья (11; 28.38), нинди (11; 21.16), берничә (11; 26.21), зурлыгы (11; 110.07), дәвам (10; 89.46), азагында (10; 69.35), ике (10; 7.43), эчендә (9; 23.93), аерым (9; 21.22), беренчесе (9; 61.76), бәясә (9; 45.17), күп (9; 4.68), тарихы (9; 38.27), үзен (9; 30.2), тышлыгын (9; 143.19), исемәнә (9; 52.86), өченчә (9; 31.36), дәрәжәсен (8; 48.51), рәссамы (8; 66.54), үзенә (8; 23.84), чыгуы (8; 57.27), тышлыгына (7; 98.96), идея (7; 47.72), буеннан-буена (7; 78.26), илаһи (7; 49.64), тулы (7; 19.42), тышына (7; 80.96), максаты (7; 28.99), кулъязмасы (7; 74.51), тәржемәсен (7; 82.86), концепциясе (7; 61.05), дүртенчә (7; 35.75), ролен (6; 34.36), алгы (6; 40.17), берсен (6; 33.5), тәүгә (6; 28.69), битләрен (6; 57.76), бәясән (6; 41.32), нигезәнә (6; 48.04), урыны (6; 21.71), жаваплы (6; 23.14), нигезән (6; 43.45), авырлыгы (6; 44.28), барлык (6; 7.04), калган (6; 9.22), кинәйтелгән (5; 36.03), битләре (5; 44.65), тәржемәсе (5; 42.59), архив (5; 34.78), озакламый (5; 31.25), тышында (5; 56.4), баштагы (5; 41.35), кырыена (5; 45.43), россиягә (5; 34.23), эчтөлөгәнә (5; 61.1), рецензенты (5; 86) [Показать все](#)

ЛЕВЫЙ КОНТЕКСТ: %^% (903; 1453.07), бу (305; 1270.4), , (224; 4.69)

Лаборатория экспериментальной фонетики

Главная
Поиск в Корпусе
Статистика
Описание корпуса
Использование корпуса
Публикации
Благодарим...
О проекте

Контакты:
tatcorpus@gmail.com

Авторы:
Сайхунов М.Р.
Ибрагимов Т.И.

А внизу в разделе «Примеры» появится список предложений, где используется данное словосочетание:

The screenshot shows a web browser window with the URL 'corpus.tatfolk.ru'. The page title is 'П И С Ъ М Е Н Н Ы Й К О Р П У С Т А Т А Р С К О Г О Я З Ы К А'. On the left, there are navigation menus for language selection (TAT, РУС, ENG), a main menu with links like 'Главная', 'Поиск в Корпусе', and 'Статистика', and contact information for 'tatcorpus@gmail.com'. The main content area is titled 'ПРИМЕРЫ:' and contains six examples of the phrase 'китапның эчтәлеген' used in various contexts, each with a source citation and a link to the full text.

П И С Ъ М Е Н Н Ы Й К О Р П У С Т А Т А Р С К О Г О Я З Ы К А

TAT РУС ENG

Главная
Поиск в Корпусе
Статистика
Описание корпуса
Использование корпуса
Публикации
Благодарим...
О проекте

Контакты:
tatcorpus@gmail.com

Авторы:
Сайхунов М.Р.
Ибрагимов Т.И.

ПРИМЕРЫ:

- ▶ Аджтимушкой турындагы бу язмамның яртысы диярлек Керчтә сатып алынган **китапның эчтәлеген** сөйләп чыгу булуга карамастан, «сөрген» халкына бик сокландыргыч тәэсир ясады ул.
(источник: "Идел" журналы (web-сайт)) [Найти текст](#)
- ▶ Элекке хәрбинең адвокаты исә **китапның эчтәлеген** рәсмиләр белән килештерергә кирәклеген кире кага.
(источник: "Азатлык Радиосы" (web-сайт)) [Найти текст](#)
- ▶ **Китапның эчтәлеген** бик матур, сәнгатьле сөйли белгән ул.
(источник: "Яшел Үзән" газетасы (web-сайт)) [Найти текст](#)
- ▶ Ә Шөгәр төбөгөндә гомер иткән Наилә Сиражетдин кызы Бикчуринаның язып калдырган истәлектәре, нәсел шәжәрәсе **китапның эчтәлеген** тарихи әһәмияткә ия мөһлүмат, документлар белән баетты.
(источник: "Самара региональ «Дуслык» һәм Самара шәһәр «Ак бәхет» нәшриятының уртак сайты" (web-сайт)) [Найти текст](#)
- ▶ Китап исеме **китапның эчтәлеген** сөйләп бирергә тиеш түгел.
(источник: Рабит Батулла. Кәбир Бәкернең тууы)
- ▶ Шулай дип башлап китте дә ул кичә укыган **китапның эчтәлеген** сөйләргә тотынды.
(источник: Разил Вәлиев. Кунак булып килде, хужа булып китте...)
- ▶ Әйе, без синең хәтер күзәнәкләреңә йөз меңнәрчә **китапның эчтәлеген** язып куйдык.
(источник: Адлер Тимергалин. КОХАУ РОНГО-РОНГО)

Лаборатория экспериментальной фонетики

2) **Через левый контекст.** В данном случае на странице поиска в текстовом поле пишем правое (второе) слово, т.е. «эчтәлеген».

ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА

TAT РУС ENG на главную

Главная
Поиск в Корпусе
Статистика
Описание корпуса
Использование корпуса
Публикации
Благодарим...
О проекте

Поиск в Письменном корпусе татарского языка

Выберите тип поиска и введите необходимое слово:

эчтәлеген

Поиск коллокаций в контекстном (статистическом) корпусе по словоформе (например, китапны, авылларга, килмәдегез)

Поиск в морфологически размеченном корпусе по лемме (например, китап, авыл, кил)

Поиск по шаблону (указывается начало или конец слова длиной 3 и более символов, например, китап*, авыл*, *гез, *ләргә)

Контакты:
tatcorpus@gmail.com

Авторы:
Сайхунов М.Р.
Ибрагимов Т.И.

Лаборатория экспериментальной фонетики

В разделе «Левый контекст» находим слово «*китапның*»:

ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА

TAT РУС ENG

Главная
Поиск в Корпусе
Статистика
Описание корпуса
Использование корпуса
Публикации
Благодарим...
О проекте

Контакты:
tatcorpus@gmail.com

Авторы:
Сайхунов М.Р.
Ибрагимов Т.И.

ПРАВЫЙ КОНТЕКСТ: , (146; 5.21), төшкил (71; 568.22), һәм (65; 60.1), сөйләп (62; 609.69), дә (53; 78.42), яңарту (42; 473.06), сөйләү (38; 452.26), тагын (35; 143.58), ачып (33; 255.32), аңлау (21; 215.15), билгели (19; 171.77), аңлап (17; 129.27), ачу (16; 94.23), тулысынча (16; 101.47), яхшырту (14; 116.92), сөйләргә (14; 120.56), кыскача (14; 130.53), үзгәрту (14; 131.99), баету (14; 169.71), ача (13; 104.56), искә (12; 52.02), сөйли (12; 65.12), баета (11; 129.14), үз (10; 9.06), язып (10; 52.7), ачыклау (9; 65.94), камилләштерү (9; 73.21), дәрәс (9; 29.53), табу (9; 60.23), аңлату (9; 74.93), аңлатып (9; 67.58), үзгәртәргә (8; 67.1), ничек (8; 17.93), аңларга (8; 56.47), үзләштерү (7; 59.95), аңлый (7; 43.1), яхшы (7; 11.76), хасил (7; 58.28), шактый (6; 16.77), төгәл (6; 30.11), бәян (6; 38.33), ачарга (6; 40.23), тирәнәйтәргә (6; 91.36), сөйләгез (6; 66.54), тулырак (6; 46.85), әйтеп (6; 20.64), табабыз (5; 55.02), мәктәптә (5; 19.67), үзенчәлекле (5; 23.57), ачуда (5; 45.24), тикшереп (5; 31.86), югалта (5; 40.28), кыска (5; 23.24), аңлауга (5; 59.92), ачар (5; 52.74), баетырга (5; 59.78), язу (5; 27.21), сайлау (5; 18.86), эшләү (5; 19.28), билгеләү (5; 29.75), бермә-бер (5; 41.76), тулыландырып (4; 39.82), бирә (4; 7.02), болай (4; 10.56), ачуга (4; 33.13), язарга (4; 20.58), истәлекләр (4; 30.09), саклаган (4; 30.5), яттан (4; 32.53), үзгәртә (4; 32.33), тулыландыру (4; 38.61), тирән (4; 15.33), саклап (4; 13.36), баетып (4; 42.29), халыкка (4; 13.77), тулы (4; 10.48), табарга (4; 19.72), сөйләтү (4; 60.72), аңлата (4; 19.67), сөйлиләр (4; 27.55), өйрәнә (4; 22.21), белү (4; 20.17), баетуга (4; 46.86), укып (4; 13.67), ачыклай (4; 33.08), төкъдим (4; 5.24), ; (4; 4.68), түбәндәге (3; 14.01), алдагы (3; 10.51), телдән (3; 22.33), сөйләде (3; 8.55), үзгәртеп (3; 14.58), беләргә (3; 11.78), яңартуга (3; 27.54), авторның (3; 17.9), микән (3; 11.53), яхшырак (3; 14.17), салырга (3; 14.49), баеткан (3; 34.99), янадан (3; 8.48) [Показать все](#)

ЛЕВЫЙ КОНТЕКСТ: , (179; 26.12), аның (114; 521.98), әсәрнең (91; 1154.86), төп (72; 466.93), һәм (59; 46.83), идея (59; 710.65), кыскача (45; 525.52), текстның (41; 640.25), аларның (38; 173.86), **китапның (26; 285.34)**, әсәрләренң (25; 312.73), әсәр (25; 215.59), хатның (22; 311.81), бирүнең (21; 233.2), бирү (18; 82.78), тулы (14; 69.61), газетаның (14; 137.22), жырының (13; 146.04), хикәянең (13; 182.25), китапларның (12; 138.96), спектакльнең (11; 116.53)

Лаборатория экспериментальной фонетики

После клика мышью данное слово становится красным:

ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА

TAT РУС ENG

Главная
Поиск в Корпусе
Статистика
Описание корпуса
Использование корпуса
Публикации
Благодарим...
О проекте

Контакты:
tatcorpus@gmail.com

Авторы:
Сайхунов М.Р.
Ибрагимов Т.И.

ПРАВЫЙ КОНТЕКСТ: , (146; 5.21), төшкил (71; 568.22), һәм (65; 60.1), сөйләп (62; 609.69), дә (53; 78.42), яңарту (42; 473.06), сөйләү (38; 452.26), тагын (35; 143.58), ачып (33; 255.32), аңлау (21; 215.15), билгели (19; 171.77), аңлап (17; 129.27), ачу (16; 94.23), тулысынча (16; 101.47), яхшырту (14; 116.92), сөйләргә (14; 120.56), кыскача (14; 130.53), үзгәртү (14; 131.99), баету (14; 169.71), ача (13; 104.56), искә (12; 52.02), сөйли (12; 65.12), баета (11; 129.14), үз (10; 9.06), язып (10; 52.7), ачыклау (9; 65.94), камилләштерү (9; 73.21), дәрәс (9; 29.53), табу (9; 60.23), аңлату (9; 74.93), аңлатып (9; 67.58), үзгәртәргә (8; 67.1), ничек (8; 17.93), аңларга (8; 56.47), үзләштерү (7; 59.95), аңлый (7; 43.1), яхшы (7; 11.76), хасил (7; 58.28), шактый (6; 16.77), төгәл (6; 30.11), баян (6; 38.33), ачарга (6; 40.23), тирәнәйтергә (6; 91.36), сөйләгез (6; 66.54), тулырак (6; 46.85), әйтеп (6; 20.64), табабыз (5; 55.02), мәктәптә (5; 19.67), үзенчәлекле (5; 23.57), ачуда (5; 45.24), тикшереп (5; 31.86), югалта (5; 40.28), кыска (5; 23.24), аңлауга (5; 59.92), ачар (5; 52.74), баетырга (5; 59.78), язу (5; 27.21), сайлау (5; 18.86), эшләү (5; 19.28), билгеләү (5; 29.75), бермә-бер (5; 41.76), тулыландырып (4; 39.82), бирә (4; 7.02), болай (4; 10.56), ачуга (4; 33.13), язарга (4; 20.58), истәлекләр (4; 30.09), саклаган (4; 30.5), яттан (4; 32.53), үзгәртә (4; 32.33), тулыландыру (4; 38.61), тирән (4; 15.33), саклап (4; 13.36), баетып (4; 42.29), халыкка (4; 13.77), тулы (4; 10.48), табарга (4; 19.72), сөйләтү (4; 60.72), аңлата (4; 19.67), сөйлиләр (4; 27.55), өйрәнә (4; 22.21), белү (4; 20.17), баетуга (4; 46.86), укуп (4; 13.67), ачыкый (4; 33.08), төкәдим (4; 5.24), ; (4; 4.68), түбәндәгә (3; 14.01), алдагы (3; 10.51), телдән (3; 22.33), сөйләде (3; 8.55), үзгәртеп (3; 14.58), белергә (3; 11.78), яңартуга (3; 27.54), авторның (3; 17.9), микән (3; 11.53), яхшырак (3; 14.17), салырга (3; 14.49), баеткан (3; 34.99), яңадан (3; 8.48) [Показать все](#)

ЛЕВЫЙ КОНТЕКСТ: , (179; 26.12), аның (114; 521.98), әсәрнең (91; 1154.86), төп (72; 466.93), һәм (59; 46.83), идея (59; 710.65), кыскача (45; 525.52), текстның (41; 640.25), аларның (38; 173.86), **китапның** (26; 285.34), әсәрләрнең (25; 312.73), әсәр (25; 215.59), хатның (22; 311.81), бирүнең (21; 235.2), биру (18; 82.78), тулы (14; 69.61), газетаның (14; 137.22), жырның (13; 146.04), хикәянең (13; 182.25), китапларның (12; 138.96), спектакльнең (11; 116.53), текст (11; 104.1), мәгънәви (9; 97.29), мәкаләнең (9; 112.41), оригиналның (8; 128.65)

Лаборатория экспериментальной фонетики

А в разделе «Примеры» видим тот же список предложений с данным словосочетанием, что и в первом варианте поиска.

ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА

TAT РУС ENG

Главная
Поиск в Корпусе
Статистика
Описание корпуса
Использование корпуса
Публикации
Благодарим...
О проекте

Контакты:
tatcorpus@gmail.com

Авторы:
Сайхунов М.Р.
Ибрагимов Т.И.

ПРИМЕРЫ:

- ▶ Аджитмушкай турындагы бу язмамның яртысы диярлек Керчтә сатып алынган **китапның эчтәлеген** сөйләп чыгу булуга карамастан, «сөрген» халкына бик сокландыргыч тәэсир ясады ул.
(источник: "Идел" журналы (web-сайт)) [Найти текст](#)
- ▶ Элекке хәрбиң адвокаты исә **китапның эчтәлеген** рәсмиләр белән килештерергә кирәклеген кире кага.
(источник: "Азатлык Радиосы" (web-сайт)) [Найти текст](#)
- ▶ **Китапның эчтәлеген** бик матур, сәнгатьле сөйли белгән ул.
(источник: "Яшел Үзән" газетасы (web-сайт)) [Найти текст](#)
- ▶ Ә Шөгәр төбөгәндә гомер иткән Наилә Сиражетдин кызы Бикчуринаның язып калдырган истәлекләре, нәсел шәжәрәсе **китапның эчтәлеген** тарихи әһәмияткә ия мәгълүмат, документлар белән баетты.
(источник: "Самара региональ «Дуслык» һәм Самара шөһәр «Ак бәхет» ижади-иҗтимагый оешмаларының уртак сайты" (web-сайт)) [Найти текст](#)
- ▶ Китап исеме **китапның эчтәлеген** сөйләп бирергә тиеш түгел.
(источник: Рабит Батулла. Кәбир Бәкернең тууы)
- ▶ Шулай дип башлап китте дә ул кичә укыган **китапның эчтәлеген** сөйләргә тотынды.
(источник: Разил Вәлиев. Кунак булып килде, хужа булып китте...)
- ▶ Әйе, без синең хәтер күзәнәкләреңә йөз меңнәрчә **китапның эчтәлеген** язып куйдык.
(источник: Адлер Тимергалин. КОХАУ РОНГО-РОНГО)

Лаборатория экспериментальной фонетики

3) **Сложный морфологический поиск.** Если вашей целью не является получение какой-либо статистической информации, и вы просто хотите увидеть примеры употребления (предложения) с указанными комбинациями слов, тогда наиболее оптимальным вариантом является использование [Системы сложного морфологического поиска](#), описанной в следующей главе.

Система сложного морфологического поиска:

Что представляет из себя данная система?

В 2014 году была произведена морфологическая разметка Корпуса. В основу метаязыка грамматических помет положена система «тегов» для тюркских языков, разработанная международным проектом [Apertium](#). В рамках данного проекта создается система машинного перевода для большого количества языков. Основными аргументами в пользу выбора морфологического таггера Apertium'a для разметки Письменного корпуса являются:

- высокое качество морфологической аннотации;
- абсолютная открытость этого проекта: все исходные коды и наработки публично доступны всем желающим бесплатно.

Разработанная нами в 2015-2016 годах **Система сложного морфологического поиска** позволяет производить поиск в Корпусе на основе различных комбинаций таких параметров, как словоформа, лемма, набор морфологических (грамматических) тегов, начало, середина, конец слова с указанием возможных расстояний между ними. Максимальная длина поискового запроса достигает пяти токенов и, соответственно, четыре показателя расстояния между ними.

Для того чтобы открыть основную форму Сложного морфологического поиска необходимо в главном меню нажать на ссылку «Поиск в Корпусе» и в открывшемся окне выбрать пункт «Сложный морфологический поиск»:






П И С Ъ М Е Н Н Ы Й К О Р П У С Т А Т А Р С К О Г О Я З Ы К А

TAT RUS ENG на главную

Главная
Поиск в Корпусе
Синтез татарской речи
Инструкция
Статистика
Описание корпуса
Использование корпуса
Публикации
Благодарим...
О проекте
Полезная информация

Поиск в Письменном корпусе татарского языка

Выберите тип поиска и введите необходимое слово:

Слово 1:	<input type="text"/>		Расстояние 1:	<input type="text" value="1-1"/>
Слово 2:	<input type="text"/>		Расстояние 2:	<input type="text" value="1-1"/>
Слово 3:	<input type="text"/>		Расстояние 3:	<input type="text" value="1-1"/>
Слово 4:	<input type="text"/>		Расстояние 4:	<input type="text" value="1-1"/>
Слово 5:	<input type="text"/>		<input type="button" value="Найти!"/>	

Поиск коллокаций в контекстном (статистическом) корпусе по словоформе (например, **китапны, авылларга, килмәдегез**)

Поиск в морфологически размеченном корпусе по лемме (например, **китап, авыл, кил**)

Поиск по шаблону (указывается начало или конец слова длиной 3 и более символов, например, **китап*, авыл*, *гез, *ләргә**)

Сложный морфологический поиск (Инструкция)

Лаборатория экспериментальной фонетики

Пример поиска по сочетанию слов «кунакка бардык» :

П И С Ъ М Е Н Н Ы Й К О Р П У С Т А Т А Р С К О Г О Я З Ы К А

ТАТ РУС ENG

[на главную](#)

Главная

Поиск в Корпусе

Синтез татарской речи

Инструкция

Статистика

Описание корпуса

Использование корпуса

Публикации

Благодарим...

О проекте

Полезная информация

Поиск в Письменном корпусе татарского языка

Выберите тип поиска и введите необходимое слово:

Слово 1:	<input type="text" value="кунакка"/>		Расстояние 1:	<input type="text" value="1-1"/>
Слово 2:	<input type="text" value="бардык"/>		Расстояние 2:	<input type="text" value="1-1"/>
Слово 3:	<input type="text"/>		Расстояние 3:	<input type="text" value="1-1"/>
Слово 4:	<input type="text"/>		Расстояние 4:	<input type="text" value="1-1"/>
Слово 5:	<input type="text"/>		<input type="button" value="Найти!"/>	

- Поиск коллокаций в контекстном (статистическом) корпусе по словоформе (например, **китапны, авылларга, килмәдегез**)
- Поиск в морфологически размеченном корпусе по лемме (например, **китап, авыл, кил**)
- Поиск по шаблону (указывается начало или конец слова длиной 3 и более символов, например, **китап*, авыл*, *гез, *ләргә**)
- Сложный морфологический поиск (Инструкция)

Авторы:
Сайхунов М.Р.
Ибрагимов Т.И.
tatcorpus@gmail.com

Лаборатория экспериментальной фонетики

В основном окне результатов поиска в самом верху отображен набор параметров поиска. Под ним размещена информация о точном количестве найденных единиц.

П И С Ъ М Е Н Н Ы Й К О Р П У С Т А Т А Р С К О Г О Я З Ы К А

TAT RUS ENG

Главная
Поиск в Корпусе
Синтез татарской речи
Инструкция
Статистика
Описание корпуса
Использование корпуса
Публикации
Благодарим...
О проекте

Полезная информация

Авторы:
Сайхунов М.Р.
Ибрагимов Т.И.
tatcorpus@gmail.com

ЗАПРОС: {кунакка} 1-1 {бардык} 1-1 {} 1-1 {} 1-1 {}
КОЛИЧЕСТВО СОВПАДЕНИЙ: 20

▶ Икенче көн дә күтәрэнке күңел белән узды: хатынның туганнарына **кунакка бардык**, яисә үзләре безгә килделәр.
(источник: "Безнең гәҗит" газетасы (web-сайт)) [Найти текст](#)

▶ Менә без дә Чаллыга **кунакка бардык**, шәһәр карап, аның матурлыгына сокланып йөрдек.
(источник: "Безнең гәҗит" газетасы (web-сайт)) [Найти текст](#)

▶ Шактый еллар узгач, аның илле яшьлек юбилеена **кунакка бардык**.
(источник: "Безнең гәҗит" газетасы (web-сайт)) [Найти текст](#)

▶ – Кунаклар килде, **кунакка бардык**, – дип җавап кайтаралар.
(источник: "Безнең гәҗит" газетасы (web-сайт)) [Найти текст](#)

▶ Иртүк мөезин хатынына **кунакка бардык**.
(источник: "Идел" журналы (web-сайт)) [Найти текст](#)

▶ Без әни белән Нәфисә апаларга **кунакка бардык**.
(источник: "Яңарыш" газетасы (web-сайт)) [Найти текст](#)

▶ Бервакыт Украинада яшәүче әтиемнең туганнарына **кунакка бардык**.
(источник: "Яңарыш" газетасы (web-сайт)) [Найти текст](#)

▶ Мәселән, узган атнада **жәйге** лагерыбызда тәрбияләнүче балаларыбыз белән аларга **кунакка бардык**.
(источник: "Азатлык Радиосы" (web-сайт)) [Найти текст](#)

▶ Без, делегация белән, Яр Чаллы каласына Советлар Союзы Герое Илдар абый Маннанов янына **кунакка бардык**.
(источник: "Әдипләр. Әхмәт Дусайлының шәхси порталы" (web-сайт)) [Найти текст](#)

Лаборатория экспериментальной фонетики

Указанные при поиске параметры в найденных предложениях выделяются красным цветом. При наведении на каждый из них курсора мыши, появляется окно с отображением леммы (начальной формы данного слова в интерпретации системы Arertium) и набора характерных для данного словоупотребления морфологических тегов.

П И С Ъ М Е Н Н Ы Й К О Р П У С Т А Т А Р С К О Г О Я З Ы К А

ТАТ РУС ENGЗАПРОС: {кунакка} 1-1 {бардык} 1-1 {} 1-1 {} 1-1 {}

Главная

Поиск в Корпусе

Синтез татарской речи

Инструкция

Статистика

Описание корпуса

Использование корпуса

Публикации

Благодарим...

О проекте

Полезная информация

КОЛИЧЕСТВО СОВПАДЕНИЙ: 20

▶ Икенче көн дә күтәрэнке күңел белән узды: хатынның туганнарына **кунакка бардык**, яисә үзләре безгә килделәр.
(источник: "Безнең гәҗит" газетасы (web-сайт)) [Найти текст](#)

▶ Менә без дә Чаллыга **кунакка бардык**, шәһәр карап, аның матурлыгына сокланып йөрдек.
(источник: "Безнең гәҗит" газетасы (web-сайт)) [Найти текст](#)

▶ Шактый еллар узгач, аның илле яшьлек юбилеена **кунакка бардык**.
(источник: "Безнең гәҗит" газетасы (web-сайт)) [Найти текст](#)

`(кунак) <dat>,<n>,<sg>`

▶ – Кунаклар килде, **кунакка бардык**, – дип җавап кайтаралар.
(источник: "Безнең гәҗит" газетасы (web-сайт)) [Найти текст](#)

▶ Иртүк мөезин хатынына **кунакка бардык**.
(источник: "Идел" журналы (web-сайт)) [Найти текст](#)

▶ Без әни белән Нәфисә апаларга **кунакка бардык**.
(источник: "Яңарыш" газетасы (web-сайт)) [Найти текст](#)

▶ Бервакыт Украинада яшәүче әтиемнең туганнарына **кунакка бардык**.
(источник: "Яңарыш" газетасы (web-сайт)) [Найти текст](#)

▶ Мәсәлән, узган атнада җәйге лагерыбызда тәрбияләнүче балаларыбыз белән аларга **кунакка бардык**.
(источник: "Азатлык Радиосы" (web-сайт)) [Найти текст](#)

▶ Без, делегация белән, Яр Чаллы каласына Советлар Союзы Герое Илдар абый Маннанов янына **кунакка бардык**.
(источник: "Әдиләр. Әхмәт Дусайлының шәхси порталы" (web-сайт)) [Найти текст](#)

Лаборатория экспериментальной фонетики

Пример всплывающего окна при наведении курсора на слово «бардык»:

П И С Ъ М Е Н Н Ы Й К О Р П У С Т А Т А Р С К О Г О Я З Ы К А

TAT RUS ENG

Главная
Поиск в Корпусе
Синтез татарской речи
Инструкция
Статистика
Описание корпуса
Использование корпуса
Публикации
Благодарим...
О проекте
Полезная информация

Авторы:
Сайхунов М.Р.
Ибрагимов Т.И.
tatcorpus@gmail.com

ЗАПРОС: {кунакка} 1-1 {бардык} 1-1 {} 1-1 {} 1-1 {}

КОЛИЧЕСТВО СОВПАДЕНИЙ: 20

▶ Икенче көн дә күтәренке күңел белән узды: хатынның туганнарына **кунакка бардык**, яисә үзләре безгә килделәр.
(источник: "Безнең гәҗит" газетасы (web-сайт)) [Найти текст](#)

▶ Менә без дә Чаллыга **кунакка бардык**, шәһәр карап, аның матурлыгына сокланып йөрдек.
(источник: "Безнең гәҗит" газетасы (web-сайт)) [Найти текст](#)

▶ Шактый еллар узгач, аның илле яшьлек юбилеена **кунакка бардык**.
(источник: "Безнең гәҗит" газетасы (web-сайт)) [Найти текст](#)

▶ – Кунаклар килде, **кунакка бардык**, – дип җавап кайтаралар.
(источник: "Безнең гәҗит" газетасы (web-сайт)) [Найти текст](#)

▶ Иртүк мөезин хатынына **кунакка бардык**.
(источник: "Идел" журналы (web-сайт)) [Найти текст](#)

▶ Без әни белән Нәфисә апаларга **кунакка бардык**.
(источник: "Яңарыш" газетасы (web-сайт)) [Найти текст](#)

▶ Бервакыт Украинада яшәүче әтиемнең туганнарына **кунакка бардык**.
(источник: "Яңарыш" газетасы (web-сайт)) [Найти текст](#)

▶ Мәсәлән, узган атнада **жәйге** лагерыбызда тәрбияләнүче балаларыбыз белән аларга **кунакка бардык**.
(источник: "Азатлык Радиосы" (web-сайт)) [Найти текст](#)

▶ Без, делегация белән, Яр Чаллы каласына Советлар Союзы Герое Илдар абый Маннанов янына **кунакка бардык**.
(источник: "Әдипләр. Әхмәт Дусайлының шәхси порталы" (web-сайт)) [Найти текст](#)

(бар) <if>, <iv>, <p1>, <pl>, <v>

Лаборатория экспериментальной фонетики

Следует помнить: в связи с тем, что морфологическая разметка Корпуса произведена в автоматическом режиме, в ней могут содержаться ошибки!

Как найти в Корпусе комбинации с указанием словоформ (словосочетания)?

Для того, чтобы выполнить поиск в Корпусе какого-либо сочетания слов необходимо ввести в текстовые поля с названиями «Слово 1», «Слово 2», «Слово 3», «Слово 4», «Слово 5» по одному слову. Таким образом, максимально возможная длина искомой фразы составляет пять слов.

П И С Ъ М Е Н Н Ы Й К О Р П У С Т А Т А Р С К О Г О Я З Ы К А

ТАТ РУС ENG

на главную

Главная

Поиск в Корпусе

Синтез татарской речи

Инструкция

Статистика

Описание корпуса

Использование корпуса

Публикации

Благодарим...

О проекте

Полезная информация

Поиск в Письменном корпусе татарского языка

Выберите тип поиска и введите необходимое слово:

Слово 1:	<input type="text" value="яңа"/>		Расстояние 1:	<input type="text" value="1-1"/>
Слово 2:	<input type="text" value="ел"/>		Расстояние 2:	<input type="text" value="1-1"/>
Слово 3:	<input type="text" value="бэйрәмендә"/>		Расстояние 3:	<input type="text" value="1-1"/>
Слово 4:	<input type="text"/>		Расстояние 4:	<input type="text" value="1-1"/>
Слово 5:	<input type="text"/>		<input type="button" value="Найти!"/>	

- Поиск коллокаций в контекстном (статистическом) корпусе по словоформе (например, **китапны, авылларга, килмәдегез**)
- Поиск в морфологически размеченном корпусе по лемме (например, **китап, авыл, кил**)
- Поиск по шаблону (указывается начало или конец слова длиной 3 и более символов, например, **китап*, авыл*, *гез, *ләргә**)
- Сложный морфологический поиск (Инструкция)

Авторы:
Сайхунов М.Р.
Ибрагимов Т.И.
tatcorpus@gmail.com

Лаборатория экспериментальной фонетики

Результаты поиска фразы «яңа ел бәйрәмендә»:

П И С Ъ М Е Н Н Ы Й К О Р П У С Т А Т А Р С К О Г О Я З Ы К А

TAT РУС ENG

ЗАПРОС: {яңа} 1-1 {ел} 1-1 {бәйрәмендә} 1-1 {} 1-1 {}

КОЛИЧЕСТВО СОВПАДЕНИЙ: 55

Главная

Поиск в Корпусе

Синтез татарской речи

Инструкция

Статистика

Описание корпуса

Использование корпуса

Публикации

Благодарим...

О проекте

Полезная информация

Авторы:

Сайхунов М.Р.

Ибрагимов Т.И.

tatcorpus@gmail.com

▶ Алар **Яңа ел бәйрәмендә** балаларны һәм, гомумән, шундагы һәркемне чын Кыш бабай һәм Кар кызы белән очрашу оештыралар.

(источник: "Безнең гәҗит" газетасы (web-сайт)) [Найти текст](#)

▶ Инде китү ягына борылгач та, Наилә апа артымнан килеп: "Кызым, тагын кил. **Яңа ел бәйрәмендә** бик күңелле була", – дип кочып алды.

(источник: "Безнең гәҗит" газетасы (web-сайт)) [Найти текст](#)

▶ Бу – **Яңа ел бәйрәмендә** 600 млн литр исерткеч елга булып агачак дигән сүз.

(источник: "Безнең гәҗит" газетасы (web-сайт)) [Найти текст](#)

▶ **Яңа ел бәйрәмендә** һәм Сабантуйда.

(источник: "Безнең гәҗит" газетасы (web-сайт)) [Найти текст](#)

▶ **Яңа ел бәйрәмендә** уңышлы гына чыгыш ясадык.

(источник: "Безнең гәҗит" газетасы (web-сайт)) [Найти текст](#)

▶ **Яңа ел бәйрәмендә** бик күңелле була", – дип кочып алды.

(источник: "Безнең гәҗит" газетасы (web-сайт)) [Найти текст](#)

▶ **Яңа ел бәйрәмендә,**

(источник: "Туган як" газетасы (web-сайт)) [Найти текст](#)

▶ – Эдуард Анатольевич, халыкны **Яңа ел бәйрәмендә** чыршылар белән тәэмин итүгә әзерлек мәсьәләсе ничек тора?

(источник: "Зәй офыклары" газетасы (web-сайт)) [Найти текст](#)

▶ Республика **Яңа ел бәйрәмендә** катнашучыларны тәбрикләргә Татарстан Президенты Рөстәм Миңнеханов килде.

(источник: Дәүләт хакимияте һәм җирле үзидарә органнарының бердәм "Рәсми Татарстан" порталы (web-сайт)) [Найти текст](#)

Лаборатория экспериментальной фонетики

Если пользователю нужно выполнить поиск слов, находящихся не только рядом, но и разделенных другими словами, то для этого справа от слова в текстовом поле «**Расстояние**» можно указать необходимые числа. Например, числа «**1-1**» или просто «**1**» означают, что следующее слово расположено непосредственно после текущего, «**2-2**» или «**2**» означают, что они разделены одним словом и т.д. Если указать «**1-5**», то система будет искать все случаи, где искомые слова расположены как рядом, так и разделены одним, двумя, тремя и четырьмя словами. Максимально возможная дистанция между словами не ограничена технически, она определяется лишь длиной предложения.

П И С Ъ М Е Н Н Ы Й К О Р П У С Т А Т А Р С К О Г О Я З Ы К А

TAT
РУС
ENG

[на главную](#)

Главная

Поиск в Корпусе

Синтез татарской речи

Инструкция

Статистика

Описание корпуса

Использование корпуса

Публикации

Благодарим...

О проекте

Полезная информация

Поиск в Письменном корпусе татарского языка

Выберите тип поиска и введите необходимое слово:

Слово 1:	бик		Расстояние 1:	2-4
Слово 2:	алдылар		Расстояние 2:	1-1
Слово 3:			Расстояние 3:	1-1
Слово 4:			Расстояние 4:	1-1
Слово 5:			<input type="button" value="Найти!"/>	

Поиск коллокаций в контекстном (статистическом) корпусе по словоформе (например, **китапны, авылларга, килмәдәгез**)

Поиск в морфологически размеченном корпусе по лемме (например, **китап, авыл, кил**)

Поиск по шаблону (указывается начало или конец слова длиной 3 и более символов, например, **китап*, авыл*, *гез, *ләргә**)

Сложный морфологический поиск (Инструкция)

Лаборатория экспериментальной фонетики

Результаты:

П И С Ъ М Е Н Н Ы Й К О Р П У С Т А Т А Р С К О Г О Я З Ы К А










TAT RUS ENG

Главная
Поиск в Корпусе
Синтез татарской речи
Инструкция
Статистика
Описание корпуса
Использование корпуса
Публикации
Благодарим...
О проекте

Полезная информация

ЗАПРОС: {бик} 2-4 {алдылар} 1-1 {} 1-1 {} 1-1 {}

КОЛИЧЕСТВО СОВПАДЕНИЙ: 179

-  Японнар тупны һәм уен кырының үзәген **бик** тиз яулап **алдылар**.
(источник: "2013 Универсиадасы блогы" (web-сайт)) [Найти текст](#)
-  Безнең спортчыларыбыз монда яхшы чыгыш ясый, ләкин алар результатларын яхшыртуны күздә тоталар һәм шуның өчен дә монда **бик** шәп тәҗрибә **алдылар**.
(источник: "2013 Универсиадасы блогы" (web-сайт)) [Найти текст](#)
-  Аккредитация үзәгендә безне **бик** ачык каршы **алдылар**, документларны тикшерделәр, банк картасы белән бейджлар тапшырдылар.
(источник: "2013 Универсиадасы блогы" (web-сайт)) [Найти текст](#)
-  Корылманың персоналы күзәтүчеләр төркемен **бик** мөлаем каршы **алдылар** һәм үз объектларындагы функцияләре турында бик теләп сөйләде.
(источник: "2013 Универсиадасы блогы" (web-сайт)) [Найти текст](#)
-  "Безне монда **бик** әйбәт каршы **алдылар**, хезмәт күрсәтү бик югары дәрәҗәдә оештырылган.
(источник: "2013 Универсиадасы блогы" (web-сайт)) [Найти текст](#)
-  "Кызларга рәхмәт, **бик** яхшы каршы **алдылар**.
(источник: "2013 Универсиадасы блогы" (web-сайт)) [Найти текст](#)
-  Ут Эстафетасының фирма поезды Курск вокзалына килде, монда Бөтендөнья студентлар Уеннарының төп символын мәскәүлеләр фанфаралар тавышы астында **бик** жылы каршы **алдылар**.
(источник: "2013 Универсиадасы блогы" (web-сайт)) [Найти текст](#)
-  Йөзләрчә төмәнлеләр Эстафета колоннасын бар хәрәкәт маршрутында **бик** жылы каршы **алдылар**.
(источник: "2013 Универсиадасы блогы" (web-сайт)) [Найти текст](#)
-  Атташе булу өчен чит телләргә әйбәт белү генә аз, шуңа да Универсиада өчен аларны **бик** җентекләп сайлап **алдылар**.

Лаборатория экспериментальной фонетики

Как найти в Корпусе комбинации с указанием леммы?

В случае, если пользователь хочет искать не по конкретной словоформе, а по всем формам определенной леммы, то для этого лемму необходимо заключить в скобки, например, «(китап)». Если система не найдет в базе данную лемму, то поиск будет произведен по словоформе.

Выполним следующий запрос: «{теге} {(китап)}»:

П И С Ъ М Е Н Н Ы Й К О Р П У С Т А Т А Р С К О Г О Я З Ы К А

ТАТ РУС ENG

на главную

Главная

Поиск в Корпусе

Синтез татарской речи

Инструкция

Статистика

Описание корпуса

Использование корпуса

Публикации

Благодарим...

О проекте

Полезная информация

Поиск в Письменном корпусе татарского языка

Выберите тип поиска и введите необходимое слово:

Слово 1:	<input type="text" value="теге"/>		Расстояние 1:	<input type="text" value="1-1"/>
Слово 2:	<input type="text" value="(китап)"/>		Расстояние 2:	<input type="text" value="1-1"/>
Слово 3:	<input type="text"/>		Расстояние 3:	<input type="text" value="1-1"/>
Слово 4:	<input type="text"/>		Расстояние 4:	<input type="text" value="1-1"/>
Слово 5:	<input type="text"/>		<input type="button" value="Найти!"/>	

- Поиск коллокаций в контекстном (статистическом) корпусе по словоформе (например, **китапны**, **авылларга**, **килмәдегез**)
- Поиск в морфологически размеченном корпусе по лемме (например, **китап**, **авыл**, **кил**)
- Поиск по шаблону (указывается начало или конец слова длиной 3 и более символов, например, **китап***, **авыл***, ***гез**, ***ләргә**)
- Сложный морфологический поиск (Инструкция)

Авторы:
Сайхунов М.Р.
Ибрагимов Т.И.
tatcorpus@gmail.com

Лаборатория экспериментальной фонетики

Результат:

П И С Ь М Е Н Н Ы Й К О Р П У С Т А Т А Р С К О Г О Я З Ы К А




TAT RUS ENG

Главная
Поиск в Корпусе
Синтез татарской речи
Инструкция
Статистика
Описание корпуса
Использование корпуса
Публикации
Благодарим...
О проекте

Полезная информация

ЗАПРОС: {теге} 1-1 {китан} 1-1 {} 1-1 {} 1-1 {}

КОЛИЧЕСТВО СОВПАДЕНИЙ: 22

-  - Булат, **теге китан** синдэме эле? - диде Харис.
(источник: Заһит Мурсиев. Ачылмаган тәрәзә)
-  Балачагым күз алдыма килде, "Путивльдән Карпатка кадәр" дигән **теге китан та** искә төште.
(источник: "Роберт Миңнуллин (официаль сайт)" (web-сайт)) [Найти текст](#)
-  Әмма **теге китан** күңелдән китми генә.
(источник: "Роберт Миңнуллин (официаль сайт)" (web-сайт)) [Найти текст](#)
-  Тагын туп-туры **теге китан** торган бүлеккә ашыгасың.
(источник: "Роберт Миңнуллин (официаль сайт)" (web-сайт)) [Найти текст](#)
-  Кави, син минем **теге китанны** укыдыңмы?
(источник: Рабит Батулла. Тузга язылган хәлләр (Тарихи шәхесләребез, замандашларыбыз турында мәэжләр, хикәяләр, фаҗигале хәбәрләр))
-  Кави, син минем **теге китанны** укыдыңмы?
(источник: Рабит Батулла. Тузга язылган хәлләр (Тарихи шәхесләребез, замандашларыбыз турында мәэжләр, хикәяләр, фаҗигале хәбәрләр))
-  Хәер, теге төрек тарихын, **теге китанны** биш-алты кат укып чыктым.
(источник: Гаяз Исхакый. Тормышмы бу?)
-  - Каенанам хатынга, кияү **теге китабын** биреп торсын эле, мин дә укып чыгар идем, дигән.
(источник: "Мәдәни Жомга" газетасы (web-сайт)) [Найти текст](#)
-  Ярты сәгать тә үтми, томшыгына кыстырып **теге китан** битен алып та килә.
(источник: "Татарча текстлар, программалар" (web-сайт)) [Найти текст](#)
-  **Теге китан**, дәфтәрләр энә шул атларның аяк астында тапталып ятканнар.
(источник: "Әдипләр. Әхмәт Дусайлының шәхси порталы" (web-сайт)) [Найти текст](#)

Лаборатория экспериментальной фонетики

В качестве другого примера рассмотрим «{(авыл)} 1-3 {(кил)}». Здесь поиск производится по двум леммам с возможным расстоянием между словами до трех единиц.

П И С Ъ М Е Н Н Ы Й К О Р П У С Т А Т А Р С К О Г О Я З Ы К А

TAT RUS ENG

на главную

Главная

Поиск в Корпусе

Синтез татарской речи

Инструкция

Статистика

Описание корпуса

Использование корпуса

Публикации

Благодарим...

О проекте

Полезная информация

Поиск в Письменном корпусе татарского языка

Выберите тип поиска и введите необходимое слово:

Слово 1: (авыл)	<input type="text"/>		Расстояние 1: 1-3
Слово 2: (кил)	<input type="text"/>		Расстояние 2: 1-1
Слово 3:	<input type="text"/>		Расстояние 3: 1-1
Слово 4:	<input type="text"/>		Расстояние 4: 1-1
Слово 5:	<input type="text"/>		<input type="button" value="Найти!"/>

- Поиск коллокаций в контекстном (статистическом) корпусе по словоформе (например, **китапны, авылларга, килмәдегез**)
- Поиск в морфологически размеченном корпусе по лемме (например, **китап, авыл, кил**)
- Поиск по шаблону (указывается начало или конец слова длиной 3 и более символов, например, **китап*, авыл*, *гез, *ләргә**)
- Сложный морфологический поиск (Инструкция)

Авторы:
Сайхунов М.Р.
Ибрагимов Т.И.
tatcorpus@gmail.com

Лаборатория экспериментальной фонетики

П И С Ъ М Е Н Н Ы Й К О Р П У С Т А Т А Р С К О Г О Я З Ы К А

TAT
РУС
ENG

ЗАПРОС: {авыл} 1-3 {кил} 1-1 {} 1-1 {} 1-1 {}

КОЛИЧЕСТВО СОВПАДЕНИЙ: 4302

Главная

Поиск в Корпусе

Синтез татарской речи

Инструкция

Статистика

Описание корпуса

Использование корпуса

Публикации

Благодарим...

О проекте

Полезная информация

Авторы:

Сайхунов М.Р.

Ибрагимов Т.И.

tatcorpus@gmail.com

- ▶ – Мәсәде **авылына** язучылар **килгән**, барабызмы?
(источник: Бари ИСЛАМ, БИК ГАДИ ДӘ, СЕРЛЕ ДӘ СИН, ТОРМЫШ)
- ▶ Чөнки якын тирә **авылларга килеп** утырган Гәрәйләр, Меңнәр, Кара бәк нәселләренең шәжәрәләрендәгә кешеләр исемнәре бу авылларда әле дә байтак.
(источник: Илдус Хужин. Топонимнар)
- ▶ Урта Бәяк турында 17 гасыр документларында мәгълүматлар юк, ахрысы **авыл** соңрак барлыкка **килгән**.
(источник: Илдус Хужин. Топонимнар)
- ▶ Бу очракта **авылның** исеме **килеп** чыгышы башка күп кенә авылларныкына туры килә, чөнки Кантуган дигән шәхес тарихта бар.
(источник: Илдус Хужин. Топонимнар)
- ▶ Әлбәттә, коңгырат кабиләләреннән тыш бу **авылда** борынгыдан **килгән** төрки ырулар яшәгәндер һәм дә 400-500 еллар элек көчле гәрәйләр һәм Кара бәк кабиләләре анда килеп утырып яңа исем биргәннәр булырга тиеш.
(источник: Илдус Хужин. Топонимнар)
- ▶ Әйтергә кирәк, урыслар **авылга** соңрак **килеп** урнашкан булырга тиешләр, чөнки алар татар авылы белән аның зираты арасына кереп утырганнар.
(источник: Илдус Хужин. Топонимнар)
- ▶ Шулай итеп Уралда кыпчакларның һәм Кара бәк нәселенең бик күп **авыллары** дөньяга **килә**.
(источник: Илдус Хужин. Топонимнар)
- ▶ Ә документлар күрсәтүенчә, 1647 елда Өфе елгасы буендагы Өфе Шигер һәм Әртә Шигер **авыллары** барлыкка **килгән**.
(источник: Илдус Хужин. Топонимнар)

Лаборатория экспериментальной фонетики

Как найти в Корпусе комбинации с указанием морфологических (грамматических) тегов?

Ознакомиться со списком поддерживаемых Корпусом тегов можно пройдя по следующим ссылкам:

- 1) http://corpus.tatar/index.php?openinframe=manual/tags_uniq.pdf
- 2) <https://sourceforge.net/p/apertium/svn/61954/tree//branches/turkic/lexc.rkt>

Поиск возможен как по одному, так и по нескольким тегам. Можно использовать только комбинации типа «и», но не «или», т.е. нельзя одновременно указать взаимоисключающие теги, например «<dat><acc>» (*дательный (направительный) падеж, винительный падеж*).

Давайте произведем поиск по комбинации «{<adj>} 1-2 {<n><dat>} 1-3 {<v><past>}». Первое слово должно быть *прилагательным* (<adj>), после него на расстоянии до двух слов идет *существительное* (<n>) в *дательном (направительном) падеже* (<dat>), а после него на расстоянии до трех слов идет *глагол* (<v>) *прошедшего времени на «-ган/-гән/-кан/-кән»* (<past>).

П И С Ъ М Е Н Н Ы Й К О Р П У С Т А Т А Р С К О Г О Я З Ы К А

Т А Т Р У С Е N G

[на главную](#)

Главная

Поиск в Корпусе

Синтез татарской речи

Инструкция

Статистика

Описание корпуса

Использование корпуса

Публикации

Благодарим...

О проекте

Полезная информация

Поиск в Письменном корпусе татарского языка

Выберите тип поиска и введите необходимое слово:

Слово 1:	<input type="text" value="<adj>"/>		Расстояние 1:	<input type="text" value="1-2"/>
Слово 2:	<input type="text" value="<n><dat>"/>		Расстояние 2:	<input type="text" value="1-3"/>
Слово 3:	<input type="text" value="<v><past>"/>		Расстояние 3:	<input type="text" value="1-1"/>
Слово 4:	<input type="text"/>		Расстояние 4:	<input type="text" value="1-1"/>
Слово 5:	<input type="text"/>		<input type="button" value="Найти!"/>	

- Поиск коллокаций в контекстном (статистическом) корпусе по словоформе (например, **китапны, авылларга, килмәдегез**)
- Поиск в морфологически размеченном корпусе по лемме (например, **китап, авыл, кил**)
- Поиск по шаблону (указывается начало или конец слова длиной 3 и более символов, например, **китап*, авыл*, *гез, *ләргә**)
- Сложный морфологический поиск (Инструкция)

Лаборатория экспериментальной фонетики

П И С Ъ М Е Н Н Ы Й К О Р П У С Т А Т А Р С К О Г О Я З Ы К А

ТАТ
РУС
ENG

ЗАПРОС: {<adj>} 1-2 {<n>,<dat>} 1-3 {<v>,<past>} 1-1 {} 1-1 {}

КОЛИЧЕСТВО СОВПАДЕНИЙ: 10594

Главная

Поиск в Корпусе

Синтез татарской речи

Инструкция

Статистика

Описание корпуса

Использование корпуса

Публикации

Благодарим...

О проекте

Полезная информация

Авторы:

Сайхунов М.Р.
Ибрагимов Т.И.
tatcorpus@gmail.com

▶ Авылларыбызның килеп чыгуы вакыты турында фикер йөрткәндә, шуларны онытмау мәгъкуль: башкорт галимнәренең язмалары буенча, бөтен татар авыллары да 1650 – 1750 елларда һәм **соңрак барлыкка килгәннәр**.
(источник: Илдус Хужин. Топонимнар)

▶ Урта Бәяк турында 17 гасыр документларында мәгълүматлар юк, ахрысы авыл **соңрак барлыкка килгән**.
(источник: Илдус Хужин. Топонимнар)

▶ **Соңрак бу урынга** башкалар да **килгән**.
(источник: Илдус Хужин. Топонимнар)

▶ Бу авыл башкортлары (ягъни жирбиләүчеләре, вотчинники) **соңрак** Әртә **заводына** жирләрен **сатканнар** һәм үзләре шул ук жирләрдә яшиләр дип язылган бер документта.
(источник: Илдус Хужин. Топонимнар)

▶ Авыл халкының байтагы күмер яндырып **якындагы заводларга** ташып **сатканнар**.
(источник: Илдус Хужин. Топонимнар)

▶ **Ак шәлләре жиргә** шуып **төшкән...**
(источник: Разил Вәлиев. Шигырьләр, Жырлар)

▶ 21 яшьлек шагыйрь 1907 елда үз арамыздан чыккан маһир рәссам хакында хыяланганда, андый рәссам инде хәзерге Татарстанның Буа районы (элеккеге Тәтеш өязе) Күл Черкене авылында 1897 елның 22 февралендә Мәхжүбә абыстай һәм Идрис хәзрәт Урманчелар гаиләсендә **якты дөньяга** килгән **булган**.
(источник: "Татарстан китап нәширияты" (web-сайт)) [Найми текст](#)

▶ Соңрак рус тарихчылары, мәгънәне берәз тарайтып, барлык төркиләргә татар дип атаганнар, **төрки** сүзе **урынына** татар сүзен **кулланганнар**.
(источник: "Татарстан китап нәширияты" (web-сайт)) [Найми текст](#)

Лаборатория экспериментальной фонетики

В следующем примере производится поиск двух смежных слов, при этом первое слово должно иметь атрибут категории *принадлежности (притяжательности) 2 лица единственного числа (<px2sg>)*, а второе слово должно быть в *винительном падеже (<acc>)*.

ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА

ТАТ РУС ENG

на главную

Главная

Поиск в Корпусе

Синтез татарской речи

Инструкция

Статистика

Описание корпуса

Использование корпуса

Публикации

Благодарим...

О проекте

Полезная информация

Поиск в Письменном корпусе татарского языка

Выберите тип поиска и введите необходимое слово:

Слово 1:	<input type="text" value="<px2sg>"/>		Расстояние 1:	<input type="text" value="1-1"/>
Слово 2:	<input type="text" value="<acc>"/>		Расстояние 2:	<input type="text" value="1-1"/>
Слово 3:	<input type="text"/>		Расстояние 3:	<input type="text" value="1-1"/>
Слово 4:	<input type="text"/>		Расстояние 4:	<input type="text" value="1-1"/>
Слово 5:	<input type="text"/>		<input type="button" value="Найти!"/>	

- Поиск коллокаций в контекстном (статистическом) корпусе по словоформе (например, **китапны, авылларга, килмәдегез**)
- Поиск в морфологически размеченном корпусе по лемме (например, **китап, авыл, кил**)
- Поиск по шаблону (указывается начало или конец слова длиной 3 и более символов, например, **китап*, авыл*, *гез, *ләргә**)
- Сложный морфологический поиск (Инструкция)

Авторы:
Сайхунов М.Р.
Ибрагимов Т.И.
tatcorpus@gmail.com

Лаборатория экспериментальной фонетики

П И С Ъ М Е Н Н Ы Й К О Р П У С Т А Т А Р С К О Г О Я З Ы К А

ТАТ РУС ENG

Главная
Поиск в Корпусе
Синтез татарской речи
Инструкция
Статистика
Описание корпуса
Использование корпуса
Публикации
Благодарим...
О проекте
Полезная информация

Авторы:
Сайхунов М.Р.
Ибрагимов Т.И.
tatcorpus@gmail.com

ЗАПРОС: {<px2sg>} 1-1 {<acc>} 1-1 {} 1-1 {} 1-1 {}

КОЛИЧЕСТВО СОВПАДЕНИЙ: 4954

- ▶

Шуңа күрә, күптән инде икенче халыкка әйләнгән үз **халкыңның эзләрен** эзләү файдасыз.

(источник: Илдус Хужин. Топонимнар)
- ▶

Менә шул үзсүзлегең, кеше сүзен, **анаң сүзен** тыңламавың харап итә яздымы үзеңне?

(источник: Фәрит Яхин. СЕРЛЕ КАМКА)
- ▶

Була шундый чаклар, син якын **дустыңның ялганын** тыңлыйсың.

(источник: "Татарстан китап нәширияты" (web-сайт)) [Найми текст](#)
- ▶

Такта өстенә **жым саласың да ятьмәне** бакчага чыгарып куясың.

(источник: "Татарстан китап нәширияты" (web-сайт)) [Найми текст](#)
- ▶

Тик, зинһар өчен, шаярышмагыз анда, рәссам **абыйларыгызының кушканын** төгәл үтәгез!

(источник: "Татарстан китап нәширияты" (web-сайт)) [Найми текст](#)
- ▶

- Их сез, яшел чебешләр, яшь **чагыгызының кадерен** белеп калыгыз.

(источник: Заһит Мурсиев. Мин сезнең хатыңыгыз)
- ▶

– И улым, болары гына аның бик вак мәсьәлә, артык **исең китмәсен**, менә үз көнеңне үзең күрә башлагач, дөнья әле уңны–суңны яңаклар...

(источник: РАВИЛ ӘМИРХАН. ӘМИРХАННАР (тәфсилле шәһәрә). КАЗАН – 2005)
- ▶

Дүшәмбе көн безнең уртақ кичәдә өзелгән **аягыңның тарихын** сөйләрсен, Зәңгәр чышмә буйлары мондый хәлләрне иштергә бик сусаганнар... ярыймы? — дип, миннән кат-кат вәгдәләр алгач, шулашып чыгып киттеләр.

(источник: ГАЛИМЖАН ИБРАҖИМОВ. КЫЗЫЛ ЧӘЧӘКЛӘР)
- ▶

Ә инде семинар актив аралашуны, балалар белән аралашу өчен үз **программаңны формалаштыруны**, яңа методлар, технологияләр табуны күздә тотканың, бу мөгаен, педагогика белеменең иң зур нәтижәседер.

(источник: "2013 Универсиадасы блогы" (web-сайт)) [Найми текст](#)

Лаборатория экспериментальной фонетики

Как найти в Корпусе комбинации с указанием начала, середины и/или конца слова?

Для поиска всех слов, начинающихся с определенного сочетания букв используется шаблон вида «**ки***». В этом случае будут найдены такие словоформы, как «**китап, китапны, киштәгә, кит...**».

Поиск по середине слова возможен в случае использования запроса «***әме***»: **керәмен, әмер, үсәме...**

Шаблон поиска по концу слова выглядит следующим образом: «***рны**». В итоге получим примеры «**дусларны, тарны, барны, кулларны...**».

Давайте произведем поиск со следующими параметрами: «**{ил*} 1-1 {белән}**», т.е. первое слово должно начинаться на «**ил**», а вторым словом является «**белән**».

П И С Ъ М Е Н Н Ы Й К О Р П У С Т А Т А Р С К О Г О Я З Ы К А

ТАТ РУС ENG

на главную

Главная

Поиск в Корпусе

Синтез татарской речи

Инструкция

Статистика

Описание корпуса

Использование корпуса

Публикации

Благодарим...

О проекте

Полезная информация

Поиск в Письменном корпусе татарского языка

Выберите тип поиска и введите необходимое слово:

Слово 1:	<input type="text" value="ил*"/>		Расстояние 1:	<input type="text" value="1-1"/>
Слово 2:	<input type="text" value="белән"/>		Расстояние 2:	<input type="text" value="1-1"/>
Слово 3:	<input type="text"/>		Расстояние 3:	<input type="text" value="1-1"/>
Слово 4:	<input type="text"/>		Расстояние 4:	<input type="text" value="1-1"/>
Слово 5:	<input type="text"/>		<input type="button" value="Найти!"/>	

- Поиск коллокаций в контекстном (статистическом) корпусе по словоформе (например, **китапны, авылларга, килмәдегез**)
- Поиск в морфологически размеченном корпусе по лемме (например, **китап, авыл, кил**)
- Поиск по шаблону (указывается начало или конец слова длиной 3 и более символов, например, **китап*, авыл*, *гез, *ләргә**)
- Сложный морфологический поиск (Инструкция)

Авторы:
Сайхунов М.Р.
Ибрагимов Т.И.
tatcorpus@gmail.com

Лаборатория экспериментальной фонетики

Результат:

П И С Ъ М Е Н Н Ы Й К О Р П У С Т А Т А Р С К О Г О Я З Ы К А

ТАТ РУС ENG ЗАПРОС: {ил} 1-1 {белән} 1-1 {} 1-1 {} 1-1 {}

КОЛИЧЕСТВО СОВПАДЕНИЙ: 2695

▶ Элеге масштаблы проектны тормышка ашыру Татарстан башкаласын үстерүгә, аның Россия төбәкләре белән элемтэләрен ныгытуга зур этәргеч булды, халыкта патриотизм һәм **ил белән** хорурлык хисе артты.
(источник: "2013 Универсиадасы блогы" (web-сайт)) [Найти текст](#)

▶ Яңа ел театральләштерелгән тамашасында барс балаларының ничек итеп Универсиада Утын эзләп төрле континентлар буенча сәяхәт итүләре, чит **илләр белән** танышулары, төрле мажараларга тарулары, төрле **һөнәрләргә өйрәнүләре һәм ниһаять, Универсиада Утын кулга төшерүләре** тасвирланды.
(источник: "2013 Универсиадасы блогы" (web-сайт)) [Найти текст](#)

▶ Универсиадада Россия медальләр өчен Кытай, АКШ кебек **илләр белән** көч сынашчак.
(источник: "2013 Универсиадасы блогы" (web-сайт)) [Найти текст](#)

▶ Матди стимуллардан тыш, волонтерлар өчен бик киң күңел ачу программасы да эзерлэнгән: Универсиада **Илчеләре белән** иртэнге аш һәм очрашулар, һәр кичтә МЕГАда кичеләр һәм Уеннар тәмамланганнан соң бик якты зур йомгаклау бәйрәме.
(источник: "2013 Универсиадасы блогы" (web-сайт)) [Найти текст](#)

▶ Бу максат, әлбәттә, башка **илләр белән** очрашуда да кабатлана.
(источник: "2013 Универсиадасы блогы" (web-сайт)) [Найти текст](#)

▶ - Универсиада **илчеләре белән** килешүләр ничек бара?
(источник: "2013 Универсиадасы блогы" (web-сайт)) [Найти текст](#)

▶ 2013 елгы Универсиаданың кайда үтәчәге турындагы карарны берничә ел элек бик дулкынланып, тын да алмый **илебез белән** көткән идея.
(источник: "2013 Универсиадасы блогы" (web-сайт)) [Найти текст](#)

▶ "Казан 2013" волонтерлары өчен Универсиада **Илчеләре** белән күзгә-күз сөйләшү эш сменасына килү кебек үк гади нәрсә: нәкъ менә объектлардагы буш вакытларында егетләргә һәм кызларны теге яки бу спорт

Лаборатория экспериментальной фонетики

Чтобы задать поиск и по началу, и по концу слова, то знак * нужно расположить между ними: «**ал*лар**» (найдутся такие слова, как «**алалар, алдылар, алмалар, алмагачлар...**»).

Для поиска по началу, середине и концу слова одновременно запрос оформляется в виде «**к*әме*н**»: **керәмен, каләмен, күләменнән, кияүдәмен...**

В то время, как знак звездочка «*» совпадает с любым количеством (включая ноль) символов, знак вопроса «?» используется для замены любого одиночного символа. Например, «**т?з***» совпадет с словами «**тиз, тозны, түзде, тазарды...**», но не с «**тигез, тугызны, тәрәзә...**».

Все перечисленные параметры поиска (словоформа, лемма, грамматические теги, начало и конец слова) можно всячески комбинировать в разных полях. Например, запрос такого вида: «{<prn>} 1-1 {(кеше)} 1-1 {ал*}». Здесь первое слово должно быть *местоимением* (<prn>), следующее за ним является *леммой* «кеше», а после него находится *слово, начинающееся на «ал»*:

ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА

ТАТ РУС ENG

[на главную](#)

Главная

Поиск в Корпусе

Синтез татарской речи

Инструкция

Статистика

Описание корпуса

Использование корпуса

Публикации

Благодарим...

О проекте

Полезная информация

Поиск в Письменном корпусе татарского языка

Выберите тип поиска и введите необходимое слово:

Слово 1:	<input type="text" value="<prn>"/>		Расстояние 1:	<input type="text" value="1-1"/>
Слово 2:	<input type="text" value="(кеше)"/>		Расстояние 2:	<input type="text" value="1-1"/>
Слово 3:	<input type="text" value="ал*"/>		Расстояние 3:	<input type="text" value="1-1"/>
Слово 4:	<input type="text"/>		Расстояние 4:	<input type="text" value="1-1"/>
Слово 5:	<input type="text"/>		<input type="button" value="Найти!"/>	

- Поиск коллокаций в контекстном (статистическом) корпусе по словоформе (например, **китапны, авылларга, килмәдегез**)
- Поиск в морфологически размеченном корпусе по лемме (например, **китап, авыл, кил**)
- Поиск по шаблону (указывается начало или конец слова длиной 3 и более символов, например, **китап*, авыл*, *гез, *ләргә**)
- Сложный морфологический поиск (Инструкция)

Лаборатория экспериментальной фонетики

П И С Ъ М Е Н Н Ы Й К О Р П У С Т А Т А Р С К О Г О Я З Ы К А

TAT RUS ENG

Главная
Поиск в Корпусе
Синтез татарской речи
Инструкция
Статистика
Описание корпуса
Использование корпуса
Публикации
Благодарим...
О проекте

Полезная информация

ЗАПРОС: {<prn>} 1-1 {кеше} 1-1 {ал} 1-1 {} 1-1 {}

КОЛИЧЕСТВО СОВПАДЕНИЙ: 150

- ▶ **Андый кеше Аллаһы Тәгалә** тарафыннан бирелә, лидер мохит эчендә тәрбияләнә ала.
(источник: "Үзөбөз" яңа буын хәрәкәте (web-сайт)) [Найти текст](#)
- ▶ **Аларын кешеләр ала.**
(источник: "Безнең гәҗит" газетасы (web-сайт)) [Найти текст](#)
- ▶ Бу хатын әле дә **шулай кеше алдап** йөри икән дип бик аптырадым.
(источник: "Безнең гәҗит" газетасы (web-сайт)) [Найти текст](#)
- ▶ Тагын бер хәдистә Пәйгамбәребез (с.г.в.) әйтә: «Дөрөслектә, **шул кеше Аллаһының** сөекле колы була, кем намаз укыганда сул кулы өстенә уң кулын куйса, сәхәрне соңгарак калдырып, ифтарны вакытында кылса».

(источник: "Безнең гәҗит" газетасы (web-сайт)) [Найти текст](#)
- ▶ **Болары да кешегә Аллаһы Тәгалә** биргән сыйфатлар һәм алар икесе дә кирәк, әмма үз үлчәмнәрендә булырга тиешләр.
(источник: "Безнең гәҗит" газетасы (web-сайт)) [Найти текст](#)
- ▶ **Мондый кешеләр алдында**, гомумән, баш ияргә дә була, минемчә.
(источник: "Безнең гәҗит" газетасы (web-сайт)) [Найти текст](#)
- ▶ “Һәркем күрөп торды: бу кеше хоккейны ярата, **мондый кешеләргә алкышларны** жәлләмәскә кирәк”, – дип белдерде ул.
(источник: "Безнең гәҗит" газетасы (web-сайт)) [Найти текст](#)
- ▶ **Мондый кешеләргә Аллаһы Тәгалә** кешеләргә ярдәм иткән кешенеке чаклы савап яза.
(источник: "Безнең гәҗит" газетасы (web-сайт)) [Найти текст](#)
- ▶ Бер безне генә түгел, ничә еллар **шулай кешеләр алдап** йөргән икән ул.
(источник: "Безнең гәҗит" газетасы (web-сайт)) [Найти текст](#)

Лаборатория экспериментальной фонетики

Как найти в Корпусе комбинации с указанием группы параметров для одного слова?

Каждое слово в поиске может характеризоваться различным сочетанием таких параметров, как словоформа, лемма, теги, начало, середина и конец слова.

Рассмотрим в качестве примера следующую ситуацию: нам нужно найти словоформу «алма» (глагол, *не бери*), однако в результат поиска будут включены также предложения со словом «алма» (существительное, *яблоко*). Для того, чтобы система искала только слово «алма» в качестве глагола, необходимо оформить запись в виде «алма<v>». Здесь тег «<v>» означает *глагол*.

П И С Ъ М Е Н Н Ы Й К О Р П У С Т А Т А Р С К О Г О Я З Ы К А

ТАТ РУС ENGна главную

Главная

Поиск в Корпусе

Синтез татарской речи

Инструкция

Статистика

Описание корпуса

Использование корпуса

Публикации

Благодарим...

О проекте

Полезная информация

Поиск в Письменном корпусе татарского языка

Выберите тип поиска и введите необходимое слово:

Слово 1:	<input style="border: 1px solid red;" type="text" value="алма<v>"/>		Расстояние 1:	<input type="text" value="1-1"/>
Слово 2:	<input type="text"/>		Расстояние 2:	<input type="text" value="1-1"/>
Слово 3:	<input type="text"/>		Расстояние 3:	<input type="text" value="1-1"/>
Слово 4:	<input type="text"/>		Расстояние 4:	<input type="text" value="1-1"/>
Слово 5:	<input type="text"/>		<input type="button" value="Найти!"/>	

Поиск коллокаций в контекстном (статистическом) корпусе по словоформе (например, **китапны**, **авылларга**, **килмәдегез**)

Поиск в морфологически размеченном корпусе по лемме (например, **китап**, **авыл**, **кил**)

Поиск по шаблону (указывается начало или конец слова длиной 3 и более символов, например, **китап***, **авыл***, ***гез**, ***ләргә**)

Сложный морфологический поиск (Инструкция, Список тегов)

Авторы:

Сайхунов М.Р.
Ибрагимов Т.И.
tatcorpus@gmail.com

Лаборатория экспериментальной фонетики

В данном случае имеется другой способ решения данной проблемы. В качестве второго аргумента можно указать лемму: «алма(ал)». Т.е. система будет искать словоформу «алма», у которой леммой является «ал». В этом случае система при поиске исключит существительное «алма», потому что ее леммой является «алма».

П И С Ъ М Е Н Н Ы Й К О Р П У С Т А Т А Р С К О Г О Я З Ы К А

TAT RUS ENG на главную

Главная
Поиск в Корпусе
Синтез татарской речи
Инструкция
Статистика
Описание корпуса
Использование корпуса
Публикации
Благодарим...
О проекте
Полезная информация

Поиск в Письменном корпусе татарского языка

Выберите тип поиска и введите необходимое слово:

Слово 1:	<input type="text" value="алма(ал)"/>			Расстояние 1:	<input type="text" value="1-1"/>
Слово 2:	<input type="text"/>			Расстояние 2:	<input type="text" value="1-1"/>
Слово 3:	<input type="text"/>			Расстояние 3:	<input type="text" value="1-1"/>
Слово 4:	<input type="text"/>			Расстояние 4:	<input type="text" value="1-1"/>
Слово 5:	<input type="text"/>			<input type="button" value="Найти!"/>	

Поиск коллокаций в контекстном (статистическом) корпусе по словоформе (например, **китапны, авылларга, килмәдегез**)

Поиск в морфологически размеченном корпусе по лемме (например, **китап, авыл, кил**)

Поиск по шаблону (указывается начало или конец слова длиной 3 и более символов, например, **китап*, авыл*, *гез, *ләргә**)

Сложный морфологический поиск (Инструкция, Список тегов)

Лаборатория экспериментальной фонетики

Ввод поисковых параметров в графическом режиме

В целях облегчения ввода поисковых параметров, имеется специальный инструмент, доступ к которому можно получить нажав на указанные кнопки справа от текстовых полей:

П И С Ъ М Е Н Н Ы Й К О Р П У С Т А Т А Р С К О Г О Я З Ы К А

ТАТ РУС ENG

на главную

Главная

Поиск в Корпусе

Синтез татарской речи

Инструкция

Статистика

Описание корпуса

Использование корпуса

Публикации











Благодарим...

О проекте

Полезная информация

Поиск в Письменном корпусе татарского языка

Выберите тип поиска и введите необходимое слово:

Слово 1:	<input type="text"/>	 	Расстояние 1:	<input type="text" value="1-1"/>
Слово 2:	<input type="text"/>	 	Расстояние 2:	<input type="text" value="1-1"/>
Слово 3:	<input type="text"/>	 	Расстояние 3:	<input type="text" value="1-1"/>
Слово 4:	<input type="text"/>	 	Расстояние 4:	<input type="text" value="1-1"/>
Слово 5:	<input type="text"/>	 	<input type="button" value="Найти!"/>	

Поиск коллокаций в контекстном (статистическом) корпусе по словоформе (например, **китапны**, **авылларга**, **килмөдегез**)

Поиск в морфологически размеченном корпусе по лемме (например, **китап**, **авыл**, **кил**)

Поиск по шаблону (указывается начало или конец слова длиной 3 и более символов, например, **китап***, **авыл***, ***гез**, ***ләргә**)

Сложный морфологический поиск (Инструкция, Список тегов)

Авторы:

Сайхунов М.Р.
Ибрагимов Т.И.
tatcorpus@gmail.com

Лаборатория экспериментальной фонетики

Открывающееся при этом окно имеет следующий вид:

П И С Ъ М Е Н Н Ы Й К О Р П У С Т А Т А Р С К О Г О Я З Ы К А

TAT RUS ENG

Главная
Поиск в Корпусе
Синтез татарской речи
Инструкция
Статистика
Описание корпуса
Использование корпуса
Публикации
Благодарим...
О проекте
Полезная информация

Авторы:
Сайдунов М.Р.
Ибрагимов Т.И.
tatcorpus@gmail.com

СЛОВО #1 **Словоформа:**
Лемма:

Расстояние до следующего слова равно / от **до**

Грамматические признаки

! Part of speech categories = First-level tags

- <n> ! Noun ! Исем ! Существительное
- <np> ! Proper noun ! Ялгызлык исем ! Имя собственное
- <adj> ! Adjective ! Сыйфат ! Прилагательное
- <num> ! Numeral ! Сан ! Числительное
- <prn> ! Pronoun ! Алмашлык ! Местоимение
- <det> ! Determiner ! Детерминатив ! Детерминатив
- <v> ! Verb ! Фигыль ! Глагол
- <vaux> ! Auxiliary verb ! Ярдәмче фигыль ! Вспомогательный глагол
- <adv> ! Adverb ! Рәвеш ! Наречие
- <post> ! Postposition ! Бәйләк ! Послелог
- <postadv> ! Postadverb ! Пострәвеш ! Посленаречие (Ачыклаган сүзләренәң артыннан килүче рәвешләр (кисәкчәләр) !# гына, ук)
- <cnjcoo> ! Coordinating conjunction ! Тезүче теркәгеч ! Сочинительный союз
- <cnjsub> ! Subordinating conjunction ! Ияртүче теркәгеч ! Подчинительный союз
- <cnjadv> ! Adverbial conjunction ! Рәвеш-теркәгеч ! Наречие-союз
- <ij> ! Interjection ! Ымлык ! Междометие
- <abbr> ! Abbreviation ! Кыскартылма ! Аббревиатура
- <cop> ! Copula ! Копула ! Копула
- <ideo> ! Ideophone ! Аваз ияртеме ! Звукоподражательное слово

Лаборатория экспериментальной фонетики

В левом верхнем углу указан номер текущего слова. Справа от него размещены две кнопки, которые позволяют быстро перейти на редактирование предыдущего или следующего слова. Следует учитывать, что при переходе данные на текущей странице автоматически сохраняются!

П И С Ъ М Е Н Н Ы Й К О Р П У С Т А Т А Р С К О Г О Я З Ы К А

TAT PUC ENG

СЛОВО #1 ← →

Словоформа:

Лемма:

Расстояние до следующего слова равно / от до

Грамматические признаки

! Part of speech categories = First-level tags

- <n> ! Noun ! Исем ! Существительное
- <np> ! Proper noun ! Ялгызлык исем ! Имя собственное
- <adj> ! Adjective ! Сыйфат ! Прилагательное
- <num> ! Numeral ! Сан ! Числительное
- <prn> ! Pronoun ! Алмашлык ! Местоимение
- <det> ! Determiner ! Детерминатив ! Детерминатив
- <v> ! Verb ! Фигыль ! Глагол
- <vaux> ! Auxiliary verb ! Ярдәмче фигыль ! Вспомогательный глагол
- <adv> ! Adverb ! Рәвеш ! Наречие
- <post> ! Postposition ! Бәйләк ! Послелог
- <postadv> ! Postadverb ! Пострәвеш ! Посленаречие (Ачыклаган сүзләренәң артыннан килүче рәвешләр (кисәкчәләр) !# гына, ук)
- <cnjcoo> ! Coordinating conjunction ! Тезүче теркәгеч ! Сочинительный союз
- <cnjsub> ! Subordinating conjunction ! Ияргүче теркәгеч ! Подчинительный союз
- <cnjadv> ! Adverbial conjunction ! Рәвеш-теркәгеч ! Наречие-союз
- <ij> ! Interjection ! Ымлык ! Междометие
- <abbr> ! Abbreviation ! Кыскартылма ! Аббревиатура
- <cop> ! Copula ! Копула ! Копула
- <ideo> ! Ideophone ! Аваз ияртеме ! Звукоподражательное слово

Лаборатория экспериментальной фонетики

На рисунке красным выделена область, где можно указать словоформу, лемму и расстояние. Необходимо отметить, что здесь лемма указывается без скобок!

П И С Ъ М Е Н Н Ы Й К О Р П У С Т А Т А Р С К О Г О Я З Ы К А

TAT RUS ENG

Главная
Поиск в Корпусе
Синтез татарской речи
Инструкция
Статистика
Описание корпуса
Использование корпуса
Публикации
Благодарим...
О проекте
Полезная информация

Авторы:
Сайдунов М.Р.
Ибрагимов Т.И.
tatcorpus@gmail.com

СЛОВО #1

Словоформа:

Лемма:

Расстояние до следующего слова равно / от **до**

Грамматические признаки

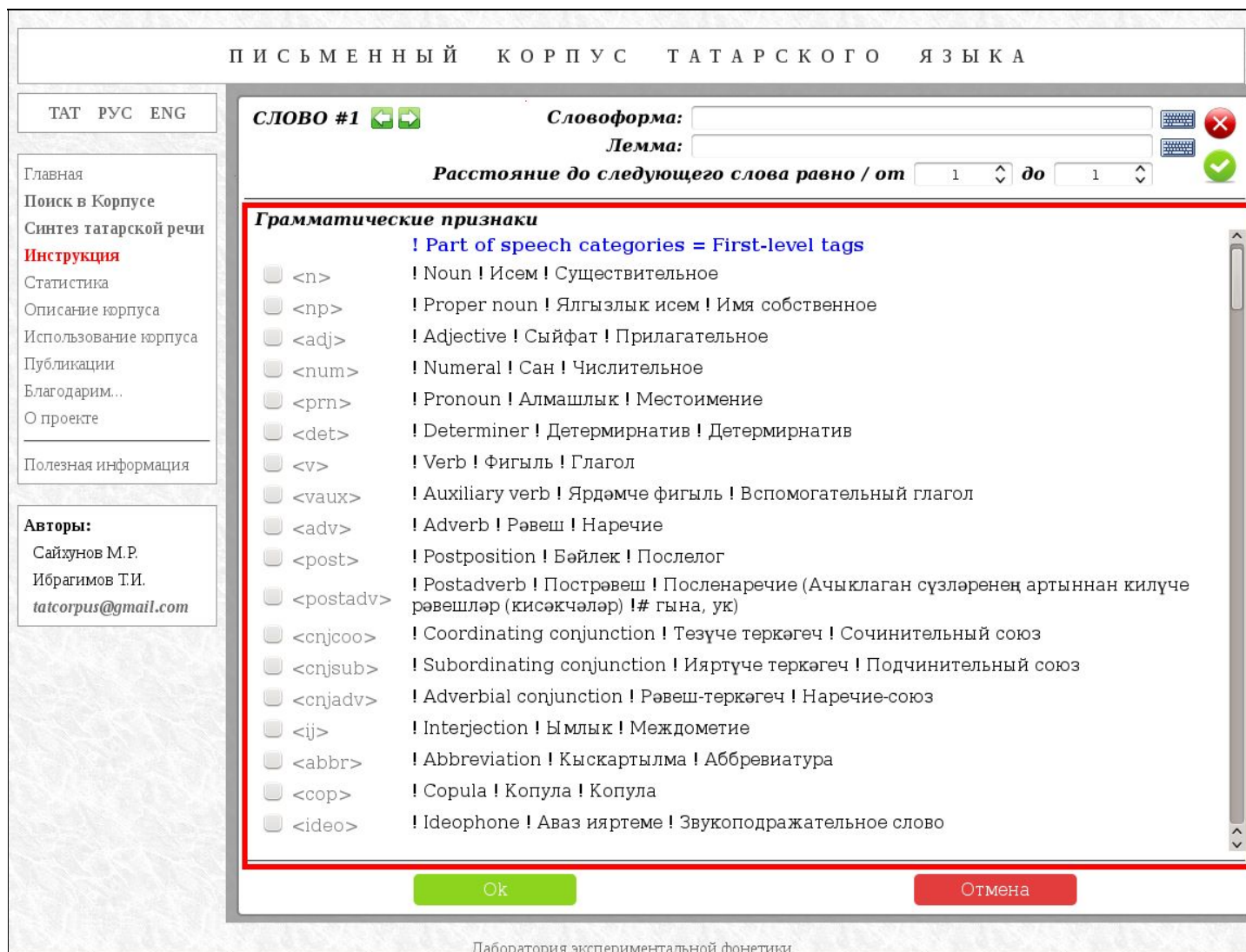
! Part of speech categories = First-level tags

- <n> ! Noun ! Исем ! Существительное
- <np> ! Proper noun ! Ялгызлык исем ! Имя собственное
- <adj> ! Adjective ! Сыйфат ! Прилагательное
- <num> ! Numeral ! Сан ! Числительное
- <prn> ! Pronoun ! Алмашлык ! Местоимение
- <det> ! Determiner ! Детерминатив ! Детерминатив
- <v> ! Verb ! Фигыль ! Глагол
- <vaux> ! Auxiliary verb ! Ярдәмче фигыль ! Вспомогательный глагол
- <adv> ! Adverb ! Рәвеш ! Наречие
- <post> ! Postposition ! Бәйләк ! Послелог
- <postadv> ! Postadverb ! Пострәвеш ! Посленаречие (Ачыклаган сүзләренәң артыннан килүче рәвешләр (кисәкчәләр) !# гына, ук)
- <cnjcoo> ! Coordinating conjunction ! Тезүче теркәгеч ! Сочинительный союз
- <cnjsub> ! Subordinating conjunction ! Ияртүче теркәгеч ! Подчинительный союз
- <cnjadv> ! Adverbial conjunction ! Рәвеш-теркәгеч ! Наречие-союз
- <ij> ! Interjection ! Ымлык ! Междометие
- <abbr> ! Abbreviation ! Кыскартылма ! Аббревиатура
- <cop> ! Copula ! Копула ! Копула
- <ideo> ! Ideophone ! Аваз ияртеме ! Звукоподражательное слово

Ok Отмена

Лаборатория экспериментальной фонетики

В основной области окна расположена область выбора грамматических признаков. Список тегов представлен по категориям. Выбор того или иного тега осуществляется путем указания «галочки» в соответствующей строке.



В правом верхнем углу и внизу имеются кнопки закрытия окна с сохранением (зеленые) и без сохранения (красные) изменений.



ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА

TAT РУС ENG

Главная
Поиск в Корпусе
Синтез татарской речи
Инструкция
Статистика
Описание корпуса
Использование корпуса
Публикации
Благодарим...
О проекте

Полезная информация

Авторы:
Сайдунов М.Р.
Ибрагимов Т.И.
tatcorpus@gmail.com

СЛОВО #1   **Словоформа:**
Лемма:
Расстояние до следующего слова равно / от **до**

Грамматические признаки

! Part of speech categories = First-level tags

- <n> ! Noun ! Исем ! Существительное
- <np> ! Proper noun ! Ялгызлык исем ! Имя собственное
- <adj> ! Adjective ! Сыйфат ! Прилагательное
- <num> ! Numeral ! Сан ! Числительное
- <prn> ! Pronoun ! Алмашлык ! Местоимение
- <det> ! Determiner ! Детерминатив ! Детерминатив
- <v> ! Verb ! Фигыль ! Глагол
- <vaux> ! Auxiliary verb ! Ярдәмче фигыль ! Вспомогательный глагол
- <adv> ! Adverb ! Рәвеш ! Наречие
- <post> ! Postposition ! Бәйләк ! Послелог
- <postadv> ! Postadverb ! Пострәвеш ! Посленаречие (Ачыклаган сүзләренәң артыннан килүче рәвешләр (кисәкчәләр) !# гына, ук)
- <cnjcoo> ! Coordinating conjunction ! Тезүче теркәгеч ! Сочинительный союз
- <cnjsub> ! Subordinating conjunction ! Ияртүче теркәгеч ! Подчинительный союз
- <cnjadv> ! Adverbial conjunction ! Рәвеш-теркәгеч ! Наречие-союз
- <ij> ! Interjection ! Ымлык ! Междометие
- <abbr> ! Abbreviation ! Кыскартылма ! Аббревиатура
- <cop> ! Copula ! Копула ! Копула
- <ideo> ! Ideophone ! Аваз ияртеме ! Звукоподражательное слово





Ok **Отмена**



Лаборатория экспериментальной фонетики



В качестве примера рассмотрим поиск словоформы «алма» (существительное). Для этого в поле словоформа вводим слово «алма», а в списке тегов выбираем соответствующее поле.

П И С Ъ М Е Н Н Ы Й К О Р П У С Т А Т А Р С К О Г О Я З Ы К А

TAT RUS ENG



СЛОВО #1   Словоформа: алма  

Лемма:  

Расстояние до следующего слова равно / от 1  до 1 

Грамматические признаки
! Part of speech categories = First-level tags

<input checked="" type="checkbox"/> <n>	! Noun ! Исем ! Существительное
<input type="checkbox"/> <np>	! Proper noun ! Ялгызлык исем ! Имя собственное
<input type="checkbox"/> <adj>	! Adjective ! Сыйфат ! Прилагательное
<input type="checkbox"/> <num>	! Numeral ! Сан ! Числительное
<input type="checkbox"/> <prn>	! Pronoun ! Алмашлык ! Местоимение
<input type="checkbox"/> <det>	! Determiner ! Детерминатив ! Детерминатив
<input type="checkbox"/> <v>	! Verb ! Фигыль ! Глагол
<input type="checkbox"/> <vaux>	! Auxiliary verb ! Ярдәмче фигыль ! Вспомогательный глагол
<input type="checkbox"/> <adv>	! Adverb ! Рәвеш ! Наречие
<input type="checkbox"/> <post>	! Postposition ! Бәйләк ! Послелог
<input type="checkbox"/> <postadv>	! Postadverb ! Пострәвеш ! Посленаречие (Ачыклаган сүзләренең артыннан килүче рәвешләр (кисәкчәләр) !# гына, ук)
<input type="checkbox"/> <cnjcoo>	! Coordinating conjunction ! Тезүче теркәгеч ! Сочинительный союз
<input type="checkbox"/> <cnjsub>	! Subordinating conjunction ! Ияртүче теркәгеч ! Подчинительный союз
<input type="checkbox"/> <cnjadv>	! Adverbial conjunction ! Рәвеш-теркәгеч ! Наречие-союз
<input type="checkbox"/> <ij>	! Interjection ! Ымлык ! Междометие
<input type="checkbox"/> <abbr>	! Abbreviation ! Кыскартылма ! Аббревиатура
<input type="checkbox"/> <cop>	! Copula ! Копула ! Копула
<input type="checkbox"/> <ideo>	! Ideophone ! Аваз ияртеме ! Звукоподражательное слово

Ok  Отмена 

Лаборатория экспериментальной фонетики

Другим вариантом поиска данной словоформы является указание словоформы «алма» и леммы «алма». В этом случае система при поиске исключит глагол «алма», потому что его леммой является «ал».

П И С Ъ М Е Н Н Ы Й К О Р П У С Т А Т А Р С К О Г О Я З Ы К А

TAT RUS ENG

Главная
Поиск в Корпусе
Синтез татарской речи
Инструкция
Статистика
Описание корпуса
Использование корпуса
Публикации
Благодарим...
О проекте

Полезная информация

СЛОВО #1 ↔ ↔

Словоформа: алма

Лемма: алма

Расстояние до следующего слова равно / от до

Грамматические признаки

! Part of speech categories = First-level tags

- <n> ! Noun ! Исем ! Существительное
- <np> ! Proper noun ! Ялгызлык исем ! Имя собственное
- <adj> ! Adjective ! Сыйфат ! Прилагательное
- <num> ! Numeral ! Сан ! Числительное
- <prn> ! Pronoun ! Алмашлык ! Местоимение
- <det> ! Determiner ! Детерминатив ! Детерминатив
- <v> ! Verb ! Фигыль ! Глагол
- <vaux> ! Auxiliary verb ! Ярдәмче фигыль ! Вспомогательный глагол
- <adv> ! Adverb ! Рәвеш ! Наречие
- <post> ! Postposition ! Бәйләк ! Послелог
- <postadv> ! Postadverb ! Пострәвеш ! Посленаречие (Ачыклаган сүзләренәң артыннан килүче рәвешләр (кисәкчәләр) !# гына, ук)
- <cnjcoo> ! Coordinating conjunction ! Тезүче теркәгеч ! Сочинительный союз
- <cnjsub> ! Subordinating conjunction ! Ияргүче теркәгеч ! Подчинительный союз
- <cnjadv> ! Adverbial conjunction ! Рәвеш-теркәгеч ! Наречие-союз
- <ij> ! Interjection ! Ымлык ! Междометие
- <abbr> ! Abbreviation ! Кыскартылма ! Аббревиатура
- <cop> ! Copula ! Копула ! Копула
- <ideo> ! Ideophone ! Аваз ияртеме ! Звукоподражательное слово

Ok
Отмена

Лаборатория экспериментальной фонетики

В качестве следующего примера рассмотрим ситуацию, когда нам необходимо в поиске задать глаголы (<v>), начинающиеся на определенное сочетание букв («ас*»). Данный запрос в графическом режиме оформляется таким образом:

П И С Ъ М Е Н Н Ы Й К О Р П У С Т А Т А Р С К О Г О Я З Ы К А

TAT RUS ENG

Главная
Поиск в Корпусе
Синтез татарской речи
Инструкция
Статистика
Описание корпуса
Использование корпуса
Публикации
Благодарим...
О проекте

Полезная информация

Авторы:
Сайхунов М.Р.
Ибрагимов Т.И.
tatcorpus@gmail.com

СЛОВО #1 ← → **Словоформа:**

Лемма:

Слово начинается на:

Слово заканчивается на:

Расстояние до следующего слова равно / от **до**

Грамматические признаки

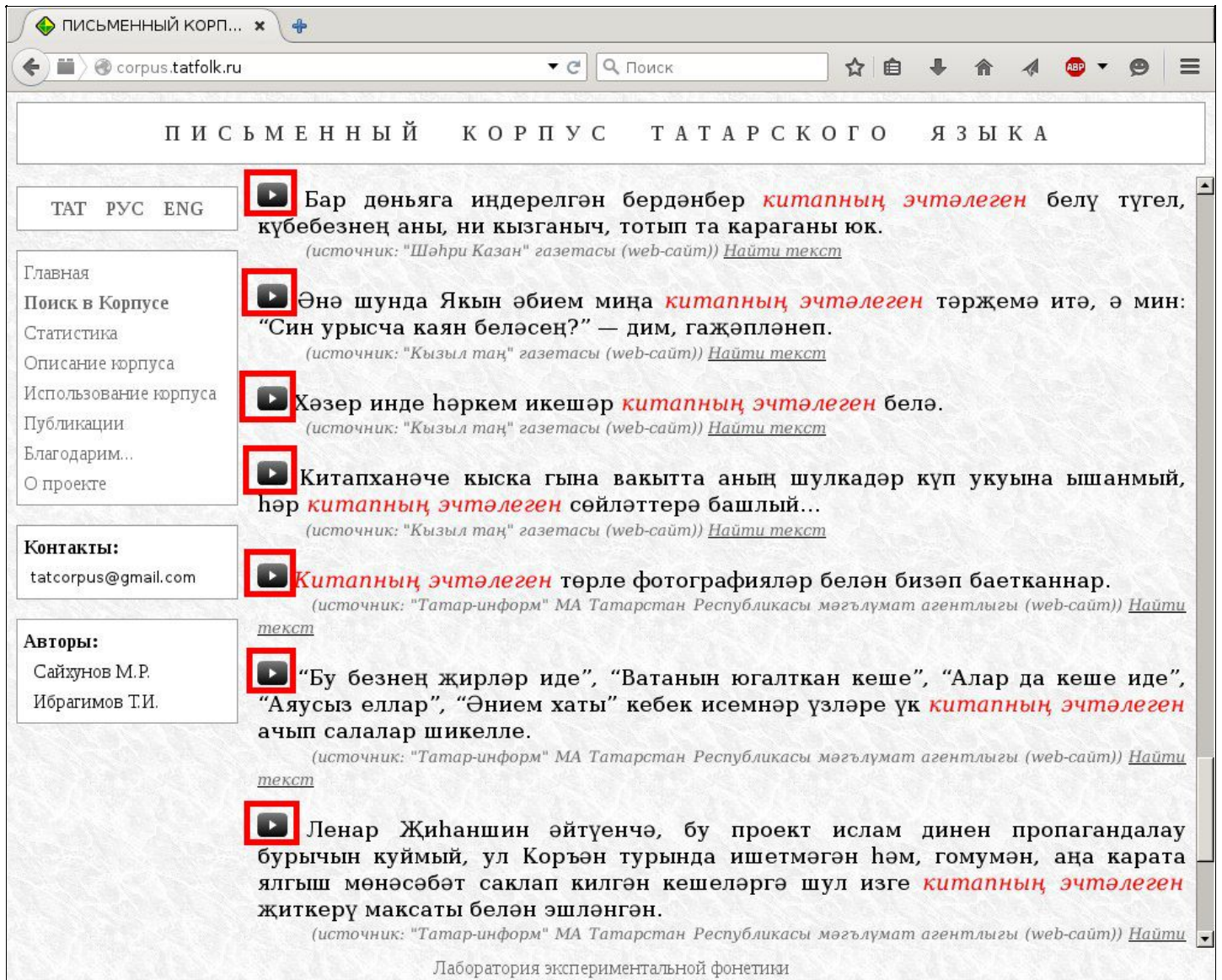
! Part of speech categories = First-level tags

<input type="checkbox"/> <n>	! Noun ! Исем ! Существительное
<input type="checkbox"/> <np>	! Proper noun ! Ялгызлык исем ! Имя собственное
<input type="checkbox"/> <adj>	! Adjective ! Сыйфат ! Прилагательное
<input type="checkbox"/> <num>	! Numeral ! Сан ! Числительное
<input type="checkbox"/> <prn>	! Pronoun ! Алмашлык ! Местоимение
<input type="checkbox"/> <det>	! Determiner ! Детерминатив ! Детерминатив
<input checked="" type="checkbox"/> <v>	! Verb ! Фигыль ! Глагол
<input type="checkbox"/> <vaux>	! Auxiliary verb ! Ярдәмче фигыль ! Вспомогательный глагол
<input type="checkbox"/> <adv>	! Adverb ! Рәвеш ! Наречие
<input type="checkbox"/> <post>	! Postposition ! Бәйләк ! Послелог
<input type="checkbox"/> <postadv>	! Postadverb ! Пострәвеш ! Посленаречие (Ачыклаган сүзләренәң артыннан килүче рәвешләр (кисәкчәләр) !# гына, ук)
<input type="checkbox"/> <cnjcoo>	! Coordinating conjunction ! Тезүче теркәгеч ! Сочинительный союз
<input type="checkbox"/> <cnjsub>	! Subordinating conjunction ! Ияртүче теркәгеч ! Подчинительный союз
<input type="checkbox"/> <cnjadv>	! Adverbial conjunction ! Рәвеш-теркәгеч ! Наречие-союз

Лаборатория экспериментальной фонетики

Как прослушать найденные в Корпусе предложения?

В Письменном корпусе татарского языка имеется уникальная возможность прослушать найденные предложения! Для этого просто надо нажать мышкой на специальной кнопке, расположенной слева от каждого предложения.




П И С Ъ М Е Н Н Ы Й К О Р П У С Т А Т А Р С К О Г О Я З Ы К А


TAT РУС ENG


Главная
Поиск в Корпусе
Статистика
Описание корпуса
Использование корпуса
Публикации
Благодарим...
О проекте


Контакты:
tatcorpus@gmail.com


Авторы:
Сайхунов М.Р.
Ибрагимов Т.И.


 Бар дөнъяга индерелгән бердәнбер **китапның эчтәлеген** белү түгел, күбеезнең аны, ни кызганыч, тотып та караганы юк.
(источник: "Шәһри Казан" газетасы (web-сайт)) [Найти текст](#)


 Әнә шунда Якын әбием миңа **китапның эчтәлеген** тәржемә итә, ә мин: "Син урысча каян беләсең?" — дим, гажәпләнеп.
(источник: "Кызыл таң" газетасы (web-сайт)) [Найти текст](#)

 Хәзер инде һәркем икешәр **китапның эчтәлеген** белә.
(источник: "Кызыл таң" газетасы (web-сайт)) [Найти текст](#)

 Китапханәче кыска гына вакытта аның шулкадәр күп укуына ышанмый, һәр **китапның эчтәлеген** сөйләттерә башлый...
(источник: "Кызыл таң" газетасы (web-сайт)) [Найти текст](#)

 **Китапның эчтәлеген** төрле фотографияләр белән бизәп баетканнар.
(источник: "Татар-информ" МА Татарстан Республикасы мәгълүмат агентлыгы (web-сайт)) [Найти текст](#)

 "Бу безнең жирләр иде", "Ватанын югалткан кеше", "Алар да кеше иде", "Аяусыз еллар", "Әнием хаты" кебек исемнәр үзләре үк **китапның эчтәлеген** ачып салалар шикелле.
(источник: "Татар-информ" МА Татарстан Республикасы мәгълүмат агентлыгы (web-сайт)) [Найти текст](#)

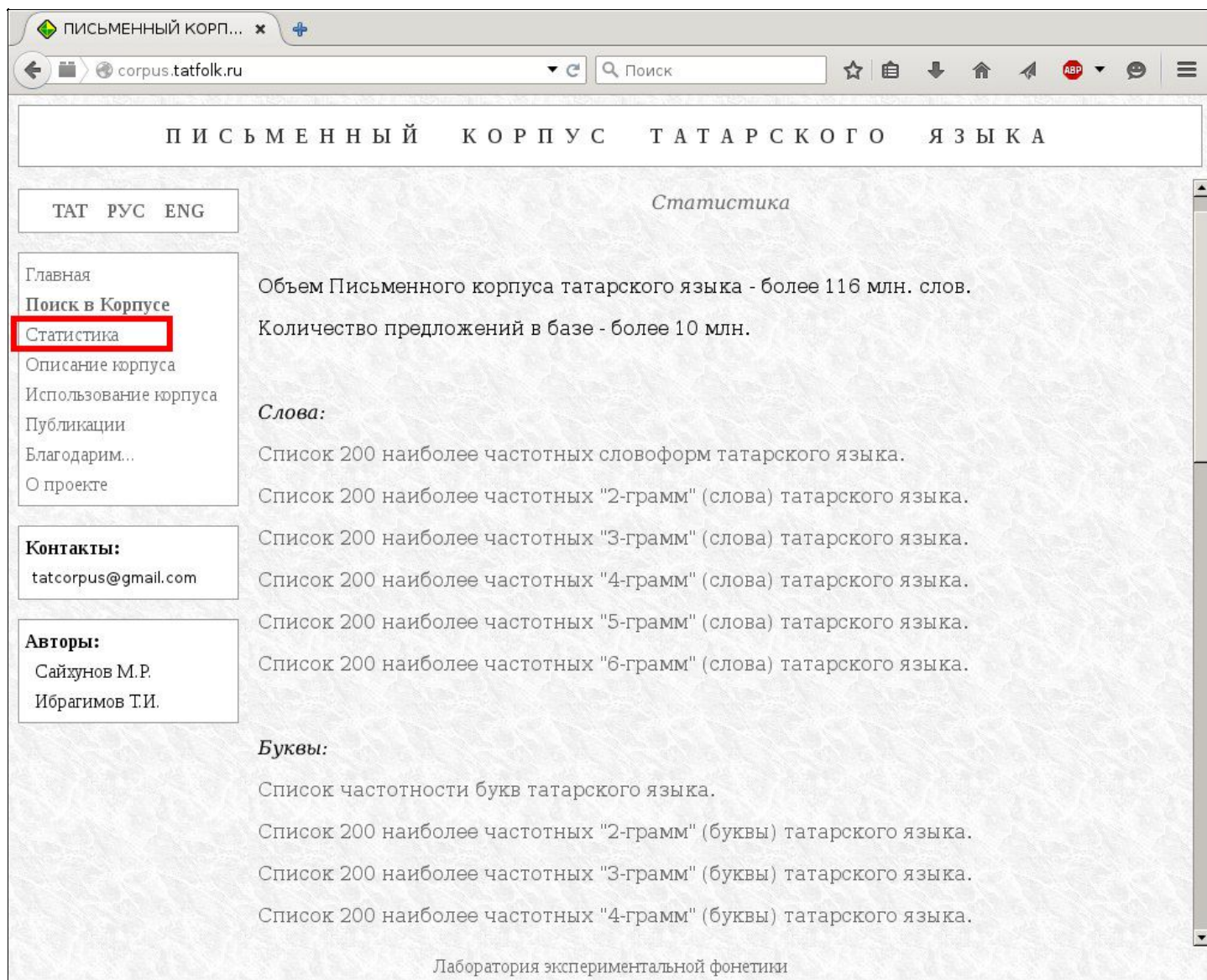
 Ленар Жиһаншин әйтүенчә, бу проект ислам динен пропагандалау бурычын куймый, ул Коръән турында ишетмәгән һәм, гомумән, аңа карата ялгыш мөнәсәбәт саклап килгән кешеләргә шул изге **китапның эчтәлеген** житкерү максаты белән эшләнгән.
(источник: "Татар-информ" МА Татарстан Республикасы мәгълүмат агентлыгы (web-сайт)) [Найти текст](#)

Лаборатория экспериментальной фонетики

Дополнительные статистические материалы!

Авторы корпуса выкладывают различные дополнительные статистические материалы по мере их получения в результате обработки Корпуса. На сегодняшний день к ним относятся:

- Список наиболее частотных словоформ татарского языка.
- Списки наиболее частотных словосочетаний, состоящих из 2, 3, 4, 5 и 6 элементов.
- Списки частотности букв и их сочетаний.



The screenshot shows a web browser window with the URL corpus.tatfolk.ru. The page title is "ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА". There are navigation tabs for "ТАТ", "РУС", and "ENG". The main content area is titled "Статистика" and contains the following information:

- Объем Письменного корпуса татарского языка - более 116 млн. слов.
- Количество предложений в базе - более 10 млн.
- Слова:**
 - Список 200 наиболее частотных словоформ татарского языка.
 - Список 200 наиболее частотных "2-грамм" (слова) татарского языка.
 - Список 200 наиболее частотных "3-грамм" (слова) татарского языка.
 - Список 200 наиболее частотных "4-грамм" (слова) татарского языка.
 - Список 200 наиболее частотных "5-грамм" (слова) татарского языка.
 - Список 200 наиболее частотных "6-грамм" (слова) татарского языка.
- Буквы:**
 - Список частотности букв татарского языка.
 - Список 200 наиболее частотных "2-грамм" (буквы) татарского языка.
 - Список 200 наиболее частотных "3-грамм" (буквы) татарского языка.
 - Список 200 наиболее частотных "4-грамм" (буквы) татарского языка.

At the bottom of the page, it says "Лаборатория экспериментальной фонетики". On the left sidebar, the "Статистика" link is highlighted with a red box.

- Списки частотности букв и их сочетаний в начальной и конечной позиции слов.
- Списки частотности фонем и их сочетаний в пределах слова и ритмической группы.

ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА

TAT RUC ENG

Главная
Поиск в Корпусе
 Статистика
 Описание корпуса
 Использование корпуса
 Публикации
 Благодарим...
 О проекте

Контакты:
 tatcorpus@gmail.com

Авторы:
 Сайхунов М.Р.
 Ибрагимов Т.И.

Буквы (в начале слова):
 Список 200 наиболее частотных "2-грамм" (буквы, начало слова) татарского языка.
 Список 200 наиболее частотных "3-грамм" (буквы, начало слова) татарского языка.
 Список 200 наиболее частотных "4-грамм" (буквы, начало слова) татарского языка.
 Список 200 наиболее частотных "5-грамм" (буквы, начало слова) татарского языка.
 Список 200 наиболее частотных "6-грамм" (буквы, начало слова) татарского языка.

Буквы (в конце слова):
 Список 200 наиболее частотных "2-грамм" (буквы, конец слова) татарского языка.
 Список 200 наиболее частотных "3-грамм" (буквы, конец слова) татарского языка.
 Список 200 наиболее частотных "4-грамм" (буквы, конец слова) татарского языка.
 Список 200 наиболее частотных "5-грамм" (буквы, конец слова) татарского языка.
 Список 200 наиболее частотных "6-грамм" (буквы, конец слова) татарского языка.

Фонемы (в пределах ритмической группы):
 Список частотности фонем татарского языка.
 Список 100 наиболее частотных "2-грамм" (фонемы) татарского языка.
 Список 100 наиболее частотных "3-грамм" (фонемы) татарского языка.

Лаборатория экспериментальной фонетики

Так выглядит список 5000 наиболее частотных словоформ татарского языка:

Rank	Word Form	Frequency
1	һәм	1630726
2	белән	1249114
3	да	1204431
4	дә	910669
5	бу	746627
6	ул	727682
7	дип	638244
8	өчен	603250
9	бер	508703
10	иде	417602
11	аның	353914
12	буенча	341434
13	ә	340479
14	татар	309106
15	генә	289810
16	бик	286219
17	инде	283772
18	шул	269870
19	бар	265782
20	түгел	264347
21	гына	264131
22	шулай	257466
23	Татарстан	255036
24	үз	251835
25	ук	251452
26	турьнда	242908
27	алар	242730
28	булган	242209
29	мин	241080
30	булып	239581
31	ел	232660
32	зур	225474
33	кеше	224115
34	алып	218246
35	иң	217754
36	яңа	213919
37	итеп	207815
38	дәүләт	207369
39	соң	204223
40	кирәк	198707
41	әлеге	198508
42	беренче	196135
43	торган	194072
44	күп	192507
45	то	101750

Лаборатория экспериментальной фонетики

Здесь в первой колонке указан порядковый номер, во второй - сама словоформа, а в третьей - количество их употреблений в Корпусе.

А это список частотности сочетаний из двух букв:

ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА

TAT РУС ENG

Главная
Поиск в Корпусе
Статистика
Описание корпуса
Использование корпуса
Публикации
Благодарим...
О проекте

Контакты:
tatcorpus@gmail.com

Авторы:
Сайхунов М.Р.
Ибрагимов Т.И.

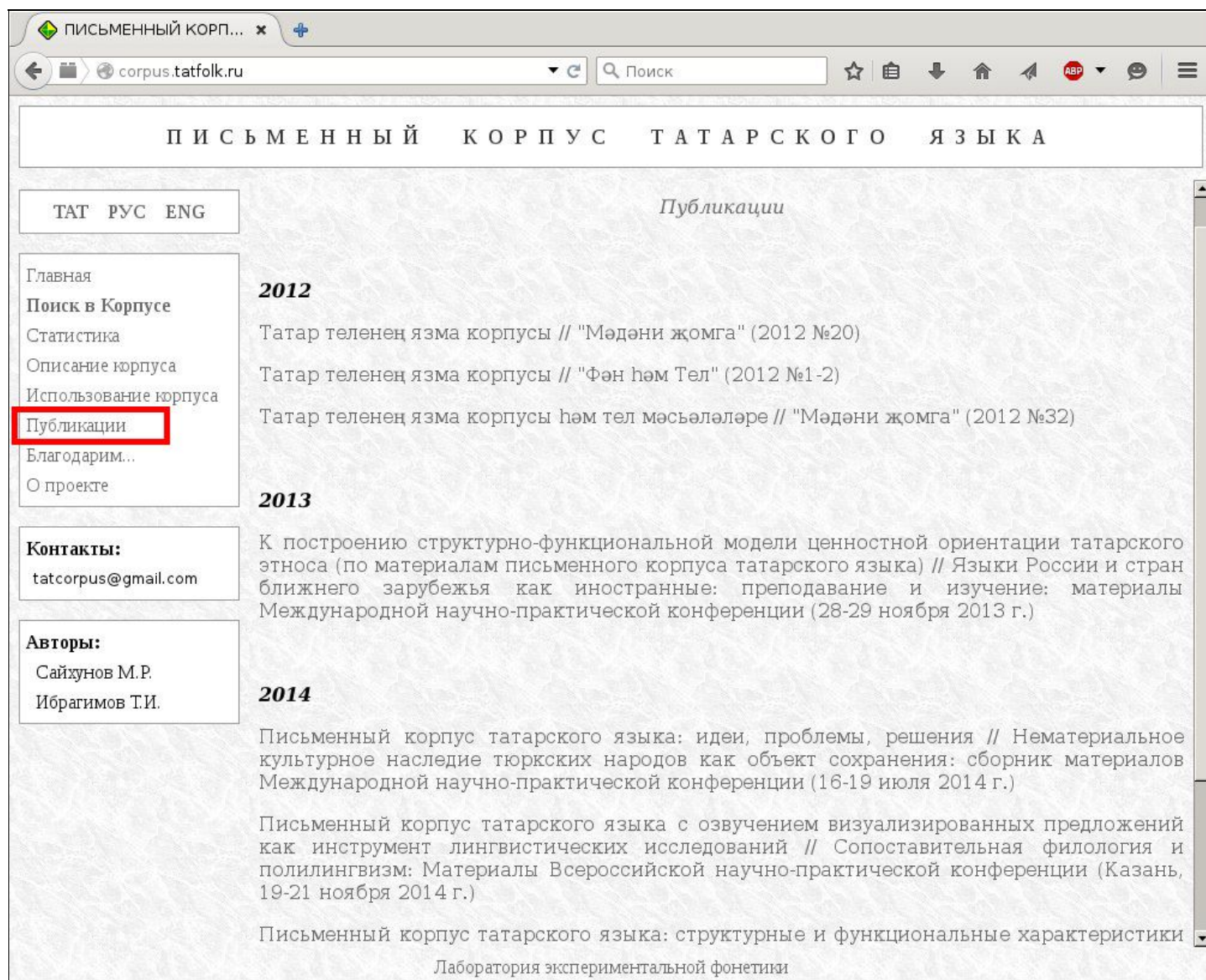
Страница: 1 из 4 Автоматически

1	13487484	ар
2	10927856	ла
3	10537847	ан
4	9389724	лә
5	8426490	ен
6	7742249	әр
7	7508939	ал
8	7476037	га
9	7226503	та
10	6947006	ын
11	6721316	да
12	6716396	ел
13	6538509	ка
14	6512923	ер
15	6420902	ән
16	5772159	ны
17	5497572	ре
18	5470991	нд
19	5292593	ры
20	5203472	дә
21	5122055	на
22	5101267	ле
23	5056469	ра
24	4975734	бе
25	4812588	лы
26	4644417	ул
27	4600942	ат
28	4458442	те
29	4039938	ба
30	3887440	не
31	3817973	ма
32	3731737	ты
33	3708165	ге
34	3700021	гә
35	3649670	бу
36	3507257	гы
37	3492769	тә
38	3456093	ин
39	3448221	ак
40	3393397	нә
41	3337527	мә
42	3275056	ки
43	3209149	ит
44	3175581	ыр
45	3146247	ог

Лаборатория экспериментальной фонетики

Публикации

В разделе «Публикации» сайта помещаются все статьи, опубликованные разработчиками Корпуса:



П И С Ъ М Е Н Н Ы Й К О Р П У С Т А Т А Р С К О Г О Я З Ы К А

TAT RUS ENG

Публикации

Главная
Поиск в Корпусе
Статистика
Описание корпуса
Использование корпуса
Публикации
Благодарим...
О проекте

Контакты:
tatcorpus@gmail.com

Авторы:
Сайхунов М.Р.
Ибрагимов Т.И.

2012

Татар теленең язма корпусы // "Мәдәни жомга" (2012 №20)
Татар теленең язма корпусы // "Фән һәм Тел" (2012 №1-2)
Татар теленең язма корпусы һәм тел мәсьәләләре // "Мәдәни жомга" (2012 №32)

2013

К построению структурно-функциональной модели ценностной ориентации татарского этноса (по материалам письменного корпуса татарского языка) // Языки России и стран ближнего зарубежья как иностранные: преподавание и изучение: материалы Международной научно-практической конференции (28-29 ноября 2013 г.)

2014

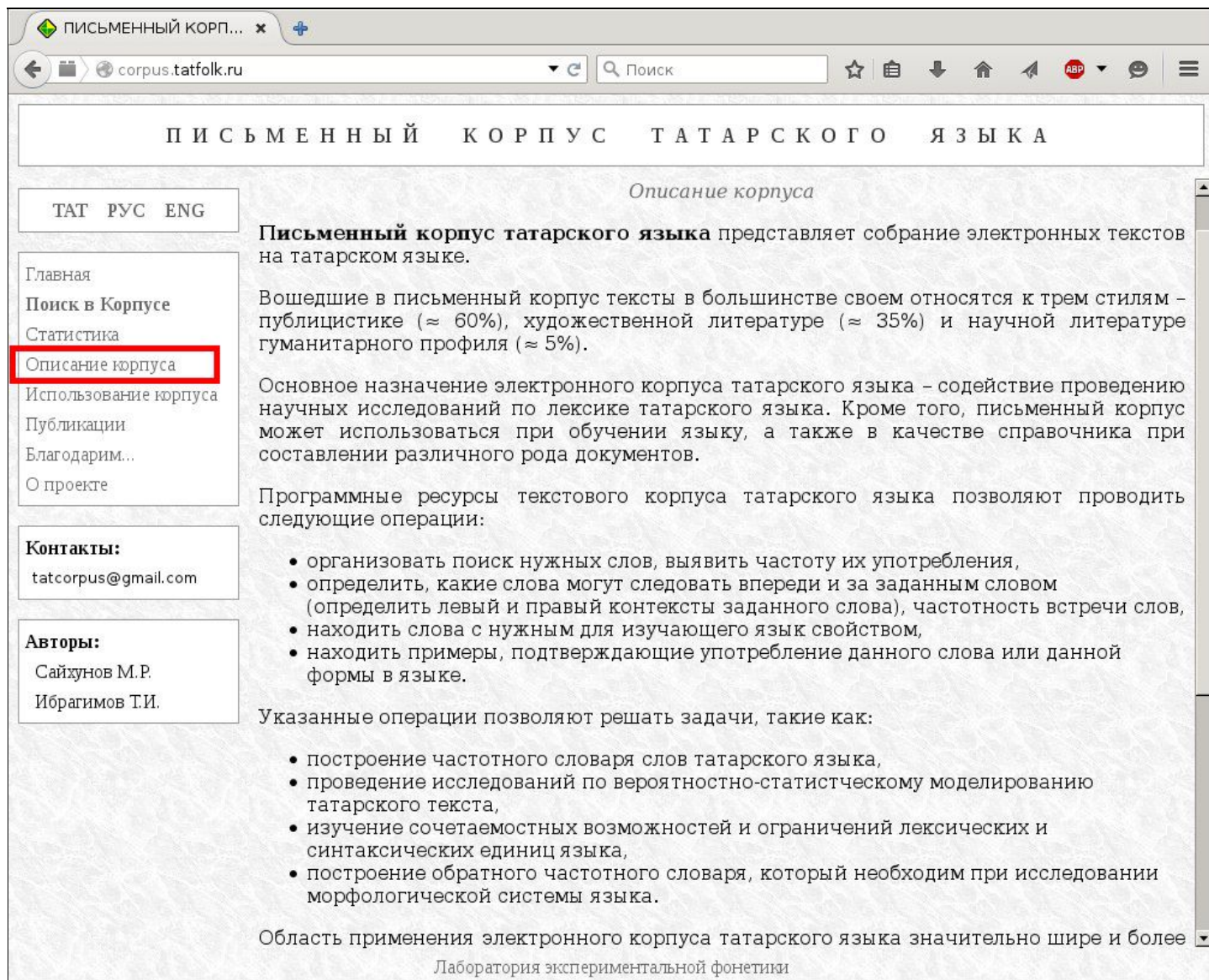
Письменный корпус татарского языка: идеи, проблемы, решения // Нематериальное культурное наследие тюркских народов как объект сохранения: сборник материалов Международной научно-практической конференции (16-19 июля 2014 г.)

Письменный корпус татарского языка с озвучением визуализированных предложений как инструмент лингвистических исследований // Сопоставительная филология и полилингвизм: Материалы Всероссийской научно-практической конференции (Казань, 19-21 ноября 2014 г.)

Письменный корпус татарского языка: структурные и функциональные характеристики
Лаборатория экспериментальной фонетики

Описание

В разделе «**Описание корпуса**» сайта представлена краткая информация по структуре и функциональности Письменного корпуса татарского языка:



The screenshot shows a web browser window with the URL corpus.tatfolk.ru. The page title is "ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА". The main navigation menu includes "TAT", "РУС", and "ENG". A sidebar menu on the left lists: "Главная", "Поиск в Корпусе", "Статистика", "Описание корпуса" (highlighted with a red box), "Использование корпуса", "Публикации", "Благодарим...", and "О проекте". Below the sidebar, there are sections for "Контакты:" (tатcorpus@gmail.com) and "Авторы:" (Сайхунов М.Р., Ибрагимов Т.И.). The main content area is titled "Описание корпуса" and contains the following text:

Письменный корпус татарского языка представляет собрание электронных текстов на татарском языке.

Вошедшие в письменный корпус тексты в большинстве своем относятся к трем стилям – публицистике (≈ 60%), художественной литературе (≈ 35%) и научной литературе гуманитарного профиля (≈ 5%).

Основное назначение электронного корпуса татарского языка – содействие проведению научных исследований по лексике татарского языка. Кроме того, письменный корпус может использоваться при обучении языку, а также в качестве справочника при составлении различного рода документов.

Программные ресурсы текстового корпуса татарского языка позволяют проводить следующие операции:

- организовать поиск нужных слов, выявить частоту их употребления,
- определить, какие слова могут следовать впереди и за заданным словом (определить левый и правый контексты заданного слова), частотность встречи слов,
- находить слова с нужным для изучающего язык свойством,
- находить примеры, подтверждающие употребление данного слова или данной формы в языке.

Указанные операции позволяют решать задачи, такие как:

- построение частотного словаря слов татарского языка,
- проведение исследований по вероятностно-статистическому моделированию татарского текста,
- изучение сочетаемостных возможностей и ограничений лексических и синтаксических единиц языка,
- построение обратного частотного словаря, который необходим при исследовании морфологической системы языка.

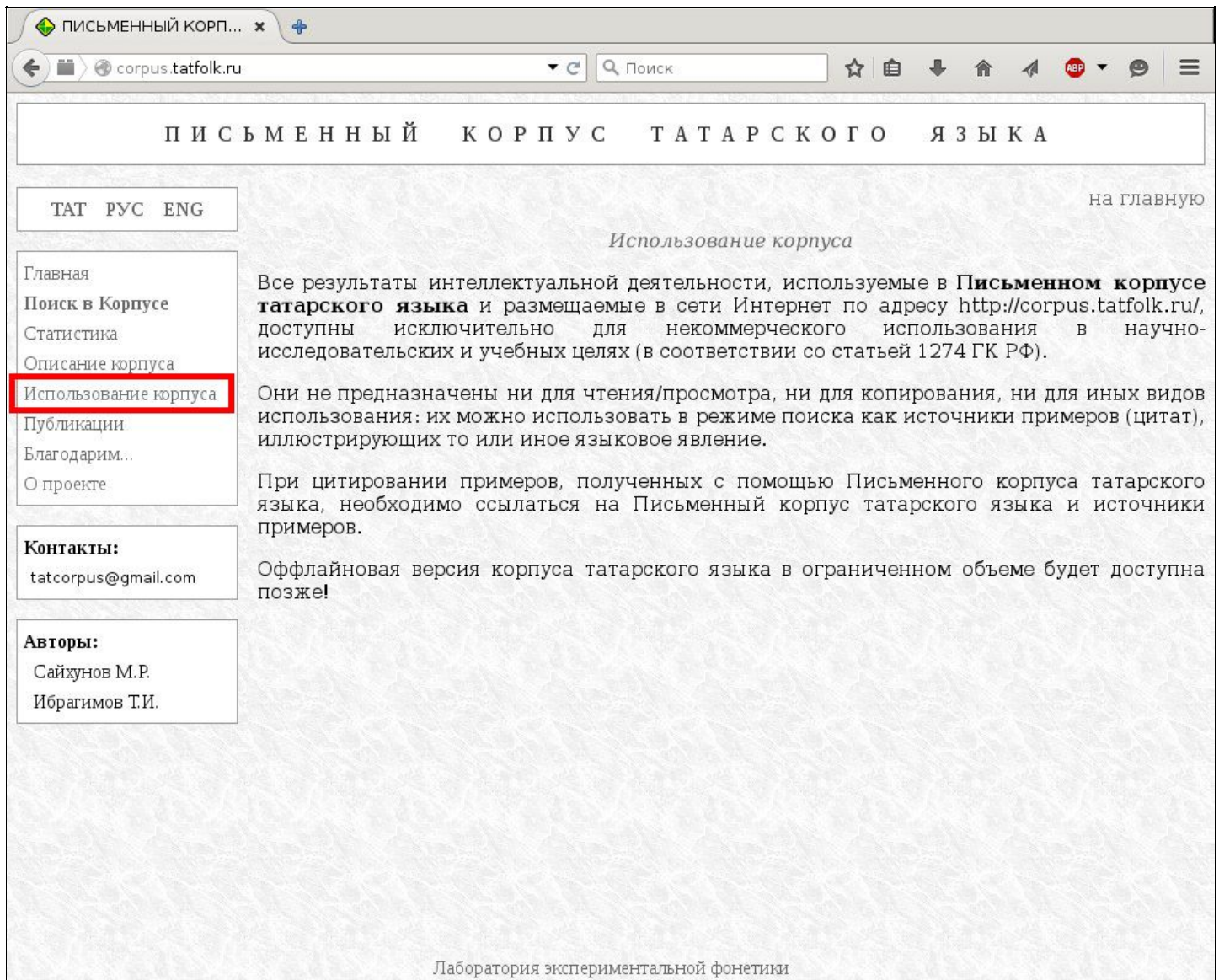
Область применения электронного корпуса татарского языка значительно шире и более

Лаборатория экспериментальной фонетики

Использование и цитирование

В разделе «**Использование корпуса**» сайта пользователь может ознакомиться с правовыми нормами, касающимися использования Корпуса в работе и рекомендациями по оформлению цитат и ссылок на материалы и сам Корпус.

Если Вы пользуетесь Письменным корпусом татарского языка, то просим делать в своих работах ссылки на сам Корпус и на статьи, написанные авторами Корпуса.



The screenshot shows a web browser window with the address bar displaying 'corpus.tatfolk.ru'. The page title is 'ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА'. The main content area is titled 'Использование корпуса' and contains three paragraphs of text. The first paragraph states that results of intellectual activity are available for non-commercial use in scientific and educational purposes. The second paragraph notes that these results are not intended for reading, copying, or other uses, but can be used as search sources for quotes and illustrations. The third paragraph explains that when citing examples from the corpus, users should refer to the corpus and the source of the examples. A footer at the bottom of the page reads 'Лаборатория экспериментальной фонетики'.

ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА

ТАТ РУС ENG на главную

Использование корпуса

Все результаты интеллектуальной деятельности, используемые в **Письменном корпусе татарского языка** и размещаемые в сети Интернет по адресу <http://corpus.tatfolk.ru/>, доступны исключительно для некоммерческого использования в научно-исследовательских и учебных целях (в соответствии со статьей 1274 ГК РФ).

Они не предназначены ни для чтения/просмотра, ни для копирования, ни для иных видов использования: их можно использовать в режиме поиска как источники примеров (цитат), иллюстрирующих то или иное языковое явление.

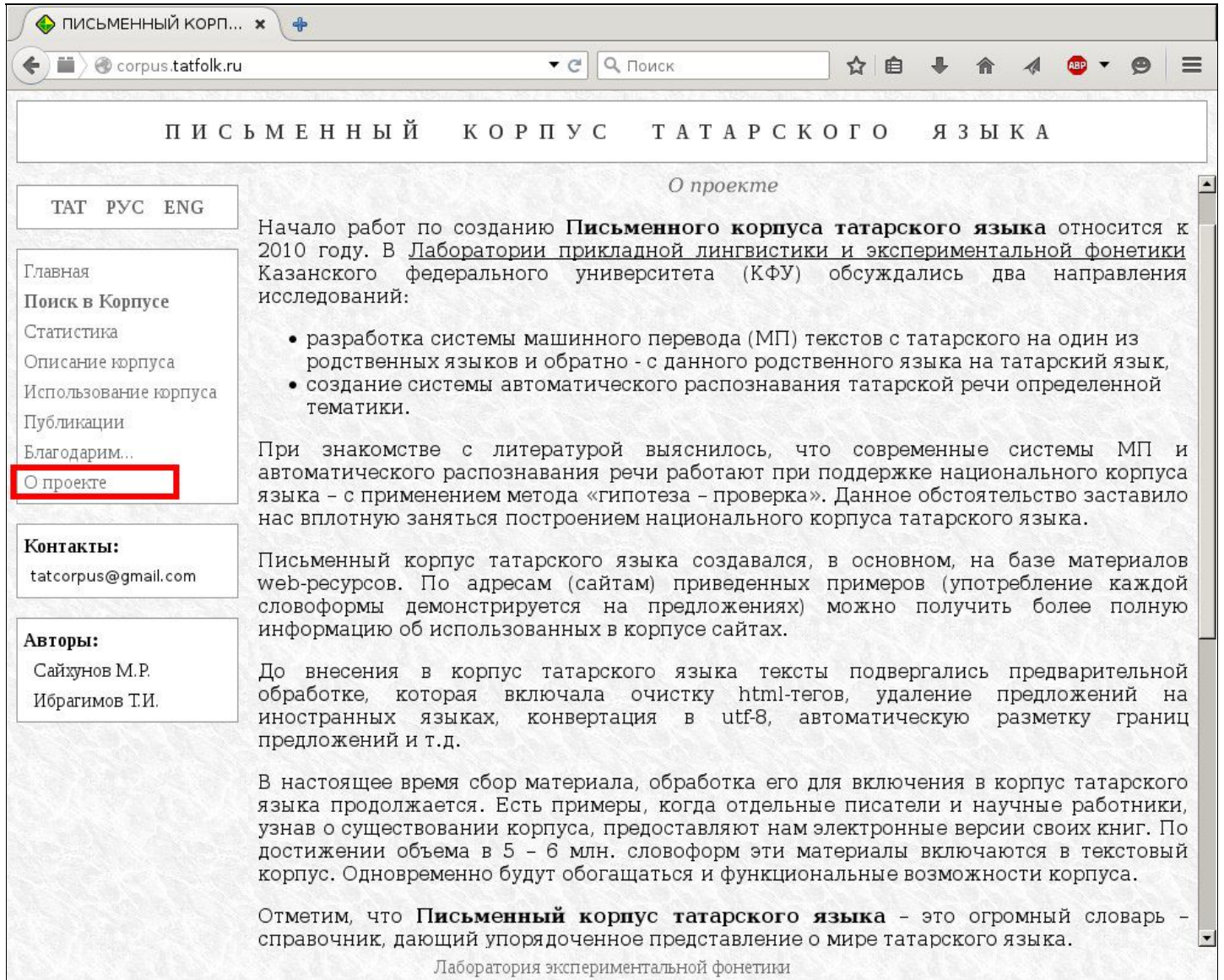
При цитировании примеров, полученных с помощью Письменного корпуса татарского языка, необходимо ссылаться на Письменный корпус татарского языка и источники примеров.

Офлайн-версия корпуса татарского языка в ограниченном объеме будет доступна позже!

Лаборатория экспериментальной фонетики

О проекте

Раздел сайта «**О проекте**» рассказывает об истории создания Корпуса и основных разработчиках:



П И С Ъ М Е Н Н Ы Й К О Р П У С Т А Т А Р С К О Г О Я З Ы К А

О проекте

TAT RUS ENG

Главная
Поиск в Корпусе
Статистика
Описание корпуса
Использование корпуса
Публикации
Благодарим...
О проекте

Контакты:
tatcorpus@gmail.com

Авторы:
Сайхунов М.Р.
Ибрагимов Т.И.

Начало работ по созданию **Письменного корпуса татарского языка** относится к 2010 году. В Лаборатории прикладной лингвистики и экспериментальной фонетики Казанского федерального университета (КФУ) обсуждались два направления исследований:

- разработка системы машинного перевода (МП) текстов с татарского на один из родственных языков и обратно - с данного родственного языка на татарский язык,
- создание системы автоматического распознавания татарской речи определенной тематики.

При знакомстве с литературой выяснилось, что современные системы МП и автоматического распознавания речи работают при поддержке национального корпуса языка - с применением метода «гипотеза - проверка». Данное обстоятельство заставило нас вплотную заняться построением национального корпуса татарского языка.

Письменный корпус татарского языка создавался, в основном, на базе материалов web-ресурсов. По адресам (сайтам) приведенных примеров (употребление каждой словоформы демонстрируется на предложениях) можно получить более полную информацию об использованных в корпусе сайтах.

До внесения в корпус татарского языка тексты подвергались предварительной обработке, которая включала очистку html-тегов, удаление предложений на иностранных языках, конвертация в utf-8, автоматическую разметку границ предложений и т.д.

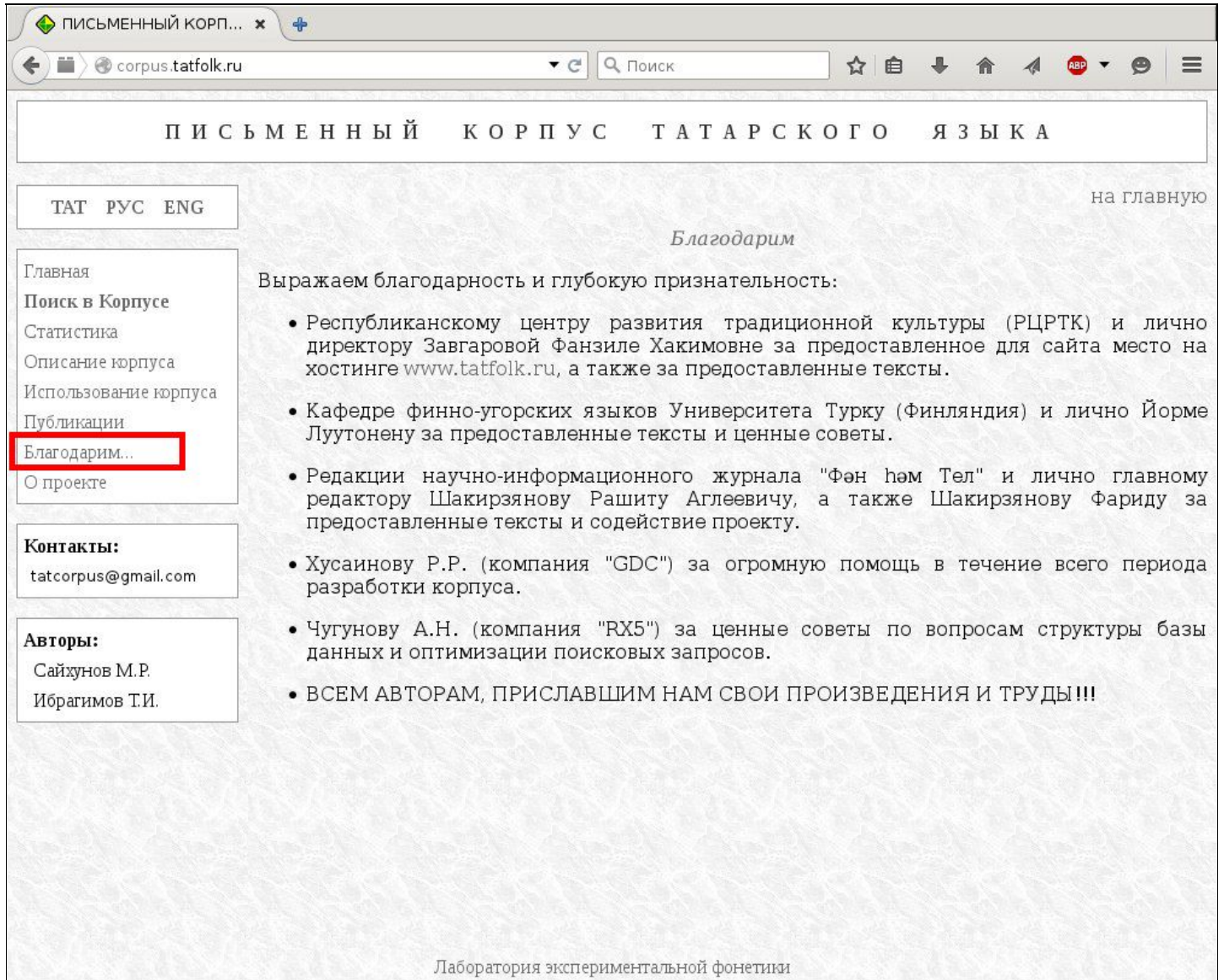
В настоящее время сбор материала, обработка его для включения в корпус татарского языка продолжается. Есть примеры, когда отдельные писатели и научные работники, узнав о существовании корпуса, предоставляют нам электронные версии своих книг. По достижении объема в 5 - 6 млн. словоформ эти материалы включаются в текстовый корпус. Одновременно будут обогащаться и функциональные возможности корпуса.

Отметим, что **Письменный корпус татарского языка** - это огромный словарь - справочник, дающий упорядоченное представление о мире татарского языка.

Лаборатория экспериментальной фонетики

Благодарности

В разделе «**Благодарим...**», как видно из названия, авторы корпуса выражают слова признательности людям и организациям, оказавшим большую помощь в данной работе:



The screenshot shows a web browser window with the address bar displaying 'corpus.tatfolk.ru'. The page title is 'ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА'. The main content area is titled 'Благодарим' and contains a list of acknowledgments. The left sidebar has a menu with 'Благодарим...' highlighted in a red box. The footer of the page reads 'Лаборатория экспериментальной фонетики'.

ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА

TAT РУС ENG на главную

Благодарим

Выражаем благодарность и глубокую признательность:

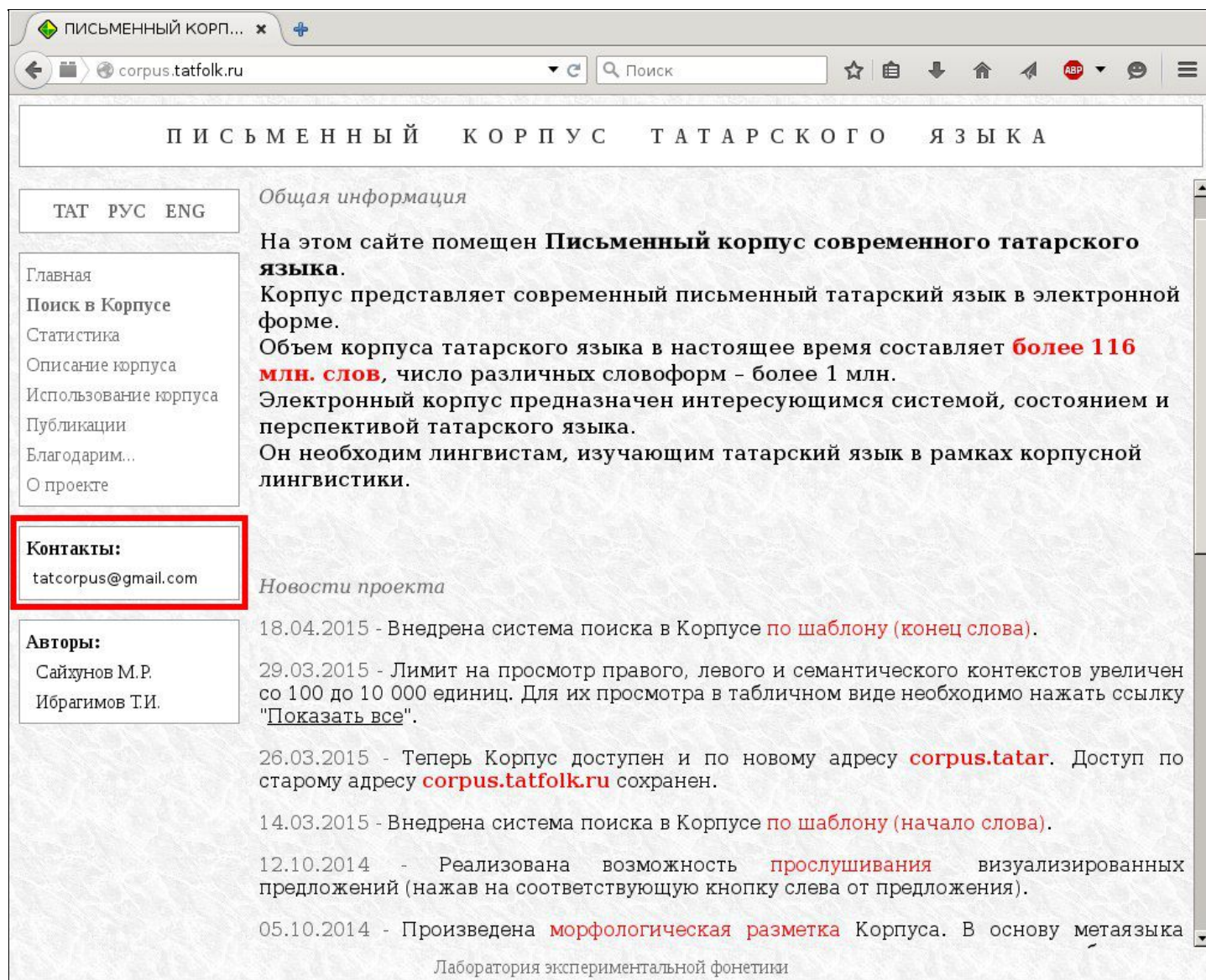
- Республиканскому центру развития традиционной культуры (РЦРТК) и лично директору Завгаровой Фанзиле Хакимовне за предоставленное для сайта место на хостинге www.tatfolk.ru, а также за предоставленные тексты.
- Кафедре финно-угорских языков Университета Турку (Финляндия) и лично Йорме Луутонену за предоставленные тексты и ценные советы.
- Редакции научно-информационного журнала "Фән һәм Тел" и лично главному редактору Шакирзянову Рашиту Аглеевичу, а также Шакирзянову Фариду за предоставленные тексты и содействие проекту.
- Хусаинову Р.Р. (компания "GDC") за огромную помощь в течение всего периода разработки корпуса.
- Чугунову А.Н. (компания "RX5") за ценные советы по вопросам структуры базы данных и оптимизации поисковых запросов.
- ВСЕМ АВТОРАМ, ПРИСЛАВШИМ НАМ СВОИ ПРОИЗВЕДЕНИЯ И ТРУДЫ!!!

Лаборатория экспериментальной фонетики

Контакты для Ваших предложений

E-mail авторов Корпуса указан слева под основным меню: tatcorpus@gmail.com

Просим Вас обращаться к нам с любыми вопросами и предложениями по использованию Письменного корпуса татарского языка!



The screenshot shows a web browser window with the URL corpus.tatfolk.ru. The page title is "ПИСЬМЕННЫЙ КОРПУС ТАТАРСКОГО ЯЗЫКА". The main content area is titled "Общая информация" and contains the following text:

На этом сайте помещен **Письменный корпус современного татарского языка**.
Корпус представляет современный письменный татарский язык в электронной форме.
Объем корпуса татарского языка в настоящее время составляет **более 116 млн. слов**, число различных словоформ - более 1 млн.
Электронный корпус предназначен интересующимся системой, состоянием и перспективой татарского языка.
Он необходим лингвистам, изучающим татарский язык в рамках корпусной лингвистики.

Below this text is a section titled "Новости проекта" with several entries:

- 18.04.2015 - Внедрена система поиска в Корпусе **по шаблону (конец слова)**.
- 29.03.2015 - Лимит на просмотр правого, левого и семантического контекстов увеличен со 100 до 10 000 единиц. Для их просмотра в табличном виде необходимо нажать ссылку "[Показать все](#)".
- 26.03.2015 - Теперь Корпус доступен и по новому адресу **corpus.tatar**. Доступ по старому адресу **corpus.tatfolk.ru** сохранен.
- 14.03.2015 - Внедрена система поиска в Корпусе **по шаблону (начало слова)**.
- 12.10.2014 - Реализована возможность **прослушивания** визуализированных предложений (нажав на соответствующую кнопку слева от предложения).
- 05.10.2014 - Произведена **морфологическая разметка** Корпуса. В основу метаязыка

At the bottom of the page, it says "Лаборатория экспериментальной фонетики".

On the left side of the page, there is a navigation menu with the following items:

- Главная
- Поиск в Корпусе
- Статистика
- Описание корпуса
- Использование корпуса
- Публикации
- Благодарим...
- О проекте

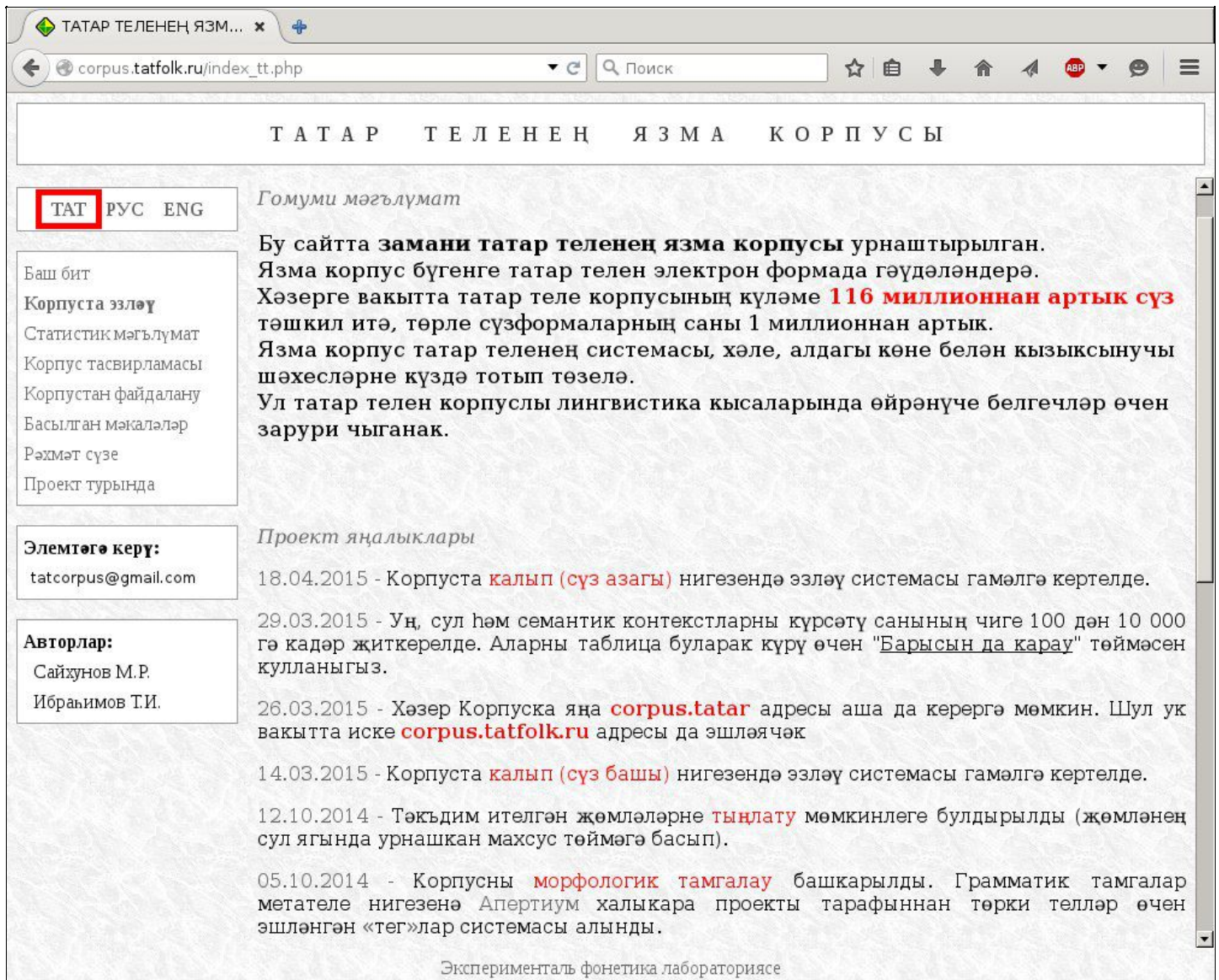
Below the navigation menu, there are two boxes:

- Контакты:**
tatcorpus@gmail.com
- Авторы:**
Сайхунов М.Р.
Ибрагимов Т.И.

На других языках

Для удобства наших пользователей все разделы сайта Письменного корпуса татарского языка на данный момент представлены на трех языках: русский, татарский и английский. Ведутся работы по переводу и на другие языки!

Выбрать язык интерфейса можно, кликнув на соответствующей ссылке над основным меню слева экрана (в данном случае на татарском языке):



The screenshot shows a web browser window with the URL `corpus.tatfolk.ru/index_tt.php`. The page title is "ТАТАР ТЕЛЕНЕЦ ЯЗМА КОРПУСЫ". At the top left, there is a language selection menu with three buttons: "TAT", "РУС", and "ENG". The "TAT" button is highlighted with a red box. Below the menu, there are several sections: "Баш бит", "Корпуста эзләү", "Статистик мәгълүмат", "Корпус тасвирламасы", "Корпустан файдалану", "Басылган мәкаләләр", "Рәхмәт сүзе", and "Проект турында". There is also a contact section "Элемтәгә керү:" with the email `tatcorpus@gmail.com` and an "Авторлар:" section listing Сайхунов М.Р. and Ибраһимов Т.И. The main content area is titled "Гомуми мәгълүмат" and contains a welcome message in Tatar, mentioning that the corpus has 116 million words. Below this is a "Проект яңалыклары" section with a list of updates from 2014 to 2015, including the development of a search system and the addition of new words.

ТАТАР ТЕЛЕНЕЦ ЯЗМА КОРПУСЫ

TAT РУС ENG

Баш бит
Корпуста эзләү
Статистик мәгълүмат
Корпус тасвирламасы
Корпустан файдалану
Басылган мәкаләләр
Рәхмәт сүзе
Проект турында

Элемтәгә керү:
tatcorpus@gmail.com

Авторлар:
Сайхунов М.Р.
Ибраһимов Т.И.

Гомуми мәгълүмат

Бу сайтта замани татар теленең язма корпусы урнаштырылган. Язма корпус бүгенге татар телен электрон формада гәүдәләндерә. Хәзерге вакытта татар теле корпусының күләме **116 миллионнан артык сүз** тәшкил итә, төрле сүзформаларның саны 1 миллионнан артык. Язма корпус татар теленең системасы, хәле, алдагы көне белән кызыксынучы шөхесләргә күздә тотып төзелә. Ул татар телен корпуслы лингвистика кысаларында өйрәнүче белгечләр өчен зарури чыганаҡ.

Проект яңалыклары

18.04.2015 - Корпуста **калып (сүз азагы)** нигезендә эзләү системасы гамәлгә кертелде.

29.03.2015 - Уң, сул һәм семантик контекстларны күрсәтү санының чиге 100 дән 10 000 гә кадәр житкерелде. Аларны таблица буларак күрү өчен "Барысын да карау" төймәсен кулланыгыз.

26.03.2015 - Хәзер Корпуска яңа **corpus.tatar** адресы аша да керергә мөмкин. Шул ук вакытта иске **corpus.tatfolk.ru** адресы да эшләчәк

14.03.2015 - Корпуста **калып (сүз башы)** нигезендә эзләү системасы гамәлгә кертелде.

12.10.2014 - Тәкъдим ителгән **жөмлөләрне тыңлату** мөмкинлегә булдырылды (жөмлөнән сул ягында урнашкан махсус төймәгә басып).

05.10.2014 - Корпусны **морфологик тамгалау** башкарылды. Грамматик тамгалар метателе нигезенә Апертиум халыкара проекты тарафыннан төрки телләр өчен эшләнгән «тег»лар системасы алынды.

Эксперименталь фонетика лабораториясе

А так выглядит английский интерфейс сайта Корпуса:

The screenshot shows a web browser window with the address bar containing 'corpus.tatfolk.ru/index_en.php'. The page title is 'CORPUS OF WRITTEN TATAR'. On the left side, there is a navigation menu with 'ENG' highlighted in a red box. Below the menu, there are sections for 'Contact address:' (tatcorpus@gmail.com) and 'Developers:' (Saykhunov M.R., Ibragimov T.I.). The main content area is titled 'General Information' and contains the following text:

This website contains a **Text Corpus of the modern Tatar language** consisting of **over 116 million word occurrences**. The corpus represents modern written Tatar language in electronic form. The total count of different word forms in the Tatar corpus is more than 1 mln. This collection of Tatar texts in electronic form is intended for the use of those interested in the structure, present condition and prospects of the Tatar language. The Corpus of Written Tatar language is indispensable for everyone who wants to study Tatar by the methods of corpus linguistics.

Project news

18.04.2015 - **Template (the end of the word)** based search system in the Corpus is implemented.

29.03.2015 - Limit for viewing right, left and semantic contexts is increased from 100 to 10 000 units. To view them in a table format you should click on the link "[Show all](#)".

26.03.2015 - Now The Corpus is also available at the new address **corpus.tatar**. At the same time you can use the old address **corpus.tatfolk.ru**.

14.03.2015 - **Template (the beginning of the word)** based search system in the Corpus is implemented.

12.10.2014 - Implementation of a system for **listening** of visualized sentences (by clicking on the appropriate button on the left of sentences)

05.10.2014 - The **morphological marking** of the Corpus is made. The meta-language of

Laboratory of Experimental Phonetics