

Несмотря на процесс дерусификации в Таджикистане, русский язык продолжает играть ключевую роль в информационном пространстве страны. Согласно исследованию Навруза Каримова, подавляющее большинство поисковых запросов к Google в Таджикистане делается на русском языке. Эти данные говорят не только об отсутствии полезного контента на таджикском языке, но и о плохих перспективах таджикского языка.



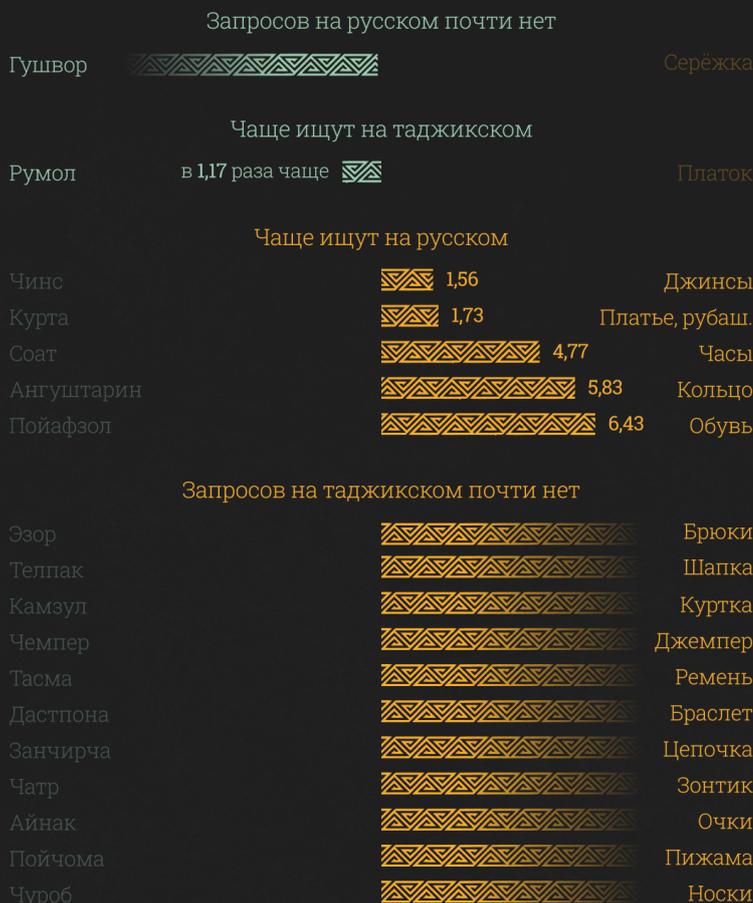
Для исследования мы отобрали 260 слов и словосочетаний в 13 разных сферах жизни. За последний год эти слова на русском языке в среднем искали в 4,13 раза чаще, чем на таджикском — не считая те из них, которые на таджикском не ищут вообще. К сожалению, есть и такие — например, в категории «одежда и аксессуары» 11 слов из 20 искали исключительно на русском языке.

# На каком языке чаще ищут информацию об одежде и аксессуарах в Google?

Среднее соотношение запросов из Таджикистана на русском и таджикском языках за последние 12 месяцев\*

\* Дата извлечения данных: 22.03.2023

Источник: Google Trends



Такая же картина наблюдается при поиске информации об объектах городской инфраструктуры. Ровно 50% терминов из этой категории почти никогда не «гуглят» на таджикском.

# На каком языке чаще делают запросы о городских объектах в Google?

Среднее соотношение запросов из Таджикистана на русском и таджикском языках за последние 12 месяцев\*

\* Дата извлечения данных: 22.03.2023

Источник: Google Trends



По мнению руководителя Общественного фонда «Гражданская инициатива политики Интернет» **Мухаммади Ибодуллоева**, низкое количество поисковых запросов на таджикском языке объясняется комплексом причин:

«Во-первых, в интернете мало контента на таджикском языке. Информационные агентства, поставляющие большую часть актуальных материалов, до сих пор чаще пишут на русском. Получается, искать что-то на таджикском — это терять доступ к части поисковой информации».

Мухаммади Ибодуллоев также обращает внимание на то, что порой пользователи бессознательно используют русский язык в интернете из-за того, что пытаются следовать правилам интерфейса:

«Пользователи обычно не идут сложными путями. Если сайт по умолчанию открылся

на русском, мало кто будет искать переключатель языка. То же и с операционными системами — несмотря на то, что в Windows уже давно есть таджикский язык, многие все равно используют русский.

Браузер и разные сайты, включая Google, считывают язык системы, переводят интерфейс и чаще предлагают контент на русском языке. В результате и пользователи чаще используют русский, так как вся коммуникация с сайтом происходит на русском. Получается замкнутый круг».

Ещё несколько лет назад по запросу «обу хаво» (погода) можно было получить устаревшие данные о погоде. Сейчас эта проблема уже в прошлом, и Google в ответ на запрос подгружает свою интерактивную панель. Однако пользователи все равно чаще ищут погоду на русском — отчасти, по привычке, отчасти, потому что из-за системных настроек информация о прогнозе погоды всё равно подгружается на русском (или на английском, если язык системы английский).

# На каком языке чаще делают запросы о природе в Google?

Среднее соотношение запросов из Таджикистана на русском и таджикском языках за последние 12 месяцев\*

\* Дата извлечения данных: 22.03.2023

Источник: Google Trends



## Особенности алгоритмов

Другие продукты компании Google также плохо поддерживают таджикский язык. Например, Google Translate «проседает» в переводе многих тем. Причина — слишком малое количество оцифрованных текстов, на которых мог бы «обучиться» алгоритм перевода. Однако эксперты отмечают, что IT-термины на таджикском языке переводчик должен обрабатывать лучше прочих:

«Из Худжанда к нам как-то обратился специалист, который занимался вопросом развития Google Translate», — вспоминает Мухаммади Ибодуллоев. — «Мы предоставили ему серию учебных текстов в сфере информационных технологий. Когда-то эти тексты специально переводили с английского языка на таджикский, чтобы помочь развитию отрасли в стране. А теперь они помогают улучшить качество перевода у Google».

По словам Ибодуллоева, проблемы Google с таджикским языком не связаны с его сложностью. «Таджикский язык более «математичен», алгоритмам наоборот легче обрабатывать данные на таджикском, чем на русском или английском языках», — говорит Ибодуллоев. — «Несмотря на это, эффективность работы алгоритмов поисковых систем также в значительной степени зависит от того, насколько популярен язык, от качества данных, доступных для обучения этих алгоритмов. На таджикском языке данных мало, и это в итоге затрудняет работу алгоритмов».

## На каком языке чаще делают запросы о еде и напитках в Google?

Среднее соотношение запросов из Таджикистана на русском и таджикском языках за последние 12 месяцев\*

\* Дата извлечения данных: 22.03.2023

Источник: Google Trends

### Чаще ищут на таджикском

Нон	в 2,85 раза чаще		Лепешка
Шурбо	2,6		Шурпа
Об	2,46		Вода
Пудина	1,33		Мята
Самбуса	1,33		Самса
Сабзи	1,30		Морковь
Ош	1,14		Плов

### Чаще ищут на русском

Шароб	 1,27	Алкоголь
Бодиринг	 1,33	Огурец
Сабзавот	 1,53	Овощи
Субхона	 1,75	Завтрак
Гушт	 1,85	Мясо
Пиёз	 2,41	Лук
Меваги	 6,8	Фрукты
Чой	 6,82	Чай

### Запросов на таджикском почти нет

Хуроки нисфирузи		Обед
Хуроки шом		Ужин
Кабуди		Зелень
Каламфур		Перец
Кахва		Кофе

Анализ запросов в Google показывает, что термины, которые относятся к традиционной культуре (например, национальные блюда) чаще ищут именно на таджикском языке. А широко распространенные слова, относящиеся к еде и напиткам, например, «чай», «мясо», «фрукты» — на русском.

### Перспективы таджикского языка

Все эксперты, с которыми мы говорили во время работы над этим материалом, считают, что отсутствие качественного контента на таджикском языке в важных областях науки может затруднить его дальнейшее развитие. И такая проблема действительно существует: только два слова в категории «медицина и болезни» из нашей выборки чаще искали на таджикском языке — «туберкулёз» и «психолог».

# На каком языке чаще ищут информацию о медицине и болезнях в Google?

Среднее соотношение запросов из Таджикистана на русском и таджикском языках за последние 12 месяцев\*

\* Дата извлечения данных: 22.03.2023

Источник: Google Trends

## Чаще ищут на таджикском

Бемории сил  
Равоншинос

в 1,25 раза чаще  
1,18

Туберкулёз  
Психолог

## Чаще ищут на русском

Эуком  
Духтур  
Дорухона  
Дорувори  
Беморхона  
Табобат  
Саратон

1,9  
2,3  
4,4  
4,4  
9,3  
15,5  
22,5

Простуда  
Врач  
Аптека  
Лекарства  
Больница  
Лечение  
Рак

## Запросов на таджикском почти нет

Фарбехи  
Бехоби  
Пахнкафии пой  
Дарунрави  
Сактаи майна  
Ёрии таъчили  
Дарди сар  
Дамиши рудахо

Ожирение  
Бессонница  
Плоскостопие  
Диарея  
Инсульт  
Скор. пом.  
Голов. боль  
Вздут. живота

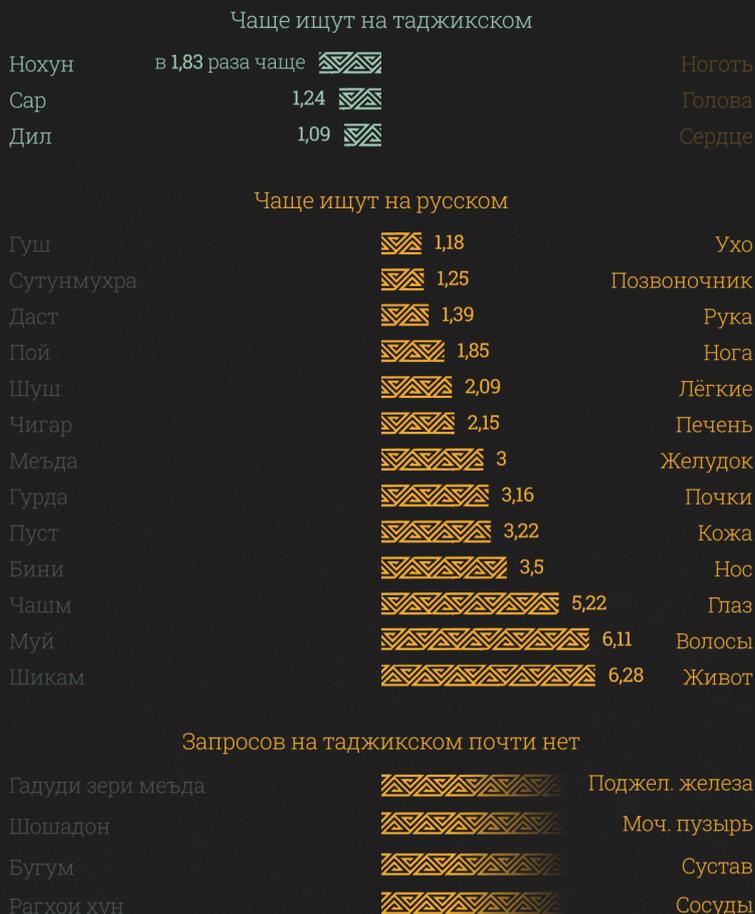
В сопутствующей теме, анатомии человека, тоже проблемы. Информацию об органах и частях тела ищут в основном на русском языке. На таджикском в незначительной степени чаще искали всего три слова.

# На каком языке чаще делают запросы об органах и частях тела в Google?

Среднее соотношение запросов из Таджикистана на русском и таджикском языках за последние 12 месяцев\*

\* Дата извлечения данных: 22.03.2023

Источник: Google Trends



Эксперт по языку **Умед Джайхони** выразил пессимистический взгляд на будущее таджикского языка, утверждая, что он застрял в развитии. В Таджикистане действуют Комитет по языку и терминологии и Институт таджикского языка и литературы, которые регулируют появление новых слов в языке, но «в том виде, в котором язык существует сейчас, он не имеет будущего» — декларирует Джайхони.

«Я разработал и предложил оригинальную таджикскую систему воинских званий для вооружённых сил Таджикистана, но ее до сих пор не утвердили и пользуются советско-российскими званиями, потому что генералитет у нас всё ещё мыслит категориями советского интернационализма».

Джайхони активно создает контент на таджикском языке, но он признается, что информацию на нём почти не ищет: «я просто знаю, что в интернете на таджикском

нет того, что мне нужно. Книги на родном языке у нас есть, но они не оцифрованы. Вот и приходится использовать русскоязычные источники».

Интересно, что, несмотря на общую тенденцию, запросы на тему религии часто задают именно на таджикском языке. Например, информацию о паломничестве и исламские атрибуты, такие как «молитвенный коврик» и «тюбетейка», на русском языке искать даже не пытаются. Это может говорить о том, что таджикоязычное население особенно интересуется исламская тематика, и контент на эту тему делают сразу на родном языке.



Еще одна особенность таджикоязычного контента — аполитичность. Даже информацию об экономике и политике, за редкими исключениями, ищут на русском. Только самые острые для Таджикистана темы — миграцию, налоги и торговлю, ищут на

родном языке. На равных в поиске идут слова «сохибкор» и «предприниматель», однако это не обязательно говорит о развитии соответствующего контента в регионе. Скорее всего, «очков» таджикоязычной версии добавляют узбекский футбольный клуб «Сохибкор» и улицы с одноименным названием.



### «Неродные» развлечения

Качественного развлекательного контента на таджикском языке тоже немного. Удивительно, что рассказать об этом могут даже данные о том, как в Таджикистане ищут животных.

# На каком языке чаще ищут информацию о ЖИВОТНЫХ в Google?

Среднее соотношение запросов из Таджикистана на русском и таджикском языках за последние 12 месяцев\*

\* Дата извлечения данных: 22.03.2023

Источник: Google Trends

## Чаще ищут на таджикском

Мохи	в 3 раза чаще	Рыба
Фил	1,15	Слон
Тути	1,1	Попугай
Маймун	1,09	Обезьяна
Тимсох	1,05	Крокодил

## Чаще ищут на русском

Асп	1,21	Лошадь
Укоб	1,22	Орёл
Гов	1,57	Корова
Харгуш	1,72	Кролик
Санчоб	2,28	Белка
Гусфанд	2,42	Баран
Мург	2,44	Курица
Паранда	2,7	Птица
Саг	3,23	Собака
Гург	3,44	Волк
Хирс	18	Медведь
Гурба	26,5	Кот

## Запросов на таджикском почти нет

Мургоби		Утка
Хуг		Свинья
Зог		Ворон

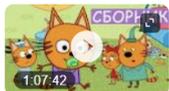
Котов в 18 раз, а медведей — в 26 раз чаще ищут именно на русском языке. Это объясняется тем, что при вводе слова «кот» в Google среди первых результатов поиска появляется мультфильм «Три кота». В случае запроса «медведь» — «Маша и Медведь». Вот так российские мультфильмы смогли потеснить из поиска национальные названия.

## Языковые предпочтения в Таджикистане: что говорят данные поисковых запросов?

Результатов: примерно 21 400 000 (0,30 сек.)

www.youtube.com › watch

### Три Кота | Сборник Лучших Серий - YouTube



С конструкторами ТРИ КОТА играть очень интересно, а с друзьями ещё и весело! Конструкторы ТРИ КОТА от 1TOY покупайте онлайн: ...

YouTube · Три Кота · 13 апр. 2021 г.



15 ключевых моментов в этом видео

www.youtube.com › watch

### Какой кот самый популярный в мире? - YouTube



... попал в книгу рекордов Гиннеса. Видео с ним в Интернете посмотрели более 330 миллионов раз. В чем секрет популярности кота и сколько ...

YouTube · Сейчас · 28 мар. 2017 г.

www.youtube.com › watch

### Три Кота | Сборник невероятных серий - YouTube



В этом сборнике вас ждут нешуточные приключения Коржика, Карамельки и Компота: от соревнований в меткости и настоящего...

YouTube · Три Кота · 2 мар. 2021 г.

Топ поисковой выдачи по запросу «кот» на русском языке.

Результатов: примерно 48 800 (0,26 сек.)

www.youtube.com › watch

### Суоли-26 Гурба оё начас аст? - YouTube



Суоли-26 Гурба оё начас аст? 9.5K views 3 years ago. Абу Мухаммад Мадани [ابو محمد المدني](#). Абу Мухаммад Мадани [ابو محمد المدني](#).

YouTube · [ابو محمد المدني](#) · Абу Мухаммад Мадани · 18 июн. 2019 г.

www.ozodi.org › ... · [Перевести эту страницу](#)

### гурба - Радиои Озодӣ



гурба. gurbai da fikr furu rafta Гузошта шуд аз ҷониби илья. Идома. Ҳамаи хабарҳои имрӯро дар инҷо бихонед. январ 11, 2022 ...

Радиои Озодӣ · 13 янв. 2021 г.

www.youtube.com › watch

### Гурба тифли навзодро наҷот дод - YouTube



Гурба тифли навзодро наҷот дод. 182 views 1 year ago. Садои БОРОН. Садои БОРОН. 4.24K subscribers. [Subscribe](#). 0. [I like this](#). [I dislike this](#).

YouTube · Садои БОРОН · 30 окт. 2021 г.

Топ поисковой выдачи по запросу «кот» на таджикском языке.

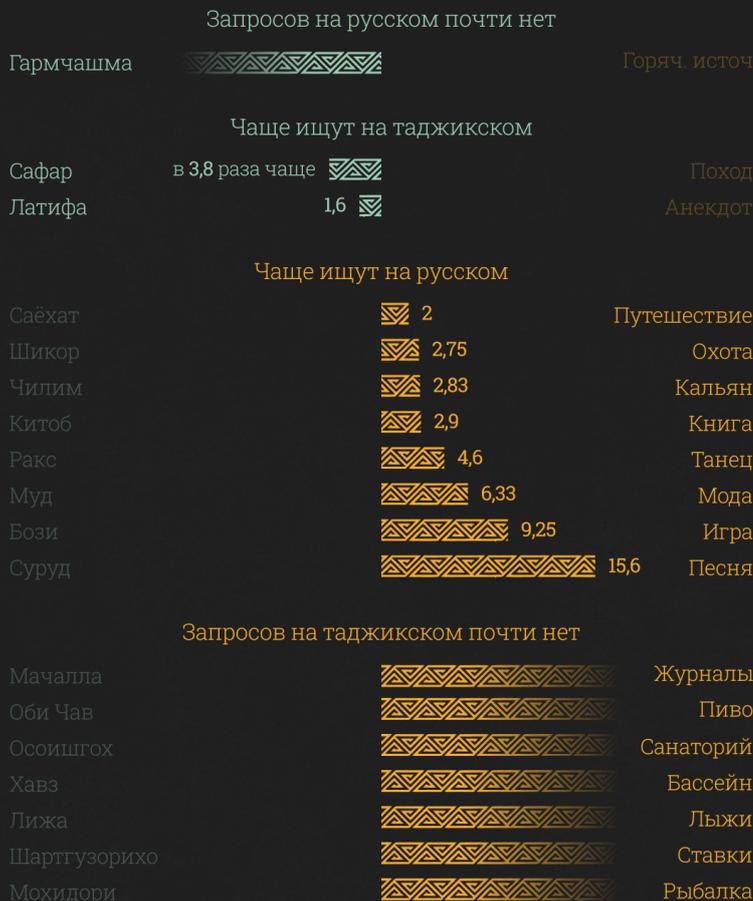
Из развлечений на таджикском ищут «анекдоты», «горячие источники» и «походы». «Песни», «танцы» и даже «книги» в основном ищут на русском языке.

# На каком языке чаще делают запросы о развлечениях и отдыхе в Google?

Среднее соотношение запросов из Таджикистана на русском и таджикском языках за последние 12 месяцев\*

\* Дата извлечения данных: 22.03.2023

Источник: Google Trends



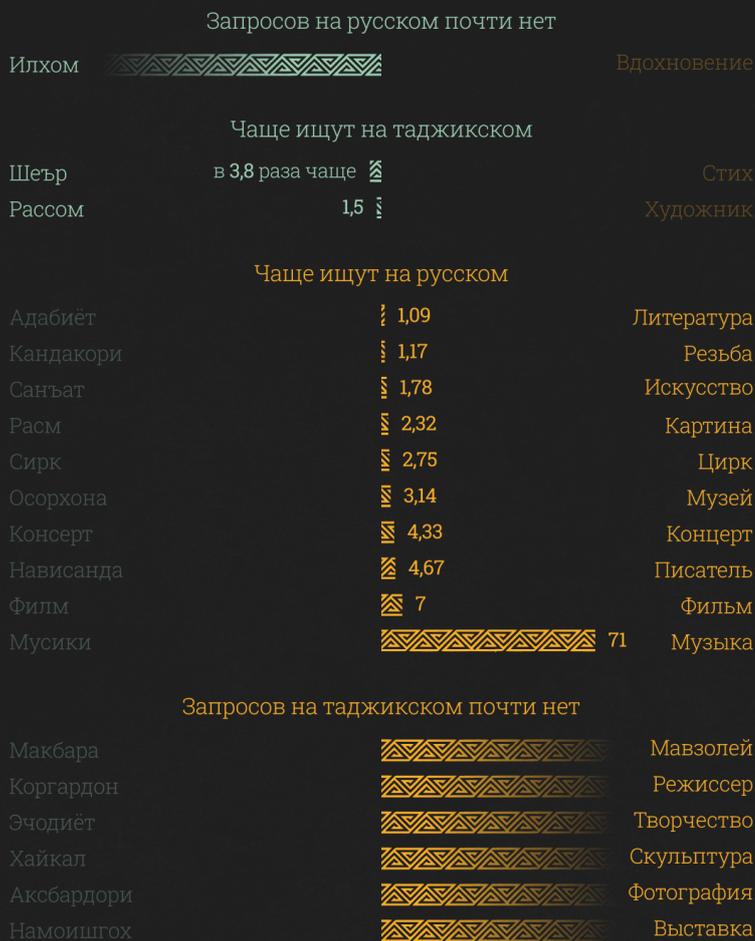
В категории «культура и искусство» выделяются два слова: «музыка» и «вдохновение». Музыку почти не ищут на таджикском, и это не удивительно: даже на таджикских музыкальных порталах вся информация публикуется на русском языке. А вот «вдохновение», напротив, почти не ищут на русском. Правда, дело не столько в том, что о вдохновении ищут только на родном языке. Запрос «илхом» — это не только «вдохновение», но и популярное имя. На первых строчках Google по нему — исполнитель Илхом Муродов с песнями о мигрантах и чужестранцах — это один из популярных нарративов в таджикской музыкальной индустрии.

# На каком языке чаще делают запросы о культуре и искусстве в Google?

Среднее соотношение запросов из Таджикистана на русском и таджикском языках за последние 12 месяцев\*

\* Дата извлечения данных: 22.03.2023

Источник: Google Trends



## Возвращение к корням

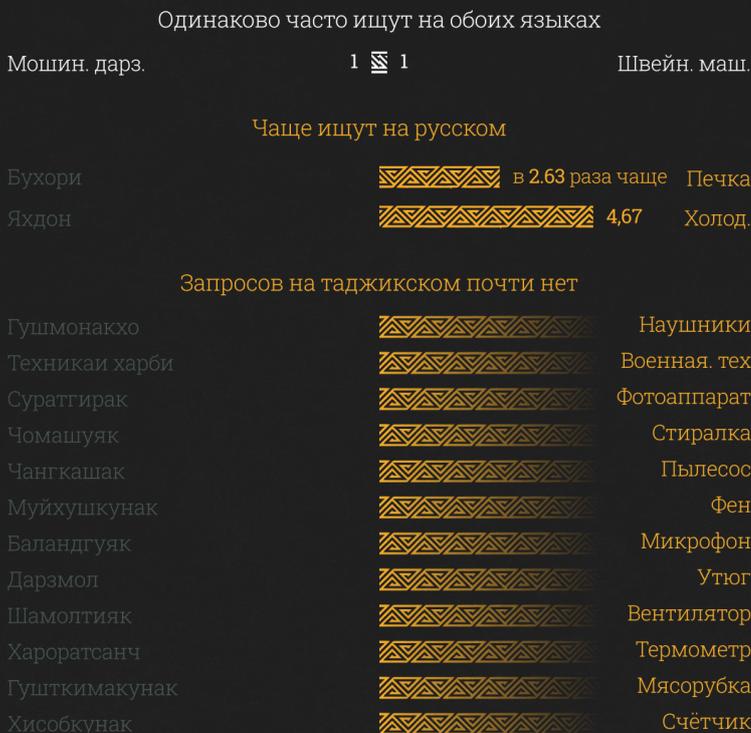
Актуальной и полезной информации на таджикском языке крайне мало. Однако ярче всего эта проблема заметна, если искать в Google названия техники и предметов быта.

# На каком языке чаще ищут информацию о технике в Google?

Среднее соотношение запросов из Таджикистана на русском и таджикском языках за последние 12 месяцев\*

\* Дата извлечения данных: 22.03.2023

Источник: Google Trends



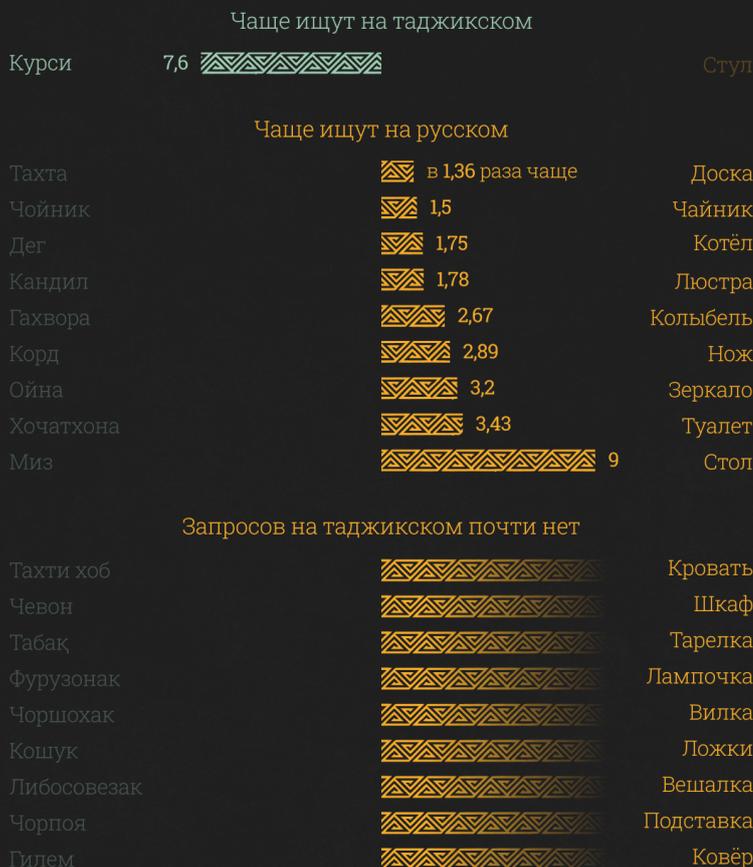
Ни одно слово в категории «техника» не искали чаще на таджикском языке. Что хуже, большинство из этих слов на таджикском искать даже не пытаются. Единственное слово в категории «предметы быта», которое встречалось чаще в таджикоязычном написании, вероятно, получило свой результат по ошибке. При запросе «курси» (стул) в результатах Google чаще всего появляются ссылки на курс валют, образовательные курсы и главу из Корана. Google не смог корректно идентифицировать таджикское слово, вместо этого предполагая, что это либо опечатка в русском слове, либо арабское название.

# На каком языке чаще делают запросы о предметах быта в Google?

Среднее соотношение запросов из Таджикистана на русском и таджикском языках за последние 12 месяцев\*

\* Дата извлечения данных: 22.03.2023

Источник: Google Trends



Это вскрывает одну из проблем использования кириллицы для языков, на которых говорит относительно небольшое количество людей: алгоритмы поисковых систем как будто игнорируют их существование.

Но стоит ли тогда переходить на персидскую или какую-то иную письменность? Как отмечает медиалингвист **Кутбиддин Мухтори**, этот вопрос остается спорным и требует дополнительного обсуждения:

«Многие персоязычные страны мира пользуются арабским алфавитом, но арабский алфавит всё-таки не является таджикским. Если мы вообще хотим менять письменность, если хотим возродить свою идентичность, возвращаться стоит к нашему древнему согдийскому алфавиту. Именно это был наш язык, который мы не смогли сохранить, и в итоге перешли на арабский.

Но любое масштабное изменение языка обойдётся нам крайне недёшево. В течение

70-80 лет основную часть научного и литературного наследия мы уже перевели на кириллицу. Население придётся переучивать, придётся отстаивать важность таких изменений.

С другой стороны, доступа к современным знаниям на родном языке у нас нет. А другие персоязычные страны в этом преуспели — например, Иран быстро переводит новую мировую литературу на персидский. С этой стороны переход на персидскую письменность принёс бы людям большую пользу».

Ещё один аргумент против смены языка — на персидском будет по-прежнему сложно продвигать местную культуру в мире. Мухтори считает, что рассказывать другим странам об истории и достижениях Таджикистана лучше всё-таки на русском или английском.

## **Главные выводы**

Таджикистанцы привыкли использовать Google на русском, и наше исследование это ярко подтвердило. Единственная сфера жизни, где ощущимую долю контента ищут на родном языке — это религия.

Такая ситуация сложилась по нескольким причинам:

Качественных текстов на родном языке очень мало, а те сайты, которых их публикуют, зачастую плохо индексируются в поиске. Пользователи поисковых систем используют русский, чтобы не потерять часть информации.

На многих сайтах и в операционных системах русский язык выставлен по умолчанию. Пользователь, привыкая к русскоязычному интерфейсу, и сам начинает «разговаривать» с сайтами по-русски.

Поисковые системы «пессимизируют» таджикские слова, похожие на русскоязычные. Как со словом «курси», по которому Google предлагает посмотреть курсы валют. В таком случае поисковая выдача не соответствует тому запросу, который ей задают.

*Можем ли мы это изменить?*

Чем больше хорошего контента в интернете будет на родном языке, тем более вероятно, что пользователь привыкнет его искать. Чем больше будет появляться текстов на таджикском, тем меньше вероятность того, что поисковые системы будут считать наши слова «ошибками» в русских. Именно большое количество контента о религии стало причиной интереса к этой теме на родном языке. То же самое может

произойти с другими сферами жизни, если приложить к этому много усилий.

Как мы считали

Для анализа поисковых запросов в Таджикистане мы использовали [Google Trends](#). Это официальный портал компании, на котором можно узнать относительную частотность любого запроса. Google не раскрывает точного числа запросов по ключевым словам, однако позволяет сравнивать разные поисковые запросы и их динамику с течением времени.

Мы выделили 13 сфер, описывающих повседневную жизнь, и для каждой сферы придумали по 20 слов, имеющих разное написание в русском и таджикском языках. Список слов, которые мы использовали для анализа, вы можете найти в Excel-файле, доступном [по этой ссылке](#).

Для каждой пары поисковых запросов (на русском и на таджикском) мы получили пару индексов (от 1 до 100) и посчитали отношение индексов друг к другу. Если индекс таджикского слова был выше, то делили таджикский индекс на русский, а если нет, — то наоборот. В некоторых случаях по одному из языков не было данных — это означало, что запросов на этом языке практически не было.

*Также в ходе анализа мы столкнулись с рядом сложностей:*

Некоторые слова пишутся одинаково в обоих языках, но имеют разные значения. Например, таджикское слово «май» (вино) является омонимом русского слова, обозначающего месяц. Мы старались не брать такие слова в исследование, заменяя их на похожие. Например, вместо «май» брали «шароб» (алкоголь).

Кроме того, сложности возникали при переводе некоторых слов на таджикский без контекста. Например, слова «платье» и «рубашка» на таджикском языке пишутся одинаково — «курта». В таких случаях в русском языке мы сравнивали оба русских слова (платье+рубашка) с таджикским («курта»).

В ряде случаев для таджикского языка мы брали важные словоформы. Например, для сопоставления термина «завтрак» мы использовали таджикские термины «субхона» и «субхона». Таким образом, Google Trends объединял поисковые запросы, включающие оба этих ключевых слова, и сравнивал их с запросом «завтрак». В некоторых случаях, из-за ошибок на платформе Google, объединение запросов давало отсутствие результата. Тогда мы оставляли только одну таджикскую словоформу.

*Отсутствие данных:*

По некоторым словам Google Trends выводил сообщение «данных по этому запросу слишком мало». Это означает, что за последние 12 месяцев запросов на обоих языках было недостаточно. Посчитать их соотношение мы не смогли. Далее мы укажем полный список запросов, которые нам пришлось исключить из исследования:

что запросов за последние 12 месяцев было очень мало как на русском, так и на таджикском языках чтобы вывести их соотношение. Запросы с недостаточным количеством данных были исключены с инфографик, однако мы указываем их здесь с разбивкой на категории.

*Медицина и болезни:* повышенное давление (фишори баланди хун), сердечный приступ (сактаи дил), иглоукалывание (сузанзани).

*Одежда и атрибутика:* женское платье (либоси занона), нижнее бельё (либоси таг).

*Экономика и политика:* нижняя палата (мачлиси намояндагон + маҷлиси намояндагон).

*Развлечения и отдых:* компьютерные игры (бозихои компютери + бозихои компютерӣ), развлекательный центр (маркази дилхуши + маркази дилхушӣ).

*Религия:* язычество (бутпарасти + бутпарастӣ).

*Культура и искусство:* каллиграфия (хушнависи + хушнависӣ).

*Предметы быта:* половник (кафлез).

*Техника:* кофеварка (кахвачушонак + қаҳвачӯшонак), соковарка (шарбатафшурак), посудомойка (зарфшуяк), электропила (арраи электрики + арраи электрикӣ), электробритва (риштароши барки + риштароши барқӣ).



Если вы нашли ошибку, пожалуйста, выделите фрагмент текста и нажмите *Ctrl+Enter*.