

Despite the process of derussification in Tajikistan, the Russian language plays a crucial role in the country's information space. According to a study by Navruz Karimov, the overwhelming majority of search queries to Google in Tajikistan are made in Russian. This data speaks not only to the lack of useful content in Tajik but also to the poor prospects for the Tajik language.



For this study, we selected 260 words and phrases from 13 different spheres of life. In the last year, these words were searched on average 4.13 times more often in Russian than in Tajik, excluding those not searched in Tajik at all. Unfortunately, there are also a few of these, for example, in the category «clothes and accessories» 11 out of 20 words were searched exclusively in Russian.

# What language is more often used for searches about clothes and accessories on Google?

The average ratio of queries from Tajikistan in Russian and Tajik languages over the last 12 months\*

\*Data extracted on 22.03.2023

Source: Google Trends



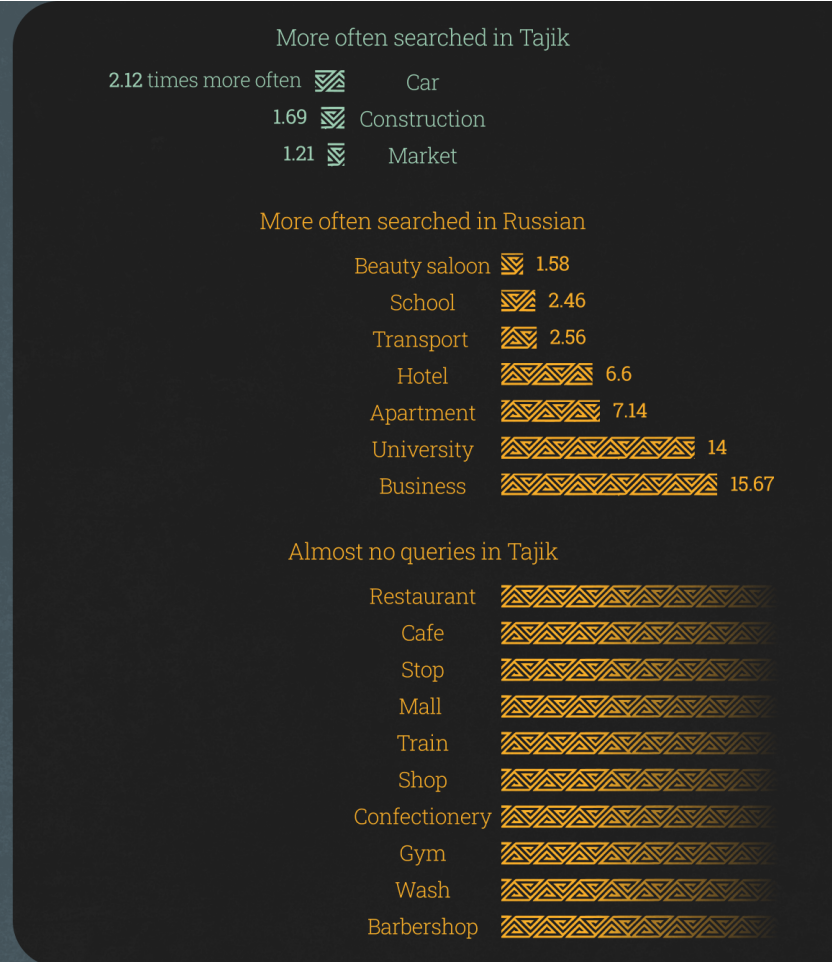
The same pattern is observed when searching for information about urban infrastructure. Exactly 50% of the terms in this category are almost never 'googled' in Tajik.

# In what language municipal facilities are searched more often on Google?

The average ratio of queries from Tajikistan in Russian and Tajik languages over the last 12 months\*

\*Data extracted on 22.03.2023

Source: Google Trends



According to **Muhammadi Ibodulloyev**, head of the Civil Initiative for Public Fund Civil Internet Policy Initiative, the low number of search queries in the Tajik language is caused by a complex set of reasons:

«First, there is little Tajik-language content on the internet. News agencies, which supply most of the relevant material, still mostly write in Russian. So to look for something in Tajik is to lose access to some of the search information.»

Ibodulloyev also drew attention to the fact that users sometimes unconsciously use Russian on the internet because they are trying to follow interface rules:

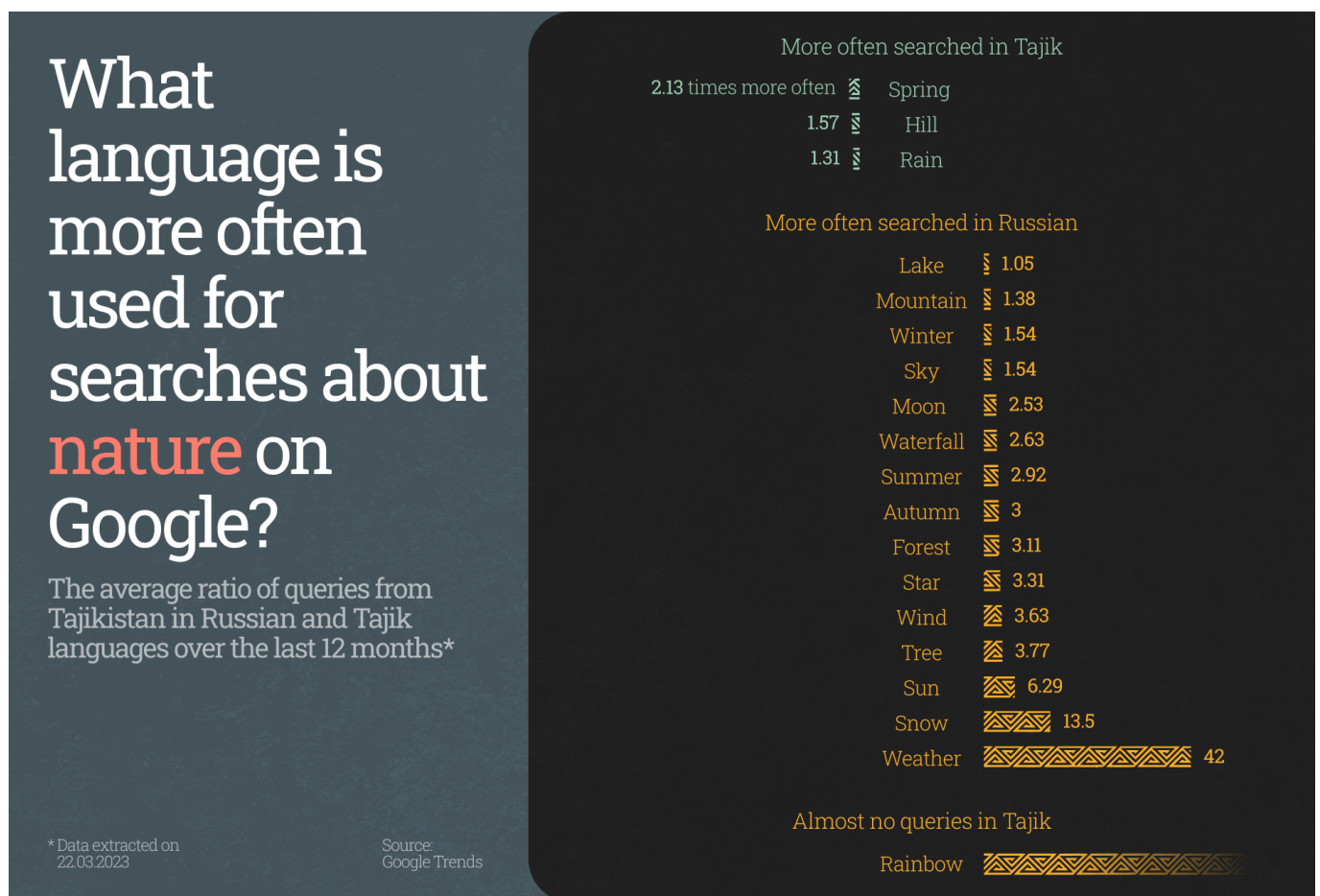
«Users usually don't go the hard way. If a website opens in Russian by default, few people will look for the language switch. The same goes for operating systems — even though Windows has had a Tajik interface for a long time, many still use Russian.»

The browser and various websites, including Google, read the system language, interpret



the interface and offer content in Russian more often. As a result, users are more likely to use Russian, because all communication with the website is in Russian. It's a vicious circle.»

A few years ago, a search for «obu havo» (weather) could return outdated weather data. That problem is a thing of the past now, and Google loads its interactive dashboard in response to that query. However, users are still more likely to search for the weather in Russian – partly out of habit, partly because due to system settings the weather forecast information is still loaded in Russian (or in English if the system language is English).



## Features of algorithms

Other Google products also poorly support the Tajik language. For example, Google Translate fails in the translation of many topics. The reason is that the number of digitized texts on which the translation algorithm can «learn» is too small. However, experts note that IT terminology in the Tajik language should be handled by the translator better than the others:



«We were once approached by a specialist from Khujand who was dealing with the development of Google Translate,» recalls Muhammadi Ibodulloev. — «We provided him with a series of training texts in the field of information technology. Once these texts were specially translated from English into Tajik to help develop the IT industry in the country. And now they are helping to improve the quality of translation at Google.»

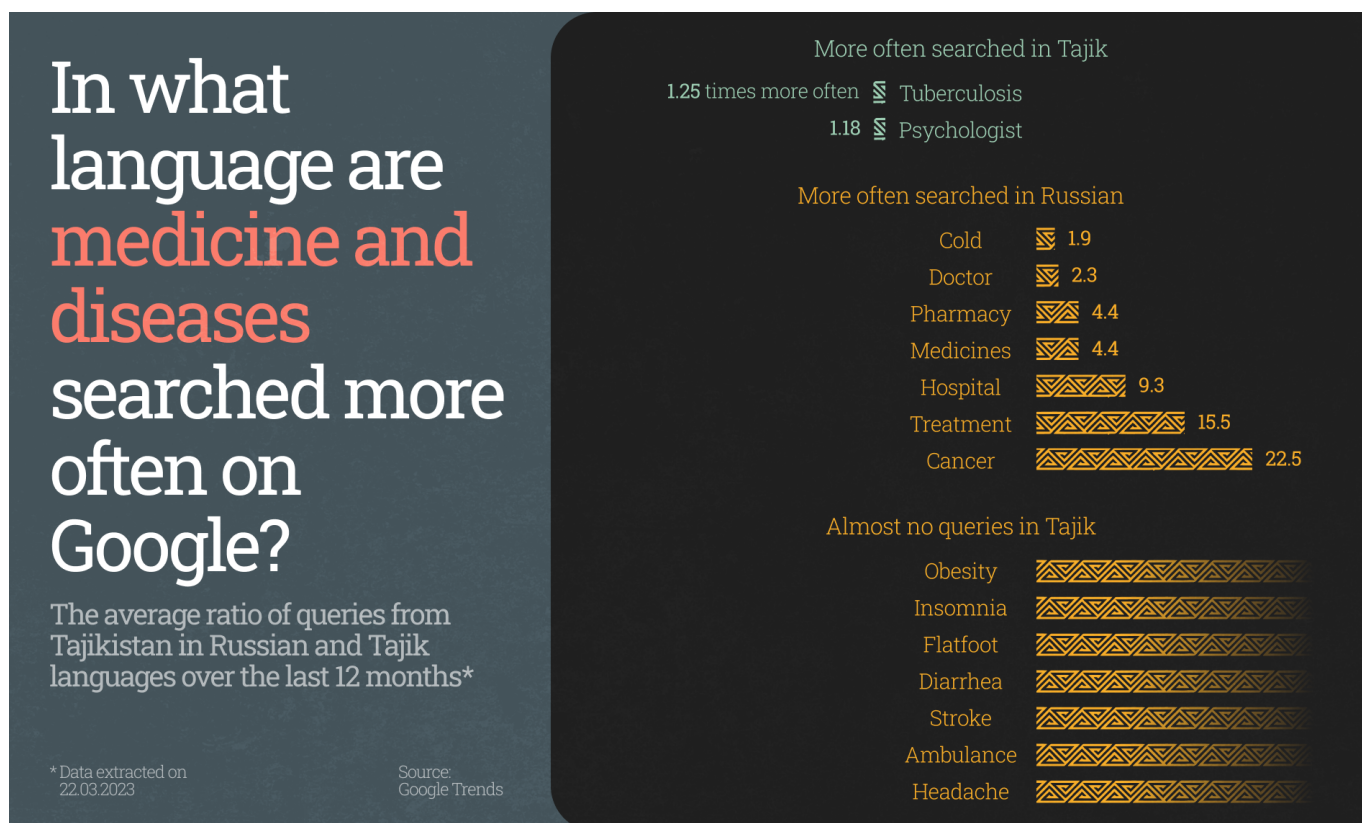
Google’s problems with Tajik are not due to its complexity, Ibodulloyev said. «The Tajik language is more ‘mathematical’; algorithms find it easier to process data in Tajik than in Russian or English,» Ibodulloyev said. — «Despite this, the efficiency of search engine algorithms also depends to a large extent on how popular the language is, and on the quality of data available to train these algorithms. There is little data in the Tajik language, and this makes it difficult for algorithms to operate.»



Analysis of queries in Google shows that terms that relate to traditional culture (e.g. national dishes) are more often searched in the Tajik language. And widespread words related to food and drink, such as «tea», «meat» and «fruit» are searched in Russian.

## Prospects for the Tajik language

All of the experts we spoke to during the writing of this piece believe that the lack of quality content in Tajik in important areas of science could hinder its further development. And such a problem does exist: only two words in the category ‘medicine and disease’ from our selection were searched more frequently in Tajik — «tuberculosis» and «psychologist».



In a related topic, human anatomy, there are also problems. Information about organs and body parts is searched mainly in Russian. Only three words were searched slightly more often in the Tajik language.

# In what language are organs and body parts searched more often on Google?

The average ratio of queries from Tajikistan in Russian and Tajik languages over the last 12 months\*

\*Data extracted on 22.03.2023

Source: Google Trends



Language expert **Umed Jayhoni** expresses a pessimistic view of the future of the Tajik language, arguing that it is stuck in development. Tajikistan has a Language and Terminology Committee and the Institute of Tajik Language and Literature, which regulate the introduction of new words in the language, but «the way the language exists now, it has no future,» Jayhoni declares.

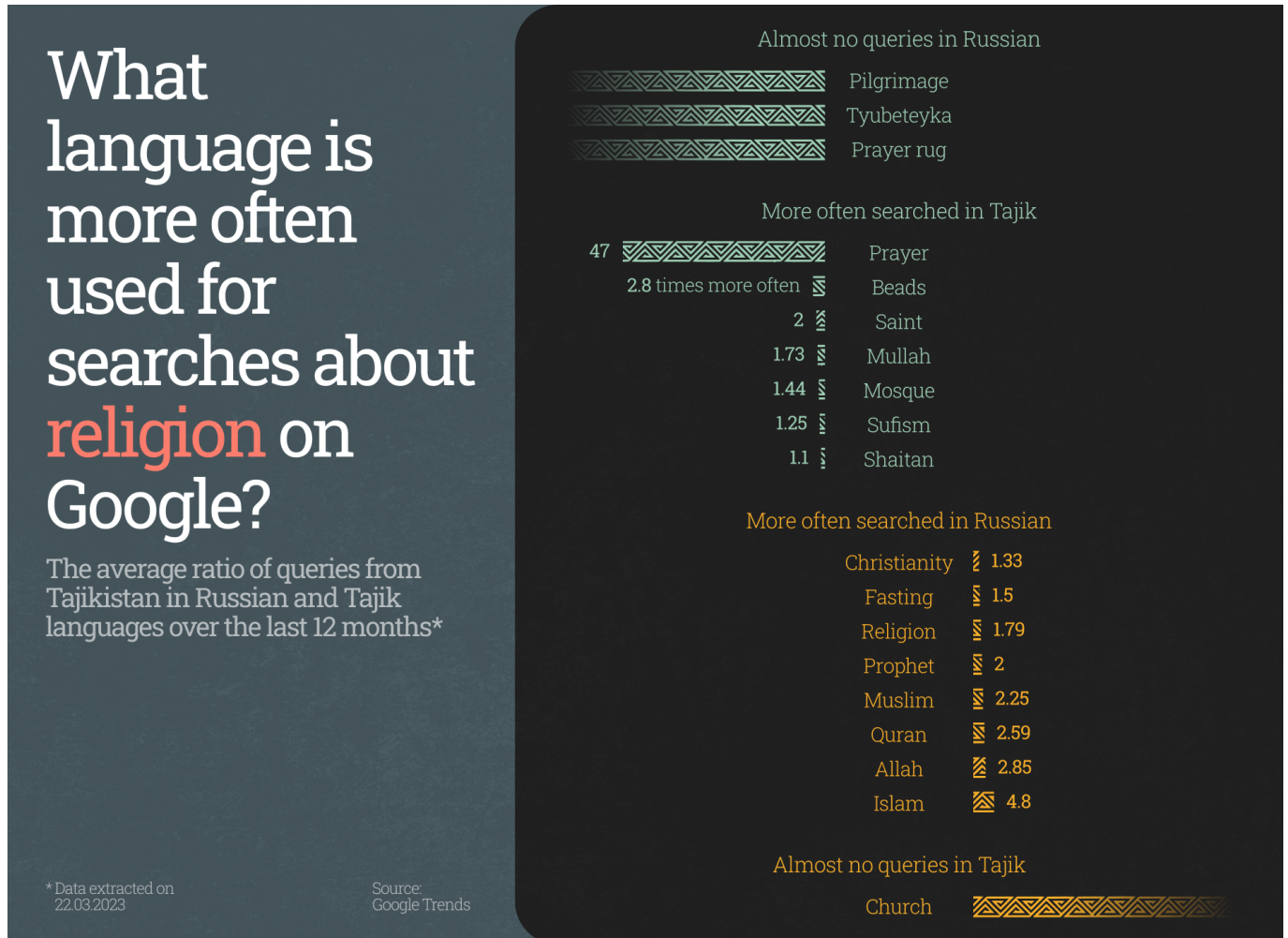
«I developed and proposed an original Tajik system of military ranks for the armed forces of Tajikistan, but it has not yet been approved and they still use Soviet-Russian ranks because our generals still think in terms of Soviet internationalism.»

Jayhoni actively creates content in Tajik, but he admits that he hardly ever seeks information in it: «I just know there is nothing I need on the internet in Tajik. We have books in our native language, but they are not digitized. So I have to use Russian-language sources.

Interestingly, despite the general trend, queries on religion are often made in Tajik. For



example, information on pilgrimages and Islamic attributes such as «prayer rug» and «tubeteyka» are not even sought in Russian. This may indicate that the Tajik-speaking population is particularly interested in Islamic topics, and content on this topic is made immediately in the mother tongue.



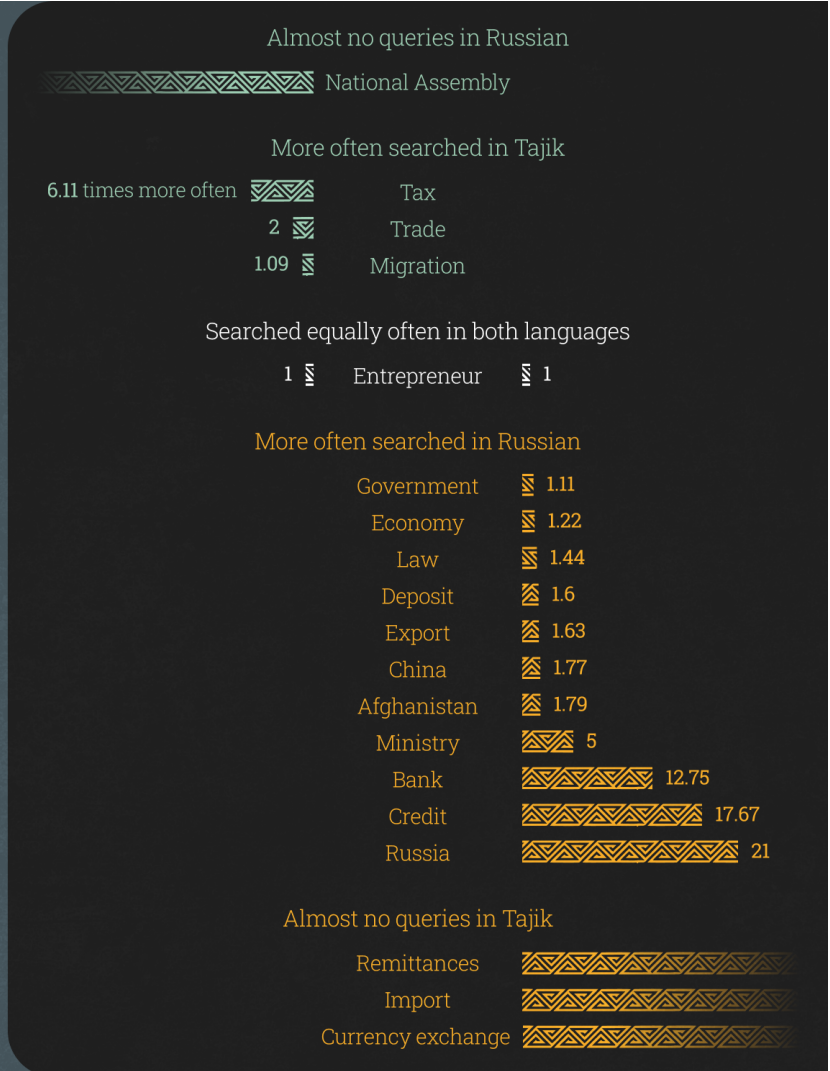
Another peculiarity of Tajik-language content is that it is apolitical. Even information about the economy and politics, with rare exceptions, is searched in Russian. Only the most pressing topics for Tajikistan — migration, taxes and trade — are searched in the native language. The word «sohibkor» (“entrepreneur”) is on an equal footing in searches, but this does not necessarily indicate the development of relevant content in the region. Rather, it is the Uzbek football club Sohibkor and streets with the same name that add ‘points’ to the queries in the Tajik language.

# What language is more often used for searches about economy and politics on Google?

The average ratio of queries from Tajikistan in Russian and Tajik languages over the last 12 months\*

\*Data extracted on 22.03.2023

Source: Google Trends



## «Non-native» entertainment

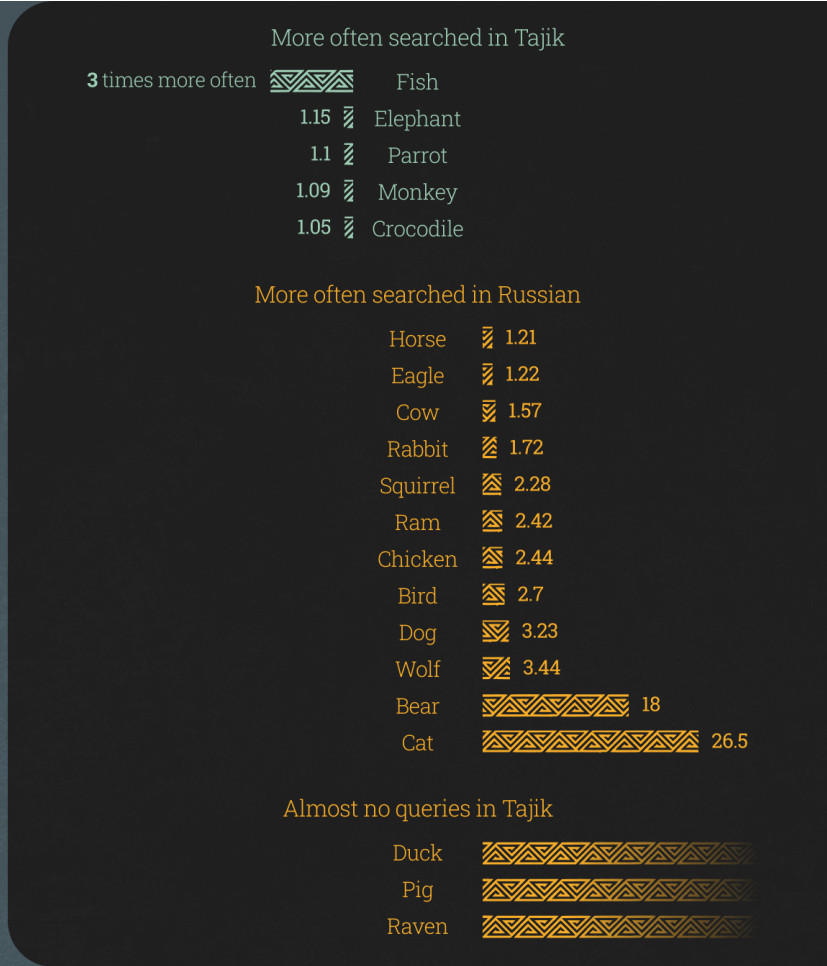
Quality entertainment content in the Tajik language is also scarce. Surprisingly, even the data on how animals are searched in Tajikistan can tell a story.

# What language is more often used for searches about animals on Google?

The average ratio of queries from Tajikistan in Russian and Tajik languages over the last 12 months\*

\*Data extracted on 22.03.2023

Source: Google Trends



Cats are 18 times and bears 26 times more frequently searched specifically in Russian. This is explained by the fact that when you type the word «cat» into Google, the cartoon «Three Cats» appears among the first search results. In the case of «bear», it is «Masha and the Bear». This is how Russian cartoons have managed to squeeze local content out of the search results.

*Top search results for «cat» in Russian. Tajik.*

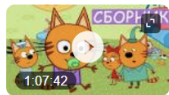
*Top search results for «cat» in*



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 1ТООУ 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

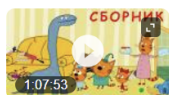


... попал в книгу рекордов Гиннеса. Видео с ним в Интернете посмотрели более 330 миллионов раз. В чем секрет популярности КОТА и сколько ...

YouTube · Сейчас · 28 мар. 2017 г.

www.youtube.com › watch

## Три Кота | Сборник невероятных серий - YouTube



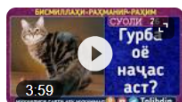
В этом сборнике вас ждут нешуточные приключения Коржика, Карамельки и Компота: от соревнований в меткости и настоящего ...

YouTube · Три Кота · 2 мар. 2021 г.

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

www.youtube.com › watch

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

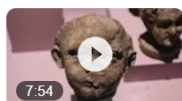


Суоли-26 Гурба оё начас аст? 9.5K views 3 years ago. Абу Мухаммад Мадани أبو محمد المدني. Абу Мухаммад Мадани أبو محمد المدني.

YouTube · أبو محمد المدني · 18 июн. 2019 г.

www.ozodi.org › ... · Перевести эту страницу

## гурба - Радиои Озодӣ

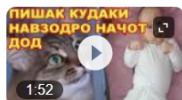


гурба. gurbai da fikr fururaft Гузошта шуд аз чониби илья. Идома. Ҳамаи хабарҳои имрӯзро дар инҷо бихонед. январ 11, 2022 ...

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.

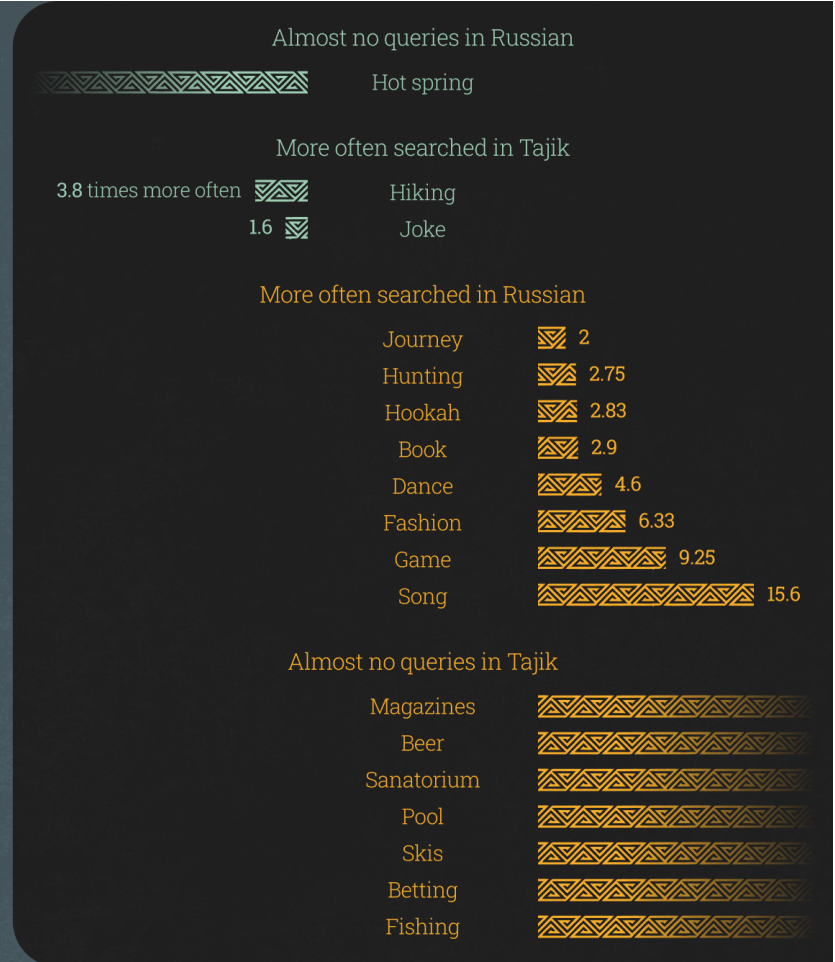
Of the entertainment options, «jokes», «hot springs» and «hiking» are sought in Tajik. «Songs», «dances» and even «books» are mostly being searched for in Russian.

# In what language entertainment & recreation are searched more often on Google?

The average ratio of queries from Tajikistan in Russian and Tajik languages over the last 12 months\*

\*Data extracted on 22.03.2023

Source: Google Trends



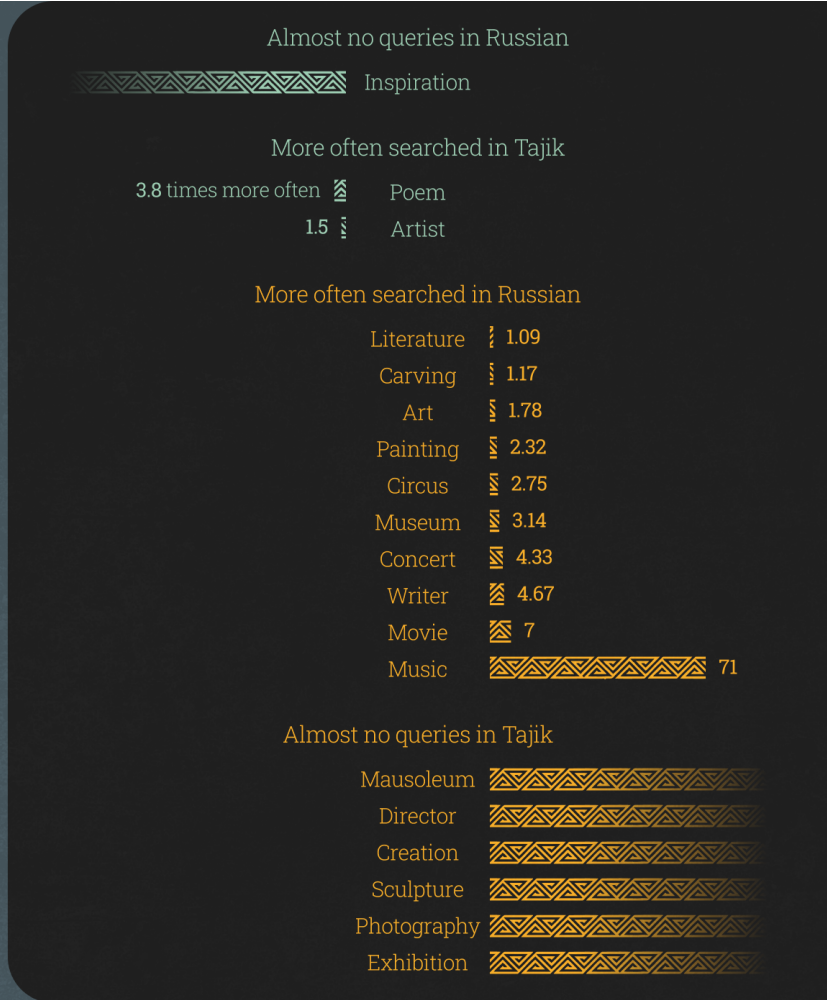
Two words stand out in the culture and art category: «music» and «inspiration». Music is hardly searched for in Tajik, which is not surprising: even Tajik music portals publish all information in Russian. On the other hand, «inspiration» is almost never searched in Russian. However, it's not necessarily because the topics on inspiration are written only in the native language. The word «ilhom» does not only mean «inspiration», it is a popular name. The singer Ilhom Murodov tops Google's list with songs about migrants and strangers — one of the popular narratives in the Tajik music industry.

# In what language culture and art are searched more often on Google?

The average ratio of queries from Tajikistan in Russian and Tajik languages over the last 12 months\*

\* Data extracted on 22.03.2023

Source: Google Trends



## Back to the roots

There is very little relevant and useful information in Tajik. However, this problem is most noticeable when searching for technology and household items on Google.



# In what language **technology** is searched more often on Google?

The average ratio of queries from Tajikistan in Russian and Tajik languages over the last 12 months\*

\* Data extracted on 22.03.2023

Source: Google Trends













Almost no queries in Russian

1 🇷🇺 Sewing machine 🇷🇺 1

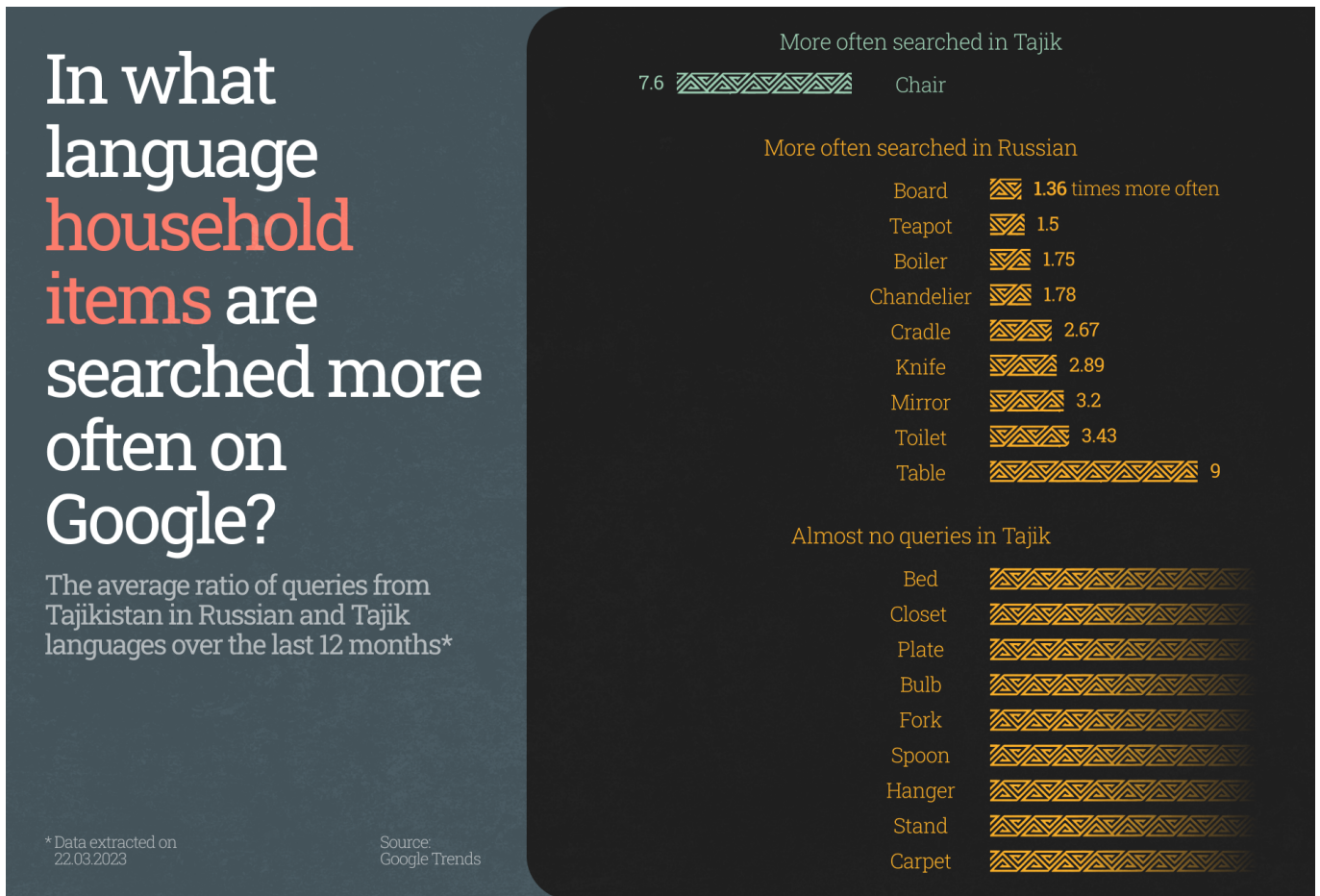
More often searched in Russian

Stove		2.63 times more often
Fridge		4.67

Almost no queries in Tajik

Headphones	
Military equipment	
Camera	
Washing machine	
Vacuum cleaner	
Hairdryer	
Microphone	
Iron	
Fan	
Thermometer	
Meat grinder	
Counter	

Not a single word in the «technology» category was searched more often in Tajik. Moreover, most of these words were never even searched in Tajik. The only word in the «household items» category, which was searched more frequently in Tajik, probably got its result by mistake. When asking for «kursi» (chair), Google results most often show links to currency exchange rates, educational courses and a chapter from the Koran. Google could not correctly identify the Tajik word, instead suggesting it was either a misprint in the Russian word or an Arabic term.



This reveals one of the problems of using Cyrillic for languages spoken by relatively few people: search engine algorithms seem to ignore their existence.

But should we then switch to Persian or some other script? As media linguist **Qutbiddin Mukhtori** points out, this remains a contentious issue and needs further discussion:

“Many Persian-speaking countries in the world use the Arabic alphabet, but the Arabic alphabet is not Tajik after all. If we want to change the script, if we want to revive our identity, we should return to our ancient Sogdian alphabet. This was our language that we were unable to preserve and we ended up switching to Arabic.

But any large-scale change in language will not come cheap. For 70-80 years, we have already translated the majority of our scientific and literary heritage into the Cyrillic alphabet. The population will have to be retrained, we will have to advocate the importance of such changes.

On the other hand, we do not have access to modern knowledge in our language. Other

Persian-speaking countries have succeeded in this — Iran, for example, is rapidly translating the world's new literature into Persian. In this respect, a switch to Persian would be of great benefit.

Another argument against changing the language is that it would still be difficult to promote local culture in Farsi to the world. Mukhtori believes it would be better to tell other countries about Tajikistan's history and achievements in Russian or English.

## **Main conclusions**

Tajiks are accustomed to using Google in Russian, and our study demonstrates this. The only sphere of life where a tangible share of content is searched in the native language is religion.

There are several reasons for this trend:

There is very little quality content in the native language, and those websites that do publish it are often poorly indexed by Google. Search engine users use Russian in order not to miss out on some of the relevant information.

On many websites and operating systems, Russian is the default language. The user gets used to the Russian-language interface and starts «talking» to websites in Russian themselves.

Search engines «pessimize» Tajik words that look like Russian words. As with the word «kursi», for which Google offers to look up currency exchange rates. In this case, the search output does not match the query it is given.

### *Can we change that?*

If there is more good content on the Internet in the native language, the more likely a user will get used to looking for it. The more texts appear in Tajik, the less likely it is that search engines will consider our words to be «mistakes» in Russian. It is the large amount of content about religion that has caused interest in this topic in the native language. The same can happen with other spheres of life if a significant amount of effort is put into it.

## **How did we count it**

We used [Google Trends](#) to analyse search queries in Tajikistan. This is the company's official

portal where it is possible to find the relative frequency of any query. Google does not reveal the exact number of keyword queries, but it does allow us to compare different search queries and their dynamics over time.

We identified 13 spheres describing everyday life, and for each sphere, we came up with 20 words with different spellings in Russian and Tajik. The list of words we used for analysis can be found in the Excel file available [here](#).

For each pair of search queries (in Russian and Tajik) we obtained a pair of indices (from 1 to 100) and calculated the ratio of indices to each other. If the Tajik word index was higher, we divided the Tajik index by the Russian one, and if not, it was vice versa. In some cases, there was no data for one of the languages - this meant that there were almost no queries in this language.

*We also encountered several difficulties during the analysis:*

Some words are spelt the same way in both languages but have different meanings. For example, the Tajik word «mai» (wine) is a homonym for the Russian word for a May month. We tried not to take such words in the study, replacing them with similar ones. For example, instead of «may» we took «sharob» (alcohol).

In addition, difficulties were encountered in translating some words into Tajik without context. For example, the words «dress» and «shirt» are spelt the same way in Tajik — «kurta». In such cases, we compared both Russian words («dress + shirt») with Tajik («kurta»).

In many cases, we have taken important word forms for the Tajik language. For example, we used the Tajik terms «subkhona» and «subhona» to match the term «breakfast». Thus, Google Trends combined searches that included both of these keywords and matched them with «breakfast». In some cases, due to bugs in Google's platform, combining the queries yielded no results. Then we left only one Tajik word form.

*Lack of data:*

For some words, Google Trends displayed the message «there is too little data for this query.» This means that there were not enough queries in both languages in the last 12 months. We were not able to calculate their ratio. Below we present the full list of queries that we had to exclude from the study:

*Medicine and diseases:* hypertension, heart attack, acupuncture.



*Clothes and accessories:* frock, underwear.

*Economy and politics:* Assembly of Representatives.

*Entertainment & recreation:* computer games, entertainment centre.

*Religion:* paganism.

*Culture and art:* calligraphy.

*Household items:* ladle.

*Technology:* coffee machine, steam juicer, dishwasher, electric saw, electric shaver.



If you have found a spelling error, please, notify us by selecting that text and pressing *Ctrl+Enter*.