

sanctions.io Conquers Cyrillic and Arabic Name Screening with Babel Street Match

Anti-money laundering and sanctions screening regulations seem like something that only banks and other money processing organizations must worry about. However, individuals and corporations small and large must also follow local sanctions regulations, from industrial manufacturers¹ to software providers.²

sanctions.io initially sought to fill an unaddressed need for commercial sanctions screening without the complicated pricing and lengthy contracts required by other providers. Seeing an opportunity, sanctions.io started offering a high-quality sanctions screening service with transparent pricing and simple pay-per-use contracts. Their service compiles data from many sources, cleans it, and makes it available in a standard format that is accessible and searchable through an API.

The Challenge

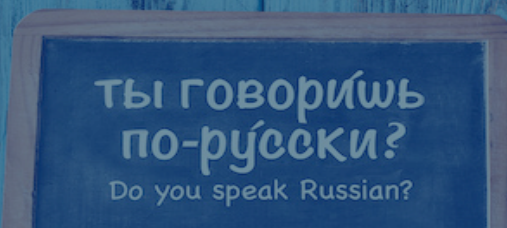
sanctions.io started with an in-house built name matching solution with the goal of offering better search than other sanctions screening solutions. What they developed was “not too bad for entity search,” but they were not satisfied.

“We started to look at name matching solutions on the market, but they were either complicated to implement or didn’t perform well,” Thorsten J. Gorny, co-founder and CEO of sanctions.io, said. “Their false positive rates were around 10% to 15% and we were looking for around 5% with zero false negatives.”

Another issue involved name transliteration. Transliteration is the process of converting names from one script to another based on pronunciation or character mappings. Arabic transliterations were a major issue for sanctions.io, and later so were Cyrillic names (found in Russian, Ukrainian, and many Central Asian languages).

¹ Ostrovsky, Simon, “American company accused of violating sanctions, doing business with Russian arms industry,” March 14, 2023 PBS News Hour. [pbs.org/newshour/show/american-company-accused-of-violating-sanctions-doing-business-with-russian-arms-industry](https://www.pbs.org/newshour/show/american-company-accused-of-violating-sanctions-doing-business-with-russian-arms-industry)

² U.S. Attorney’s Office District of Massachusetts “SAP Admits to Thousands of Illegal Exports of Its Software Products to Iran and Enters Into Non-Prosecution Agreement with DOJ,” April 29, 2021. [justice.gov/usao-ma/pr/sap-admits-thousands-illegal-exports-its-software-products-iran-and-enters-non](https://www.justice.gov/usao-ma/pr/sap-admits-thousands-illegal-exports-its-software-products-iran-and-enters-non)



Gorny gave the example of the English transliteration of the Russian surname Лукашенко, which is Lukashenko, while the German transliteration is Lukaschenkow. Name matching systems using the edit distance algorithm (which calculates the number of character differences between two words; two in this case) need to set a very low minimum score to find this type of match and consequently produce a lot of false positives.

What ultimately motivated sanctions.io to look for a better solution was the Russian invasion of Ukraine in February 2022, which caused the

U.S., EU, and G7 countries to impose sanctions on Russia.³ Suddenly sanctions.io had urgent customer demands to improve matching of names written in Cyrillic.

“Well over half of sanctions data is non-Latin scripts with Chinese, Arabic, and Cyrillic being the largest proportion,” Gorny said. “We needed accurate Cyrillic name matching right away. Our customers couldn’t wait for us to develop it.”

Too Many Transliteration Standards

Transliteration standards are intended to promote consistency in spelling, but what happens when there are too many standards?

While the Cyrillic spelling of the Boston Marathon bomber’s surname Царнаев is unambiguous, its possible transliterations include Tsarnaev, Tsarnayev, Sarnaev, and Carnaev.

The multitude of transliteration standards for Cyrillic and other scripts such as Arabic, Hebrew, Chinese, Korean, and Japanese means that inconsistent transliterations in watchlists and documents are the norm. Laypeople write names as best they can based on the sound of a name, which often does not map to a unique spelling. In addition to an ISO transliteration standard (ISO 9:1995), the U.S. government has three more: BGN (Board of Geographic Names); Undiacritized BGN; and IC (Intelligence Community).

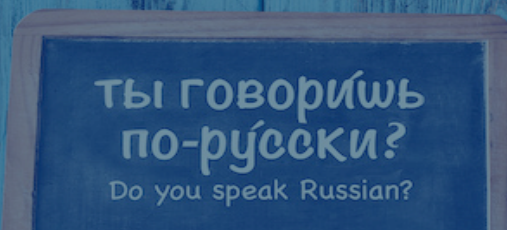
Using these four standards, Царнаев is transliterated to Latin characters as follows:

- Tsarnayev or Charnae — according to the BGN, Undiacritized BGN, and IC
- Carnaev — according to ISO 9:1995

But the passport of the Boston Marathon bomber spelled his name “Tamerlan Tsarnaev.”



³ The White House “FACT SHEET: United States, G7 and EU Impose Severe and Immediate Costs on Russia,” April 6, 2022. [whitehouse.gov/briefing-room/statements-releases/2022/04/06/fact-sheet-united-states-g7-and-eu-impose-severe-and-immediate-costs-on-russia/](https://www.whitehouse.gov/briefing-room/statements-releases/2022/04/06/fact-sheet-united-states-g7-and-eu-impose-severe-and-immediate-costs-on-russia/)



ТЫ ГОВОРИШЬ
ПО-РУССКИ?
Do you speak Russian?

The Solution

sanctions.io tested three commercial solutions searching for:

- More accurate matching for reduced false positives
- Transliteration issues, particularly for Cyrillic and Arabic
- Simpler implementation of name matching

Babel Street Match stood out for accuracy that reduced false positives and its full-featured handling of cross-lingual matching and transliteration spelling variations, particularly for Arabic and Cyrillic.

sanctions.io deals with lots of messy name fields. Consider one entered as “Tarte Normande, Inc. LLC California.” During the first part of the process, sanctions.io cleanses the entity name field to remove extraneous information. Then Match indexes the names

for search through its plug-in to Elasticsearch. Because Match works in tandem with Elasticsearch, sanctions.io benefits from the search engine’s infrastructure for scalability and deployment into a production environment.

Match features a wide range of configuration and parameter options, so users can adapt its name matching behavior to fit their matching needs. With all this, sanctions.io was able to optimize the best starting parameters for its customers.

“We decided on Match simply because it was the most comprehensive product on the market and just performed very well,” Gorny said.

“Doing a search with just a name and country of residence already performs quite well in terms of false positive rates because of our data cleansing and Match.”

Well over half of sanctions data is non-Latin scripts with Chinese, Arabic, and Cyrillic being the largest proportion. We needed accurate Cyrillic name matching right away.

— Thorsten J. Gorny, *co-founder and CEO of sanctions.io*

We decided on Match simply because it was the most comprehensive product on the market and just performed very well. Doing a search with just a name and country of residence already performs quite well in terms of false positive rates because of our data cleansing and Match.

— Thorsten J. Gorny, *co-founder and CEO of sanctions.io*

The Impact

“I believe that matching is one of the key issues of sanctions compliance,” Gorny said. “For me, the top priority was improving the quality of our product in terms of low false positives and no false negatives. We want to be able to confidently tell our customers that they have access to top-notch matching technology.”

With Match, sanctions.io’s business has grown far beyond small-scale customers to customers of all sizes who are seeking cost-effective, high-quality search results.

“Especially since late 2022, we have seen a lot of interest in our product from very large companies, regulators, and across all industries,” Gorny said.



Babel Street is the trusted technology partner for the world’s most advanced identity intelligence and risk operations. The Babel Street Insights platform delivers advanced AI and data analytics solutions to close the Risk–Confidence Gap.

Babel Street provides unmatched, analysis-ready data regardless of language, proactive risk identification, 360-degree insights, high-speed automation, and seamless integration into existing systems. We empower government and commercial organizations to transform high-stakes identity and risk operations into a strategic advantage.

Learn more at babelstreet.com