BABELSTREET

CASE STUDY

More Relevant Search for Due Diligence

Red Flag Group provides global due diligence services covering 194 countries and 54 languages. Its full menu of services offers everything the compliance officer of a Fortune 1000 company needs: watchlist and adverse media screening; vendor reports; supplier onboarding risk scores; and compliance training. Red Flag Group connects technology and human expertise to predict, identify, monitor, and manage every risk to a client's integrity.

Red Flag Group's IntegraWatch platform allows compliance officers to screen the names of people and organizations against international sanctions and watch lists, politically exposed persons databases, state-owned entity databases, and adverse media from every country in the world. Their data is summarized and categorized into risk areas by human analysts who speak the local language and understand conditions in a given country.

The Challenge

"The biggest frustration of our customers is sifting through all the irrelevant results [false positives]," Paul Johnson, Product Director of IntegraWatch at Red Flag Group, said. "The old system was a Soundex [phonetic search] that brought back a lot of false positives that seemed to have no relationship to the desired record."

Getting too many irrelevant search results was expensive and laborious for clients to review. A given client might be monitoring hundreds or thousands of names using Red Flag Group's ongoing monitoring capability, which rescreened search terms every 24 hours. This situation often resulted in new sets of "hits" or notifications that required remediation by the client.

"Clients want to be notified when 'Acme Brick' in Russia was in the news for presenting some sort of client risk, but not for 'Acme Bricklaying' in France or 'Ace Brick' in Russia," Johnson said. "Ideally, we would only return 100% matches. That'd be the perfect system. The biggest dilemma is clients can't not look at all the results."

On the other hand, a missing match (false negative) in search results could be more painful than wasted investigations. A company doing business with a bad actor could damage its reputation or risk high financial penalties from regulators.

A client might search for thousands of names at once, so there isn't time to write name variations to catch cases where a name might appear with a slightly different spelling in one database versus another. Or worse, the database might list the name of a person or company, but write it in a different language or script.

"Multilingual search was a main problem we were trying to solve," Joseph Mantuano, Red Flag Group Senior Developer, said. "For each name we would have had to collect every variation and every translation, but that's just too much work."

Red Flag Group needed a name matching solution to address these challenges:

1. Too many false positives
2. Lack of fuzzy name matching requiring clients to input names and their variations
3. Lack of cross-lingual name matching

Red Flag Group was using Elasticsearch for searches, and its default algorithm using TF/IDF (term frequency–inverse document frequency) was not effective in finding the most relevant results. Red Flag Group had a specialized thesaurus that claimed to do international name matching, but it was still only doing a straight comparison of strings.

"Our older system performed very poorly on non-Latin languages such as Chinese, Arabic, and Japanese," Tim Hawkins, Red Flag Group CTO, said. "The issue is some Chinese companies lead with their English name and some use the Chinese name, so there is a lot of inconsistency in how names are represented."

Furthermore, the name match score from Elasticsearch would change from time to time, depending on the size of the search index.

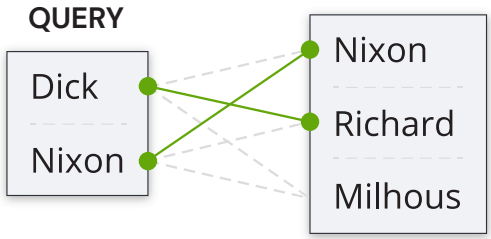
Customers felt unsatisfied that a match might be scored 95% one day, but 90% the next day, after rerunning the same search.

In addition to making clients happy, a better name search would give Red Flag Group a critical competitive edge in its industry.

The Solution

Hawkins and his team considered a larger-scale solution that offers name matching, but it was built for text mining and was not an easy drop-in solution for their search optimization problem. While looking at the text indexer in MongoDB, Hawkins came across Babel Street Match, which is used by MongoDB to expand language support of its search capability.

Match's name matching features come as a convenient Elasticsearch plugin to add fuzzy name search across multiple languages, scripts, and name variations. Essentially, at indexing time, Match encodes fuzzy matching lookup keys for each name to store in a hidden index field. Queried names are similarly encoded and enable fuzzy matching by matching the keys and calculating the similarity using machine learning models. Further, within Match, names are broken up into tokens that are compared to determine the best match alignment.



Babel Street Match compares names on a token-by-token basis, thus overcoming issues of misordered name components or names in the wrong database field.

The relative weight given to each matching or "almost matching" token depends on language - and name-specific characteristics, such as the rarity of a name, same/different gender, and missing tokens.

For matching organizational names in some languages, Match also compares the semantic similarity of words in the name using text embeddings. It allows Match to match entities based on words with similar meanings, rather than only phonetics.

Match hides all these complexities, simply returning a list of results by relevance with match scores calculated.

As each user has different data and goals, Match offers several configuration options, such as balancing the recall (comprehensiveness of the search) with performance (speed), or setting the minimum match threshold (e.g., 80% match) for a name to be returned as "a match."

The vast majority of queries from Red Flag Group clients are in English. But because Red Flag Group covers news sources in 54+ languages, name matches in multilingual content also need to be found. Some of the most important languages are Chinese, Russian, and Arabic. Match's support of multiple languages and cross-language name matching solves this problem.

The Impact

"The killer feature for us in Match was the ability to find Chinese names from English searches and vice versa," Hawkins said. "As customers do more business with China, there is more and more need to do searches on Chinese company names."

With the intelligent name matching of Match within IntegraWatch, Red Flag Group gained a significant competitive edge in customer retention and acquisition. Its customers feel confident that IntegraWatch has them covered, wherever in the world they do business.

> "Multilingual search was a main problem we were trying to solve. For each name we would have had to collect every variation and every translation, but that's just too much work."
>
> — Joseph Mantuano
> *Senior Developer, Red Flag Group*

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**

© 2024 Babel Street. All Rights Reserved.

BABELSTREET