



Strengthening U.S. Borders With Intelligent Name Matching

U.S. Customs and Border Protection (CBP) provides security and facilitation operations at 328 ports of entry throughout the country. CBP takes a comprehensive approach to border management and control, combining customs, immigration, border security, and agricultural protection into one coordinated and supportive activity. The CBP staffers are responsible for enforcing hundreds of U.S. laws and regulations. On a typical day, CBP processes nearly 1 million visitors.

The Challenge

The U.S. Department of Homeland Security (DHS) Targeting and Analysis Systems Program Office (TASPO) wanted to strengthen the screening system used by the CBP. This system scans and matches millions of names a day to secure U.S. borders and facilitate lawful international trade and travel, while enforcing U.S. laws and regulations.

Accuracy is the top priority of many missions using human language technology. For CBP's TASPO office, ensuring that people

THE BOSTON MARATHON BOMBING

In 2011, the FBI issued a detain alert for Tamerlan Tsarnaev, who later became known as one of the two Boston Marathon bomber suspects in April 2013. In January 2012, Tsarnaev traveled to Russia, and in July 2012, he returned to Boston. He was not detained on either occasion. Why?

On that January day, there were too many matching watchlist names for airport agents to process, and Tsarnaev wasn't flagged. When he returned in July, the January detain alert had expired, and a new alert spelled his last name "Tsarnayev." Because his travel documents said "Tsarnaev," his name was not matched to the FBI alert.

entering the country are properly screened ultimately saves lives.

Proper screening requires accurate name matching to flag possible hits on watchlists. To maximize that accuracy, TASPO wanted to integrate different approaches to name matching, so that it could benefit from the strengths of each.

CBP's system in 2013 used a "brute force" approach, generating all possible spelling variations in an attempt to increase possible matches. This approach has weaknesses, such as dealing with not-yet-seen names and names with added or missing spaces.

The Solution

CBP tested and added Rosette Name Indexer, a human language technology-based solution that uses machine learning and carefully weighted algorithms to understand phonetic variations, nickname formation patterns, and name variations within each language. It also links names across different languages and scripts, using knowledge of how a sound in one language may be spelled in a different one.

The Impact

By integrating Rosette technology into its screening system, CBP added greater name matching accuracy — and ultimately greater security — to its border protection efforts.

Rosette® provides businesses and government agencies with text analytics in 55 languages. www.rosette.com