

Using a Name-Matched Directory to Connect an International Community

ROSETTE

New York University (NYU) is one of the world's largest international universities with over 50,000 students, as well as thousands of faculty members and alumni. NYU has what's known as a city campus where university buildings are spread all over Manhattan and Brooklyn without a central quad. Additionally, NYU has campuses in Abu Dhabi and Shanghai, as well as 11 other study away locations throughout the world.

The challenge

The dispersed campus posed a challenge to community cohesion. Students, faculty, and alumni were less likely to bump into each other in the hallway, happen upon an event, or overhear a conversation.

To address this, NYU decided to invest in an online portal as a place for community members to find and engage with like-minded individuals within the security of the NYU community — a virtual quad of sorts.

To make the university's plan a reality, NYU administration tapped Software Architect Michael Douglass and Product Manager Madan Doraira with building and rolling out this new service. As a first step, the team created a campus-wide directory that would demonstrate immediate value for the

nascent portal. Focusing on the creation of a reliable directory to launch the portal, the NYU team set out to find the best tools for searching and matching names.

With years of experience both developing and consulting, Douglass knew that searching names supported by simple substring and synonym-based matching produces "less than overwhelmingly satisfactory results."

Because of the prevalence of open source search tools, the assumption is that the "search problem" has been solved, however searching names faces unique challenges that don't apply to traditional search. Names are short compared to documents, providing fewer opportunities to identify a positive match.

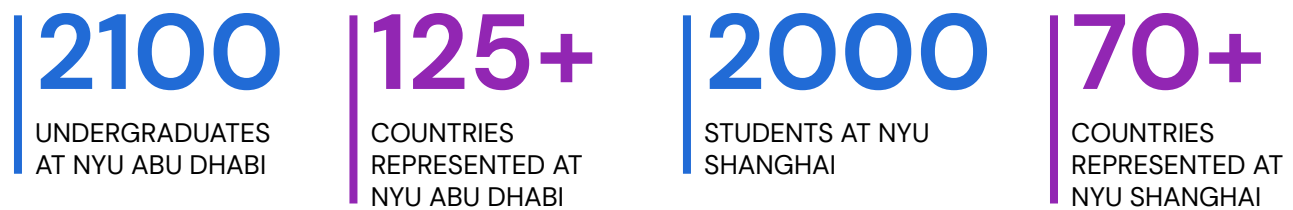


“People don’t even know a challenge in names search exists,” said Dorairaj. “Precision is important, relevancy is important. Users want to see the correct match at the top of the list, and they don’t care how difficult it is to make that happen.”

The challenge of accurately searching names multiplies when international names are added into the mix. Multilingual coverage was particularly vital for NYU, with satellite campuses in Abu Dhabi

and Shanghai, as well as the highest number of international students among U.S. universities.

“Today’s search engines have made people lazy,” said Dorairaj. “They autocorrect so much that people no longer put thought into making their search queries correct. You have to assume that they could be wrong, and hopefully your name matching service will pull the correct name to the top of the list regardless.”



Sources: <https://nyuad.nyu.edu/en/about/nyuad-at-a-glance/fast-facts.html> and <https://shanghai.nyu.edu/about>

The solution

After extensive research, the team found one tool that offered the accuracy, precision, and functionality they needed, Rosette® by Babel Street.

“When I dug into how it works, I saw that Rosette does the things experts say you should do, and doesn’t do what they say you shouldn’t,” said Douglass, “And it’s extremely fast.”

Rosette matches names based on algorithms that take into account the origins and structure of names from different languages and cultures. Its fuzzy matching recognizes differences such as phonetic variation, nicknames, swapped word order, cross-lingual matching, and transliteration variations.

<p>Phonetic similarity Jesus ↔ Heyzeus ↔ Haezoos</p> <p>Transliteration spelling differences Abdul Rasheed ↔ Abd al-Rashid</p> <p>Nicknames William ↔ Will ↔ Bill ↔ Billy</p> <p>Missing spaces or hyphens MaryEllen ↔ Mary Ellen ↔ Mary-Ellen</p> <p>Titles and honorifics Dr. ↔ Mr. ↔ Ph.D.</p> <p>Truncated name components Blankenship ↔ Blankensh</p> <p>Gender Jon Smith ↔ John Smith (but not Joan Smith)</p> <p>Missing name components Phillip Charles Carr ↔ Phillip Carr</p>	<p>Out-of-order name components Diaz, Carlos Alfonzo ↔ Carlos Alfonzo Diaz</p> <p>Initials J. E. Smith ↔ James Earl Smith</p> <p>Name split inconsistently across database fields Rip · Van Winkle ↔ Rip Van · Winkle</p> <p>Same name in multiple languages Mao Zedong ↔ Mao Цзэдуи ↔ 毛泽东 ↔ 毛澤東</p> <p>Semantically similar names PennyLuck Pharmaceuticals, Inc. ↔ PennyLuck Drugs, Co.</p> <p>Semantically similar names across languages Nippon Telegraph and Telephone Corporation ↔ 日本電信電話株式会社</p> <p>Organizational aliases Very Fine Groceries, LLC ↔ VFG</p>
---	--

The NYU directory was already hosted in Elasticsearch. Because Rosette has a pre-built plugin for Elasticsearch, integration was quicker and easier than implementing an in-house solution, or purchasing one from other providers.

NYU also faced the additional roadblock that their registry didn't support extended characters, just ASCII. Rosette catches any resulting missed matches, for example, Rosette correctly returns "Francois" even if a user searches for "François."

"There's a lot of effort that would go into creating a names search tool that really works," said Douglass. "It's likely that we would have spent as many hours working on that alone as everything else put together. Building it ourselves would have taken months, and probably cost just as much or more."

The impact

The NYUHome Portal launched in July, 2017 with Rosette under the hood to perform name matching.

User feedback was positive, with the success of the directory directly contributing to the overall usage and adoption of the portal.

With Rosette name matching, the NYUHome Portal became a thriving international online network, creating a cohesive community for a diverse and widespread population.



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.