NYU

Use Case:
People Directory Search

Segment:
Education

Products:
Name Matching

Availability:
On Premise

Building a virtual quad to connect an international community

Executive summary:

NYU is pioneering a new university intranet that blends university information with a social network, designed to build community among the widespread NYU population. Developers identified directory search as a key feature of the platform to deliver value to users. They quickly discovered that high-performing name search is no trivial task. Nicknames, spelling variations, and translated names are very difficult to match. NYU turned to Rosette name matching and Elasticsearch so that they could focus on building their platform, and get top-of-class directory search in the hands of users as fast as possible.

Key highlights:

- **Best-in-class names search:** By integrating Rosette's fuzzy name matching to their directory, NYU added support for nicknames, multilingual names, and more that their search would not have returned before.
- **Saving time and money:** Accurate names search is deeply complex. Using Rosette's Elasticsearch plugin instead of building names search from scratch saved NYU's developers months of work and the corresponding overhead costs.
- **Creating value for the end user:** With Rosette under-the-hood, the NYUHome Portal is on track to become the new standard for next-generation university intranets.

Revolutionizing the concept of a university intranet

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. Unlike many other universities, NYU has a "city campus," meaning 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 dispersed campus poses a challenge to community cohesion. Students, faculty, and alumni are less likely to bump into each other in the hallway, happen upon an event, or overhear a conversation. Connecting to a university identity factors significantly into how students reflect on their time at school, influencing loyalty to their alma mater and endowment giving.

ROSETTE
TEXT ANALYTICS

Current college students or recent graduates will tell you that their university's intranet leaves much to be desired. Slow and disorganized, they are endured when necessary (enrolling in classes, checking grades, etc.), but not considered a key service. For most universities, investing in an intranet is not a priority; however, NYU saw the portal as an opportunity. The NYUHome Portal will be 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.

One shot to get it right

To make the university's dream a reality, NYU administration tapped Software Architect Michael Douglass, Product Manager Madan Dorairaj, and Product Owner Jim Robertson with building and rolling out this new service.

The NYUHome Portal is a one-stop-shop for all NYU services and community needs. In its pilot release, the portal included over 120 services such as email and internal training programs. However in order for the portal to fulfill its high-level purpose — connecting a disparate NYU community — it would require an unprecedented volume of personal data about community members' interests, hobbies, and history.

> *"We thought to ourselves, if at the very least we can deliver a good directory system with attached profiles, that would provide significant value early on."*

The most obvious source of this data is social media, however students are unlikely to share their personal social media accounts with their school. Instead the team is relying upon users to provide personal data to self-categorize themselves. They're unlikely to do so however if their first experience with the the NYUHome Portal is unsatisfactory.

Creating value through a powerful university directory

Although the ultimate goal for the NYUHome Portal is far bigger than a simple directory, the developers determined that providing something of value early on — like a reliable community directory — would convince community members to become active portal users.

"We have lots of hope to build features on top of what we have so far," said Douglass. "We thought to ourselves, if at the very least we can deliver a good directory system with attached profiles, that would provide significant value early on. Our goal to provide a tool for linking people won't work until lots of people are in the system and adding information about themselves. The first step is to pack the system with data. To do that we need to attract users."

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

The challenge of high-quality names search

With years of experience both developing and consulting, Douglass knew that names search 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 names search faces unique challenges that don't apply to traditional search. Names are short compared to documents for example, providing fewer opportunities to identify a positive match. Furthermore, as with any user-focused app or platform, the goal is to provide as seamless and uncomplicated process as possible: make it look easy.

"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."

Names are highly variable

Unlike other search queries, names have many variations. There is no "right" way to spell a name. Is it Smith or Smythe? Katelyn or Caitlin? Similarly, nicknames may be used interchangeably with a full name. Failure to connect Chuy to Jesus and Dick to Richard will lead to missed search results.

The challenge of accurately searching names multiplies when international names are added into the mix. Often there is not an agreed-upon spelling for names transliterated to English from a foreign script. For example, the Arabic name محمد or Muhammad can also be spelled Mohammed. A high quality names search tool would return all three variations, as well as others.

Multilingual coverage is particularly vital for NYU, with satellite campuses in Abu Dhabi and Shanghai, as well as the highest number of international students among US universities.

1048
STUDENTS AT NYU
ABU DHABI

110+
COUNTRIES
REPRESENTED AT
NYU ABU DHABI

1094
STUDENTS AT NYU
SHANGHAI

69
COUNTRIES
REPRESENTED AT
NYU SHANGHAI

Mitigating the human factor

Living in the age of autocorrect, spelling ability has significantly declined.

"Google has made people lazy," said Dorairaj. "It autocorrects 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."

"Google has made people lazy."

For example, the Irish actress Saoirse Ronan's name is pronounced Sirsha. A user who had only heard the name would almost certainly spell it incorrectly. Even common names are likely to be misspelled if the user has never seen it written. They'll search for John instead of Jon or Caitlin instead of Katelyn.

Pairwise Name Matching

Score: 74.6%

Name 1: **Saoirse Ronan**   Name 2: **Sirsha Ronan**
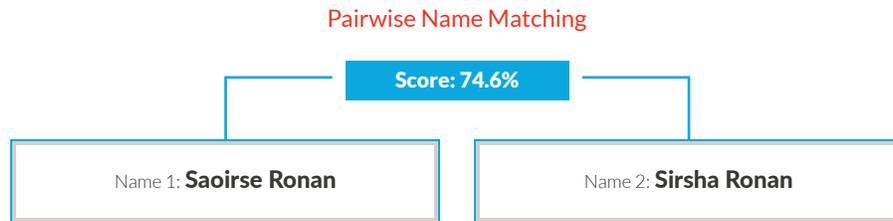
Names search, the way it should be done

After extensive research, the team found one tool that offered the accuracy, precision and functionality they needed, Rosette:

"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 ("Cairns", "Kearns", "Kerns"); nicknames ("William", "Will", "Bill", "Billy"); swapped word order ("Ichiro Suzuki" and "Suzuki Ichiro"); cross-lingual matching and transliteration variations ("Sergei Rachmaninoff," "Sergey Rachmaninov," "Серге́й Васи́льевич Рахма́нинов"), and more.

Phonetic similarity
Jesus ↔ Heyzeus ↔ Haezoos

Transliteration spelling differences
Abdul Rasheed ↔ Abdulrashid

Nicknames
William ↔ Will ↔ Bill ↔ Billy

Missing spaces or hyphens
MaryEllen ↔ Mary Ellen ↔ Mary-Ellen

Titles and honorifics
Dr. ↔ Mr. ↔ Ph.D.

Truncated name components
McDonalds ↔ McD ↔ McDonald

Missing name components
Phillip Charles Carr ↔ Phillip Carr

Same name in multiple languages
Mao Zedong ↔ Мао Цзэдун ↔ 毛泽东 ↔ 毛澤東

Initials
J. E. Smith ↔ James Earl Smith

Names split inconsistently across database fields
Dick • Van Dyke ↔ Dick Van • Dyke

Out-of-order name components
Diaz, Carlos Alfonzo ↔ Carlos Alfonzo Diaz

Semantically similar names
Eagle Pharmaceuticals, Inc. ↔ Eagle Drugs, Co.

Semantically similar names across language
Nippon Telegraph and Telephone Corporation ↔ 日本電信電話株式会社

Checking all the boxes

Before finding Rosette, the NYU directory was already hosted in Elasticsearch. Because Rosette has a pre-built names plugin for Elasticsearch, integration was quicker and easier than a names search built in-house, or purchased from other providers.

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

"Rosette does the things experts say you should do, and doesn't do what they say you shouldn't."

"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."

A modern portal for a modern university

The NYUHome Portal launched on July 1st 2017. The initial pilot of 3,000 was successful enough that the team extended access to 120,000 people, of which 60,000 have searchable profiles. University administration anticipates having as many as 500,000 users in the future once alumni are added.

User feedback has been positive so far, and the developers have already started seeing a number of users beginning to flesh-out their profiles and self-categorize themselves. After the initial pilot there were 1,800 categories that users selected to describe themselves from a dictionary of 90,000 categories and subcategories — and that number is only expected to grow.

With Rosette name matching under-the-hood, the NYUHome Portal is on its way to becoming the first of its kind: a thriving international online network, creating a cohesive community for a diverse and widespread population.

ROSETTE
TEXT ANALYTICS