

Make Sense of Your Multilingual Name Records with Rosette for RapidMiner

Try the Name Translation operator in our Rapidminer plugin and find duplicates easily, even across different languages and scripts

Guess what? We work in a global economy. You probably aren't surprised to hear that; it's quite common for companies, even tiny startups, to have partners and customers all over the world. Internationalization can be a boon for both innovation and business, but it also brings many challenges. If you have multilingual sales reps chasing leads across a wide range of countries you're likely to be confronted with a CRM database that contains a variety of different languages and scripts.

Perhaps you work in marketing and you'd like to prepare a report on all your partners in Asia-Pacific. However, your CRM data contains person, organization, and location names written in Japanese, Korean, *and* Chinese. This is an issue we often run into on the Rosette team; we have offices in Israel and Tokyo and customers all over the world. Thankfully, multilingual text analytics is also our specialty.

Today we'll show how you can use Rosette in tandem with the popular predictive analytics platform RapidMiner Studio to translate multilingual names in CRM data, using examples from our own Salesforce records.

Get the complete picture

Multilingual CRM data poses many challenges. First, it can be difficult to grasp the full scope of a business relationship when multiple languages are involved. Take, for example, the international company Hitachi. Our Salesforce contains records from discussions we've had with Hitachi in the United States are written in English, but our Japanese office tracks their work with the company under 日立. You'd need to unify both accounts to get the global picture.

While some companies might prefer that all CRM data be recorded in English, it can be valuable for sales reps and field engineers to have customer information handy in their native tongue. It's also useful to ensure that a customer's name and address is recorded properly in its original language, saving you from possible errors of reverse translation later.

Our Salesforce also contains a variety of customer interaction records, including emails, support tickets, event registrations, and requests for information. Often this data arrives in English, but not always! In order to ensure that these records are associated with the correct account, we

often need to translate the person, organization, and location names, which will allow us to correctly route the information and address it. For example, the screenshot below shows a new subscriber notification we got from Rosette API that will be automatically registered in our CRM:

北京联会's App created on Rosette API - Dear Jeanne, A nev

Feb 12

北京联会 has subscribed to your service Rosette API - Dea

Feb 12

In the next section we walk you through a simple RapidMiner process that draws data from our Salesforce into RapidMiner and uses Rosette's name translation functionality to identify records that belong to the same account.

Using Rosette Name Translation with RapidMiner

Wait! If you haven't downloaded the Rosette extension on the Rapidminer marketplace and [signed up for your free Rosette API key](#), follow the steps listed in our [Quick Start guide to extracting entities in RapidMiner Studio with Rosette](#).

Here's a sample process we've used to help us deal with multilingual Salesforce data:

The screenshot displays the RapidMiner Studio interface. The main workspace shows a process flow with the following operators: Read Salesforce, Select Attributes, Remove Duplicates, and Translate Names (2). The 'Translate Names' operator is selected, and its parameters are visible on the right: Connection (Rosette), Source Language (Unknown), Target Language (English), and Attribute Selector (LastName). A help window for 'Translate Names' is open, showing the following information:

- Translate Names**
Rosette Text Toolkit
- Tags: Rosette Text Toolkit
- Synopsis**
Translates names from unstructured text in cells in a data table.
- Description**
The Rosette name translation operator can translate the name of an entity (person, place, or location) from English to another language, or to English from another language. When you are converting from English to another language, the operator is optimized to convert the English version of a person's name back to the native language of that name. Output is in the form of an ExampleSet where additionally to the input attributes, it will also contain an attribute named after the selected

Starting with the “Read Salesforce” operator, we pulled data into RapidMiner Studio, selected the specific fields/attributes we wanted to have appear in our results, and ran Rosette Name

Translation. You'll notice that we used two "Translate Names" operators, one for each of the attributes we wanted to translate. In this example, we used "Name" and "AccountName (LocalLang)."

Row No.	InputID	Name	Name (English)	AccountName (LocalLang)	AccountName (English)
1	1	若菜 山口	Wakana Yamaguchi	[Blurred]	[Blurred]
2	2	順子 片山	Junko Katayama	[Blurred]	[Blurred]
3	3	ふじこ 山口	Fujiko Yamaguchi	[Blurred]	[Blurred]
4	4	霞王	Xia Wang	[Blurred]	[Blurred]
5	5	ヨウコ 堀内	Yuko Horiuchi	[Blurred]	[Blurred]
6	6	ひかり 中村	Hikari Nakamura	[Blurred]	[Blurred]
7	7	亜沙美 松村	Asami Matsumura	[Blurred]	[Blurred]
8	8	瀬里奈 松田	Serina Matsuda	[Blurred]	[Blurred]
9	9	ユカ 片平	Jack Katahira	[Blurred]	[Blurred]
10	10	重信 知則	Shigenobu Tomonori	[Blurred]	[Blurred]
11	11	サスケ 松田	Sisk Matsuda	[Blurred]	[Blurred]
12	12	현정 南宮	Johnson Namgung	[Blurred]	[Blurred]
13	13	秀英 张	Sung Chang	[Blurred]	[Blurred]
14	14	勝 山田	Masaru Yamada	[Blurred]	[Blurred]
15	15	ひろし 田中	Hiroshi Tanaka	[Blurred]	[Blurred]
16	16	いずみ 古川	Izumi Furukawa	[Blurred]	[Blurred]
17	17	서연 米	Sean Mi	[Blurred]	[Blurred]
18	18	早人 福田	Hayato Fukuda	[Blurred]	[Blurred]

Caption: These aren't real names from our SalesForce, but the accounts are, so we've blurred them out for confidentiality purposes.

When we run the process, our results show two new attributes "LastName (English)" and "AccountName(English)." The names above are a mix of Japanese, Chinese, and Korean. All have been translated to English and are ready for further processing and analysis.

Next Steps

When working in a complex, international environment, you need to properly route and analyze CRM (or help desk records), for which the data needs to be clean. Thanks to our name translation tool, whether you use it with RapidMiner Studio or any other platform, you can ensure that your data contains one unique record per customer or prospect alongside a standardized translation for all your cross-lingual entries.

For further information on how Name Translation could apply to your use case, don't hesitate to contact us at info@basistech.com.

French version here:

https://docs.google.com/document/d/1JZ0RY7I-NcgCRkhA__ZFD7M7XK6o74gzqyUFGd1wfg0/edit#heading=h.5g5korl925r4