

# Rosette Name Translator

## High-Accuracy Translation of Foreign Names into English

Names of people, places, and organizations are typically the most important information found within a document. Yet, translating names from a foreign language into English is quite difficult. Even the most powerful and expensive “machine translation” systems struggle when confronted with the task of accurately translating large numbers of names. Why is this so hard?

- For each name, which words should be translated *orthographically* (i.e. transliterated) and which words should be translated *semantically*? For the name 北京国家体育场, we prefer to translate the first word orthographically and the last two words semantically (**Beijing National Stadium**) rather than translate the entire name orthographically (**Beijing Guojia Tiyuchang**) or semantically (**Northern Capital National Stadium**).
- For each language, conflicting conventions for translation may be in use. For example, both “Fuji” and “Huzi” are accepted spellings of the name of the well-known Japanese mountain. The debate over whether the capital of Ukraine should be spelled “Kiev” or “Kyiv” has triggered diplomatic crises.
- Common practice or conventional wisdom may conflict with organizational standards. For example, the name of the former ruler of Iraq typically appears in the news media as “Saddam Hussein”. However, the CIA’s official spelling is “Saddam Husayn”.
- A name written in a foreign language may be native to that language, such as محمود احمدينژاد (Mahmoud Ahmadinejad), or may be an English name written in a foreign alphabet, such as جورج دبليو بوش (George W. Bush).

Names must be analyzed by individuals working in a wide range of professions, including finance, law enforcement, intelligence, border control, and litigation support. They must also be analyzed by information systems which support these professionals, and which perform such tasks as information retrieval, media exploitation, digital forensics, and document triage. Yet conventional machine translation (MT) systems are simply not up to the task.



## THE ROSETTE SOLUTION

Rosette® Name Translator (RNT) accepts names in a foreign language and produces accurate, standardized English translations suitable for use in a wide range of applications. Unique capabilities include:

- Generation of “conventional spellings” of frequently appearing foreign names;
- Robust processing of “unrecognized” names, i.e. those not appearing in any known catalog of foreign names;
- Application of complex transliteration standards (such as the U.S. Board on Geographic Names) for translating a name from a foreign alphabet into English; and
- Automatic resolution of name spelling ambiguities in the source language, such as partial vocalization of Arabic, or word segmentation in Chinese.

“We’re really impressed with the Rosette Name Translator’s capabilities and how it has improved our OFAC name checking process. It translated 330,000 Arabic names into Roman script very quickly, consistently, and accurately.”

— Peter Wilkinson, VP of Application Development, NCB Capital

## HOW IT WORKS

RNT provides multilingual name translation via a single API. It combines dictionaries, linguistic algorithms, and statistical inference to derive accurate translations. For example, names written in Chinese are converted from ideographic characters into a phonetic representation. Names written in “unvocalized” Arabic (i.e. without short vowels) are automatically vocalized to enable a phonetic translation according to any of several, user-selected standard systems.

## EXAMPLE TRANSLATIONS

The following examples illustrate the name translator's ability to translate personal and organization from Arabic, Russian, and Japanese, including names of English origin written in another language, back into the correct English.

<b>Person Name</b> - Arabic origin - Forward transliteration	أبو يوسف يعقوب	Abu Yusuf Ya'qub
<b>Place Name</b> - Arabic origin - Forward transliteration	باقة الشرقية	Baqah Al-Sharqiyyah
<b>Person Name</b> - English origin - Reverse transliteration	جون كينيدي	John Kennedy
<b>Person Name</b> - English origin - Reverse transliteration	الرئيس كلينتون	President Clinton
<b>Organization Initialism</b> - English origin - Reverse transliteration	بي بي سي	B.B.C.
<b>Person Name</b> - English origin - Reverse transliteration	Уильям Джефферсон Клинтон	William Jefferson Clinton
<b>Person Name</b> - Russian origin - Forward transliteration	Василий Иванович Баженов	Vasiliy Ivanovich Bazhenov
<b>Person Name</b> - English origin - Reverse transliteration	ビル・クリントン	Bill Clinton
<b>Person Title</b> - Japanese origin - Forward transliteration	鈴木一郎	Suzuki Ichirou

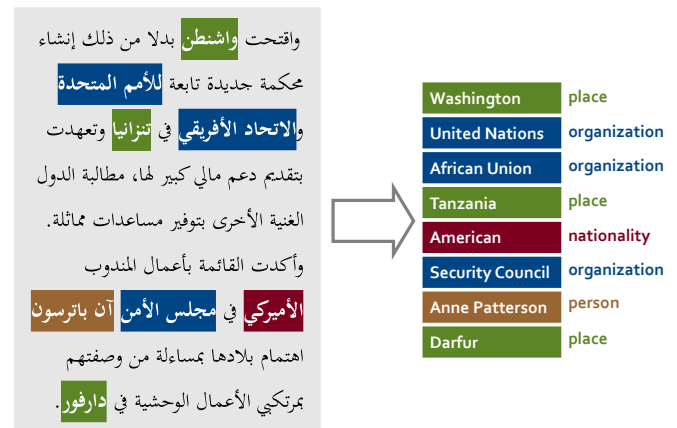
## AVAILABLE LANGUAGES AND SCRIPTS

RNT supports the following languages and writing systems, as well as those languages written using the Latin alphabet. Additional languages are available via custom development.

<b>Arabic</b>	Arabic script	العربية
<b>Chinese</b>	Simplified Hanzi, Traditional Hanzi	简体字 繁體字
<b>Dari/Farsi (Persian)</b>	Arabic script	دری فارسی
<b>Japanese</b>	Hiragana, Katakana, Kanji	ひらがな カタカナ 漢字
<b>Korean</b>	Hangul, Hanja	한글 韓文漢字
<b>Pashto</b>	Arabic script	پښتو
<b>Russian</b>	Cyrillic script	Путин
<b>Urdu</b>	Arabic script	اردو

## COMBINING EXTRACTION WITH TRANSLATION

RNT may be paired with the Rosette Entity Extractor (REX) to extract and translate key names present in a document, with accuracy superior to either statistical or rule-based MT systems. This approach may also be used to enrich or remediate the output of such systems in situations where translations of entire paragraphs or documents are required.



## SYSTEM PLATFORMS SUPPORTED

Software development kits (SDKs) and web services are available for the platforms listed below. Contact your sales representative for additional platform support.

AIX 6.1, PPC HP-UX 11i, IA64 Linux CentOS 4.x/5.x, IA32/AMD64 Linux Debian 5.x, IA32/AMD64 Linux Red Hat 4.x/5.x, IA32/AMD64	Linux Ubuntu 10.x/11.x, IA32/AMD64 MacOS Solaris 10, SPARC32/64, IA32/AMD64 Windows XP/Vista/7, IA32/AMD64 Windows Server 2003, 2008
--	--

## EXPLORE FURTHER

For more information or to request an evaluation copy, please call us at 617-386-2090 or 800-697-2062, or write to [info2011@basistech.com](mailto:info2011@basistech.com). We will be happy to assist you in evaluating the performance of our product on your data.

VISIT [www.basistech.com](http://www.basistech.com) WRITE [info2011@basistech.com](mailto:info2011@basistech.com) CALL 617-386-2090

One Alewife Center  
Cambridge, MA 02140

171 Second Street  
San Francisco, CA 94105

2553 Dulles View Drive  
Herndon, VA 20171

9-6 Nibancho, Chiyoda-ku  
Tokyo 102-0084

