

# 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 volcano. 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 احمدى نجاد (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.



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.

### HOW IT WORKS

RNT provides multilingual name translation via a single API. When presented with a name to translate, RNT first checks its dictionaries. If the name is not found in any of the dictionaries, it is analyzed according to language-specific algorithms. 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 RNT's ability to translate names between Arabic and English depending upon the name's language of origin.

<b>Person Name</b> - Arabic origin - Forward transliteration	أبو يوسف يعقوب	Abu Yusif Ya'qub
<b>Person Name</b> - English origin - Reverse transliteration	جون كينيدي	John Kennedy
<b>Place Name</b> - Arabic origin - Forward transliteration	باقة الشرقية	Baqah Al-Sharqiyyah
<b>Place Name</b> - English origin - Reverse transliteration	ديزني لاند	Disneyland
<b>Place Name</b> - English origin - Reverse transliteration	كاليفورنيا or كاليفورنيا	California
<b>Organization Name</b> - English origin - Reverse translation	البيت الأبيض	The White House
<b>Organization Initialism</b> - English origin - Reverse transliteration	بي بي سي	BBC

## AVAILABLE LANGUAGES

RNT supports the following languages. Additional languages are available via custom development.

Arabic	Dari/Farsi (Persian)	Russian
Chinese (Simplified)	Japanese	Pushto
Chinese (Traditional)	Korean	Urdu

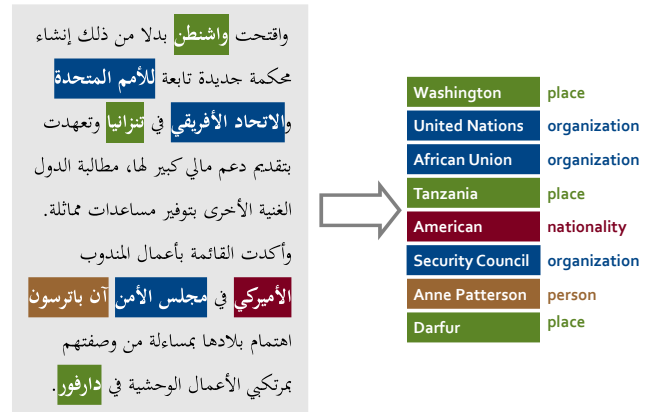
## AVAILABLE WRITING SYSTEMS

RNT supports the following writing systems. Additional writing systems are available via custom development.

العربية	Arabic, Dari, Farsi, Pushto, Urdu
漢字	Chinese Hanzi, Japanese Kanji, and Korean Hanja
かな	Japanese Hiragana
カナ	Japanese Katakana
한글	Korean Hangul

## 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) are available for popular architectures, operating systems, and development environments. Support for platforms not listed below is available by contacting your sales representative.

AIX 5.3/6.1, PPC  
 HP-UX 11i, PA-RISC/IA64  
 Linux CentOS 4.x/5.x, IA32/AMD64 (GCC 4.1/4.2)  
 Linux Debian 5.0, IA32/AMD64  
 Linux Red Hat 3.0, IA32/AMD64 (GCC 3.2/3.4/4.0)  
 Linux Red Hat 4.0, IA32/AMD64 (GCC 3.4/4.0)  
 Linux Red Hat 5.0, IA32/AMD64 (GCC 4.1/4.2)  
 Linux Ubuntu 8.04/9.x/10.x, IA32/AMD64  
 MacOS (GCC 4.0)  
 Solaris 9, SPARC32/64  
 Solaris 10, SPARC32/64  
 Solaris 10, IA32/AMD64  
 Windows XP/Vista/7, IA32 (MSVC 7.1)  
 Windows XP/Vista/7, IA32/AMD64 (MSVC 8.0)

## FOR MORE INFORMATION

For more information, please visit [www.basistech.com](http://www.basistech.com). To request an evaluation copy, please write to [info2010@basistech.com](mailto:info2010@basistech.com) or call us at 617-386-2090 or 800-697-2062.



One Alewife Center  
Cambridge, MA 02140

171 Second Street  
San Francisco, CA 94105

13800 Coppermine Road  
Herndon, VA 20171

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