

Rosette Name Indexer

Precise Search for Names in Multiple Languages and Writing Systems

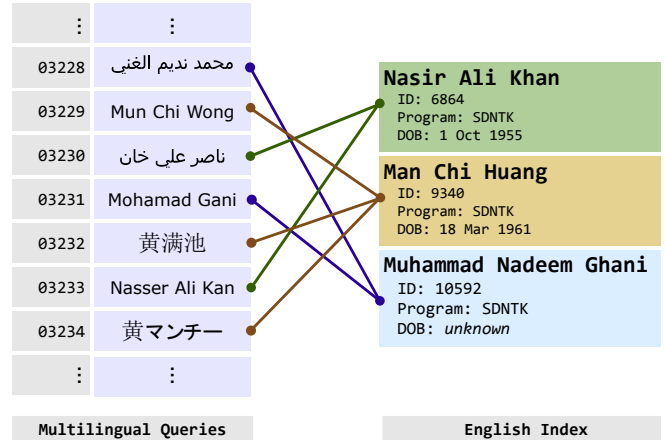
Rosette® Name Indexer (RNI) matches names of people, places, and organizations written in different languages against a single, universal index. For example, looking for “Mao Zedong” returns matches written in English (such as “Mao Ze Dong” or “Mao Tse Tung”); in Simplified Chinese (毛泽东); in Traditional Chinese (毛澤東); or even in other languages such as Arabic (ماو تسي تونغ). This capability is designed to be integrated into such applications as watch list management, fraud detection, money laundering, geospatial analysis, and document triage.

Unlike legacy solutions driven by lists containing billions of spelling variants, RNI analyzes the intrinsic structure of each name component, and performs an intelligent comparison using advanced linguistic algorithms. This approach is *not limited to a particular list of variants* and reduces the likelihood of both “false positives” (large numbers of wrong hits) and “false negatives” (zero hits, or a failure to uncover relevant matches).

NAME MATCHING CAPABILITIES

RNI handles the following name-matching problems:

Same name written in multiple languages	“Mao Zedong”, “Мао Цзэдун”, “毛泽东”, “毛澤東”
Phonetic spelling differences	“Cairns”, “Kearns”, “Kerns”
Transliteration spelling differences	“Abdul Rasheed”, “Abd-al-Rasheed”, “Abd Ar-Rashid”
Nicknames	“William”, “Will”, “Bill”, “Billy”
Initials	“J. E. Smith”, “James Earl Smith”
Titles and honorifics	“Dr.”, “Mr.”, “Esq.”, etc.
Out-of-order name components	“Diaz, Carlos Alfonso”, “Carlos Alfonso Diaz”
Missing name components	“Phillip Charles Carr”, “Phillip Carr”
Missing spaces (variable segmentation)	“MaryEllen”, “Mary Ellen”, “Mary-Ellen”
Truncated name components	“Royal Bank of Sco”, “Mcdonal”, “Stev”



EXAMPLES

RNI will score “Jesus Alfonso LOPEZ DIAZ” against each of these variations with a progressively lower confidence:

Misspelled family name	Jesus Alfonso LOBEZ DEAZ
Mother’s father’s name removed	Jesus Alfonso DEAZ
Middle name replaced with an initial	Jesus A. DEAZ
Given name replaced with a nickname	Chuy A. DEAZ
Re-ordered name components	DEAZ, Chuy A.

Names written in other languages present unique problems. For example, the complex structure of Arabic names often leads to missing name components. Consider the following name:

Title	Given Name	Patronymic	Family Name
Al-Sheikh	Abdullah	Bin Hassan	Al-Ashqar
الشيخ	عبد الله	بن حسن	الأشقر

This name may appear in various forms, including:

- Al-Sheikh Abdullah Al-Ashqar (no patronymic)
- Abdullah Al-Ashqar (no title, no patronymic)
- Al-Sheikh Abdullah Bin Hassan Bin Mohammed Al-Ashqar (with grandfather’s patronymic)

BENEFITS

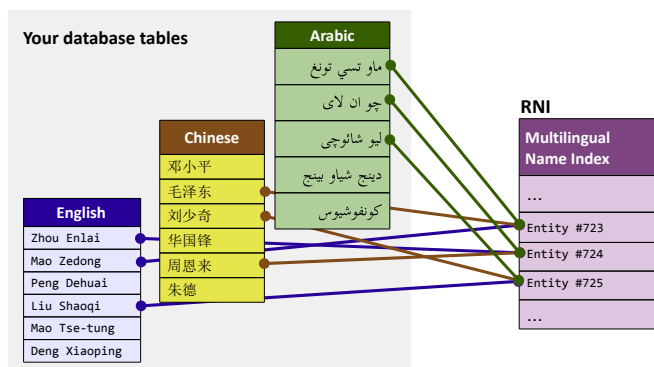
- Matches names of people, places, and organizations
- Increases name search accuracy and finds hits other systems miss
- Ranks results by relevancy with a similarity score
- Returns partial matches where data is incomplete
- Searches for names from many cultures, including Middle Eastern and East Asian, regardless of how they are written

HOW IT WORKS

RNI provides a repository for names—called a “name index”—with a fuzzy search capability. RNI may operate as a standalone name index or it may be integrated with an existing name database. It leaves your data untouched, providing pointers to the matching names within your database. When not all components of a name match, RNI aligns input names with entries to recognize partial matches. With each update of the database, RNI’s index can also be automatically updated.

When running a name search, the matching names are returned to the calling application with a confidence score from 0% to 100%. A minimum match threshold may be set to control the quality and quantity of results returned. The RNI programming interface may be used to access other information associated with each entry—such as relationships, geographic locations, and pointers to external databases—to help identify specific individuals and places.

Often an RNI name index is built using the output of the Rosette Entity Extractor (REX) to enable name matching across the names mentioned in a large collection of documents.



RNI’s multilingual name index points to matching records in your database to enable name searching across multiple languages.

CUSTOMIZABLE TO YOUR REQUIREMENTS

RNI can be customized by the user, to improve accuracy and relevancy for a particular dataset or use case:

- Ignore a given list of words (“stopwords”) with respect to matching. (e.g., titles, honorifics)
- Force two name components to always match with a given score. (e.g., “Elizabeth” and “Lisbeth” always match at 90%)
- Force two names to always match with a given score. (e.g., “John Doe” and “Joe Bloggs” always match at 95%)
- Link multiple names to a single individual. (e.g., queries for “Marilyn Monroe” and “Norma Jeane Mortensen” include the same person)

AVAILABLE LANGUAGES AND SCRIPTS

RNI matches names within the following languages, and between those languages and English. Additional languages are available via custom development.

Language	Supported Scripts	Script Samples
Arabic	Arabic script	العربية
Chinese	Simplified Hanzi, Traditional Hanzi	简体字 繁體字
English	Latin script	English
Japanese	Hiragana, Katakana	ひらがな カタカナ
Korean	Hangul, Hanja	한글 韓文漢字
Persian (Farsi)	Arabic script	دری فارسی
Pushto (Pashto)	Arabic script	پښتو
Urdu	Arabic script	اردو

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

VISIT www.basistech.com WRITE 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

