

Rosette for e-Discovery

Expand Your e-Discovery Scope to Languages Beyond English

In the age of globalization, documents created in different countries and different languages can be highly relevant to legal investigations. Litigation often spans national boundaries, and legal teams are flooded with thousands of documents in languages other than English that need to be filtered, evaluated, and analyzed.

The growing importance of multi-language discovery creates new challenges for attorneys and their technology partners. E-Discovery is already complex, and that complexity grows by orders of magnitude when it involves documents in different languages, writing systems, and character sets. Yet the need for meticulous investigation is as crucial as ever for lawyers to provide the best outcomes for their clients.

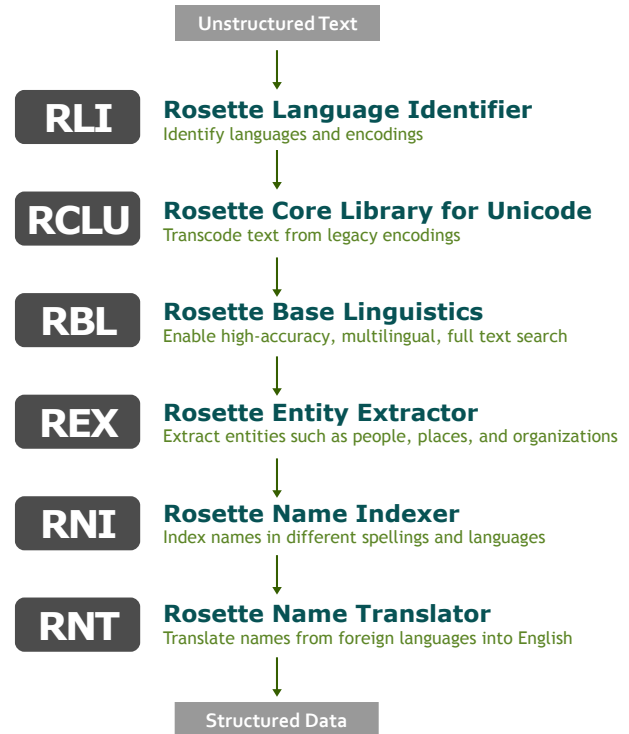
“Text analytics is no longer an academic specialty. It has become a necessary component in most search and discovery software — from selling products, tracking terrorists, delivering news, or playing music — to improving communication among people worldwide. Basis Technology’s new Rosette 7 platform ups the ante with its improvements in accuracy, enabling its customers to power a new breed of intelligent workspace applications.”

— Susan Feldman, Research Vice President, IDC

ADVANCING LEGAL TECHNOLOGY

Basis Technology helps the legal community meet its multilingual discovery challenges head-on. We provide comprehensive electronic discovery solutions that uncover evidence buried in terabytes of unstructured multilingual text — accurately, quickly, and cost-effectively. We do it using the most advanced linguistics software in the industry—software found at the core of virtually all leading multilingual search engines and information retrieval applications.

Our multilingual e-Discovery solutions are based on Rosette®, a linguistics platform proven in hundreds of commercial and government environments. Interoperable Rosette software components are configured as building blocks for multilingual e-Discovery, working seamlessly within discovery workflows and information retrieval applications while handling many different languages, character sets, and data sources.



ACHIEVE TRUE MULTILINGUAL E-DISCOVERY

Our industry-leading linguistics software is complemented by ease of integration within data mining, reviewing, search, and other discovery applications used by legal teams. By plugging the Rosette API into an application, users get instant access to unique e-Discovery tools covering major European, Asian, and Middle Eastern languages. For legal professionals, it means the ability to examine multilingual text with unparalleled accuracy and efficiency.

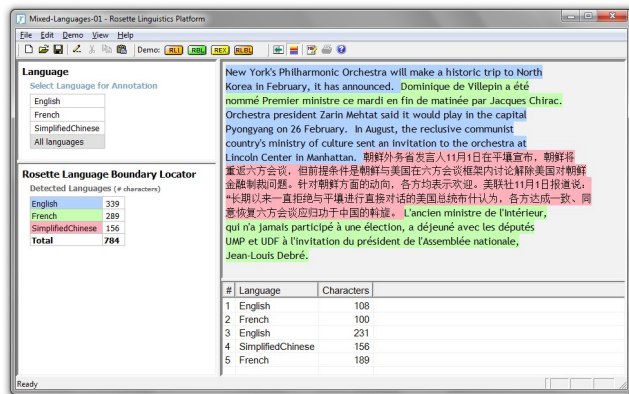
BENEFITS

- Finds evidence buried in terabytes of unstructured multilingual text
- Discovers without language barriers
- Produces all relevant multilingual documents
- Supports over 55 languages, including Asian, European, and Middle Eastern

THREE STEPS TO MULTILINGUAL E-DISCOVERY

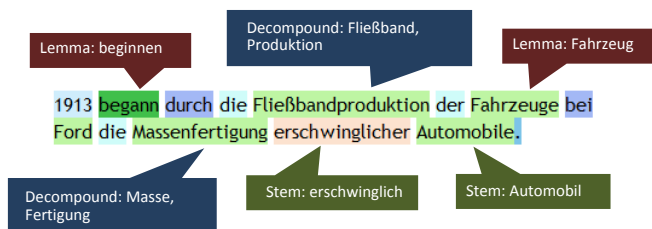
Step 1: Identify the language(s) and encoding in a document, and convert to Unicode

The *Rosette Language Identifier (RLI)* identifies the language(s) and character encoding systems present in a document so its textual content can be filtered and processed. Extracted text is converted to Unicode so that discovery and information retrieval applications can access a single data representation regardless of language. Using a module called the Language Boundary Locator, mixed-language documents may be segmented into regions so that language-specific processing may be performed on each region. RLI identifies 55 languages with high accuracy, even when presented with short strings of text.



Step 2: Apply linguistic intelligence to identify word forms, parts of speech, and sentence structure

Rosette Base Linguistics (RBL) examines documents and performs a complete morphological analysis so text can be accurately filtered, analyzed, and searched. RBL identifies parts of speech, sentence boundaries, word breaks, tokens, and other linguistic components within a document, in European, Asian, and Middle Eastern languages. The technology and linguistic data in RBL results from over ten years of development and use in web and enterprise search engines.



Step 3: Extract the items of interest (including those you didn't know about)

The *Rosette Entity Extractor (REX)* sifts through unstructured text and identifies people, places, dates, and other items that establish the true meaning of a document for further analysis. REX locates generic terms as well as custom entities such as specific names, phone numbers, and email addresses. Statistical modeling helps determine if an entity resides within a document, rather than simply referring to a list of possibilities and risk overlooking a variation. The result is entity extraction technology that lets you find what you know—and also what you didn't know.

Kofi Atta Annan GCMG is a Ghanaian diplomat who served as the seventh Secretary General of the United Nations from January 1, 1997 to January 1, 2007, serving two five-year terms. Annan was the co-recipient of the Nobel Peace Prize in October 2001.

Kofi Annan was born in April 8, 1938 to Victoria and Henry Reginald Annan in Kumasi, Ghana. He is a twin, an occurrence that is regarded as special in Ghanaian culture. Efua Atta, his twin sister, shares the same middle name, which means 'twin'. As with most Akan names, his first name indicates the day of the week he was born: Kofi denotes a boy born on a Friday. The name Annan can indicate that a child was the fourth in the family, but in his family it was simply a name which Annan inherited from his parents.

In 1962, Annan started working as a Budget Officer for the World Health Organization, an agency of the United Nations. From 1974 to 1976, he worked as the Director of Tourism in Ghana. Annan then returned to work for the United Nations as an Assistant Secretary General in three consecutive positions.

SYSTEM SPECIFICATIONS

Rosette is a portable and highly scalable software developer kit (SDK) that runs on platforms ranging from laptop PCs to multi-CPU servers processing thousands of documents per second.

A fully-documented API is provided and may be accessed from applications written in C, C++, Java, and other languages. A command-line interface is also available for testing purposes.

SDKs are available for Apple MacOS, HP-UX, IBM AIX, Microsoft Windows, Sun Solaris, and multiple Linux distributions.

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. We will be happy to assist you in evaluating the performance of our product on your data.

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

