

# Rosette Entity Extractor

## Locate and Extract Names, Places, Organizations, and Other Entities from Unstructured Text

Rosette® Entity Extractor (REX) enables search and text mining applications to automatically locate concepts, which cannot easily be found using only lists or rules, in unstructured text. REX uses advanced linguistics to find *entities* — names, places, dates, identifiers, and other unique words or phrases — that establish the real meaning in text.

Entity extraction enables the classification, management, and analysis of large volumes of unstructured text coming from sources such as e-mail, electronic documents, and the Internet. REX can be used both to locate concepts in unstructured data and as a step in preparing such data for structured analysis, a process known as “metadata annotation”.

REX locates generic terms such as “Vice President” and “Southeast Asia” as well as specific references such as “President Barack Obama” and “January 20, 2009.” These features enable the identification and extraction of important information from documents for further analysis by other applications.

REX employs three types of entity detection technology — statistical modeling, regular expressions, and lists — to achieve industry-leading accuracy.

The REX Application Programming Interface (API) is an extension of the proven Rosette API, designed to quickly and efficiently integrate into a wide variety of applications. Unlike competing extractors, REX integrates seamlessly with the Rosette Name Indexer and Rosette Name Translator, enabling fully-automatic, cross-script resolution and translation of extracted foreign names.

### APPLICATIONS

REX brings a deep level of linguistic analysis to software systems designed for information retrieval, CRM, computer forensics, intelligence analysis, competitive intelligence, social media monitoring, and countless other applications. Its functions are integral to taxonomy and categorization applications; text mining; trend analysis; and relationship extraction engines.

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.

REX is trained and tuned for each specific language, and is customizable to meet the particular entity needs of vertical applications such as health, medical, life sciences, financial, and manufacturing.

### MIXED-LANGUAGE DOCUMENTS

REX is capable of analyzing documents which may be written in more than one language. By utilizing a technique known as “language boundary location”, REX is able to apply the correct language-specific linguistics to each distinct language region of a mixed-language document, thereby achieving greater accuracy on “real world” data.

### SYSTEM SPECIFICATIONS

REX is portable and highly scalable, running 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.

### 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.

## PREDEFINED ENTITY TYPES

REX natively supports the following entity types, as well as fundamental linguistic elements, such as sentence boundaries, parts of speech, and noun phrases. User-defined entity types (such as SKU numbers) are also supported.

Person	Money
Location / GPE	Generic Number
Organization	Personal ID Number / SSN
Title	Phone Number
Nationality	Email Address / URL
Religion	Distance
Credit Card Number	Date
Latitude / Longitude / UTM	Time

## AVAILABLE LANGUAGES

REX natively supports the following languages. Additional languages are available via custom training.

Arabic	French	Persian (Farsi/Dari)
Chinese (Simplified)	German	Pushto
Chinese (Traditional)	Italian	Russian
Dutch	Japanese	Spanish
English	Korean	Urdu

## 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

(서울=연합뉴스) 이승관 김경희 기자 = 청와대가 최근 한나라당 시,도당 여성 당직자를 초청, 간담회를 갖는 과정에서 복당 의원들을 일괄 배제해 "복당파는 뒷전"이라는 불만이 제기되고 있다.

한나라당 한 관계자는 27일 연합뉴스와의 통화에서 "지난달부터 총정, 호남권 등을 시작으로 대통령 부인 김윤옥 여사가 각 시,도당별 여성 당직자들을 청와대에 초청해 다과회를 갖고 있다"면서 "당원협의회별로 여성 당직자들을 선발하다보니, 복당 의원 지역의 경우는 총선에서 한나라당 공천을 받은 당협위원장들이 추천권을 갖게 됐다"고 설명했다.

문제는 지난달 9일 총정권을 시작으로 두 달 가까이 계속돼 온 행사 자체가 복당 의원들 예겐 양해는 물론 통보조차 되지 않았다는 것.

게다가 당협위원장이 대부분 남성이어서 그 부인들이 여성 당직자들을 인솔해 청와대를 방문해야 하고, 그러다 보니 같은 한나라당 국회의원 부인 가운데에도 복당 의원들의 부인만 행사에서 소외되는 상황까지 발생하고 있다.

이제까지는 총정과 호남, 강원, 인천, 제주 등 복당 인사들이 그다지 많지 않은 지역을 중심으로 행사가 진행됐기 때문에 특별한 잡음이 발생하지 않았지만, 복당 인사들이 상당수를 차지하는 부산과 대구 지역 간담회가 오는 29일로 다가오면서 이와 관련한 불만과 문제제기가 여기 저기서 터지고 있다.

### Korean

صربيا تسحب سفيرها من واشنطن بعد اعترافها باستقلال كوسوفو  
GMT+04:00) - 18/02/08) 2358

(CNN)-- في رد فعل فوري على إعلان وزارة الخارجية الأمريكية اعتراف الولايات المتحدة بكوسوفو كدولة مستقلة عن صربيا، وعزمها إقامة علاقات دبلوماسية معها، قررت حكومة بلغراد سحب سفيرها من واشنطن.

جاءت هذه الخطوة، التي أعلنها وزير الخارجية الصربي، فوك جيرميتش الاثنين، بعد قليل من إعلان وزيرة الخارجية الأمريكية، كوندوليزا رايس، في بيان مكتوب، أن الولايات المتحدة تعترف بدولة كوسوفو، كأحد دولة مستقلة في أوروبا.

وجاء في البيان الذي صدر بواشنطن: "نتقدم بالتهنئة إلى شعب كوسوفو بهذه المناسبة التاريخية"، مضيفة قولها إن "الرئيس بوش قرر الاستجابة على الفور، لطلب من حكومة كوسوفو، بشأن إقامة

### Arabic

朝鮮外務省發言人11月1日在平壤宣布，朝鮮將重返六方會談，但前提條件是朝鮮與美國在六方會談框架內討論解除美國對朝鮮金融制裁問題。針對朝鮮方面的動向，各方均表示歡迎。美聯社11月1日報道說：“長期以來一直拒絕與平壤進行直接對話的美國總統布希認為，各方達成一致、同意恢復六方會談應歸功於中國的斡旋。

朝鮮中央通訊社11月1日報道，朝鮮外務省發言人當天在接受該社記者採訪時，對10月31日北京舉行的雙邊及多邊會晤問題作出回答。朝鮮最近對美國的金融制裁採取了自衛性措施。10月31日在北京舉行了雙邊及多邊會晤。會晤討論了尋求恢復六方會談途徑的事宜。朝鮮決定重返六方會談，前提是朝鮮與美國將在六方會談框架內就解除金融制裁問題展開討論並作出安排。在朝鮮宣布重返六方會談前，中國外交部10月31日發布消息說，根據中方建議，六方會談團長在北京舉行非正式會晤。三方就繼續推進六方會談坦率地交換了意見。三方一致同意在六方會談的近期舉行六方會談。美國總統布希10月31日對新聞界說，他對中國、美國和朝鮮三方同意於近期恢復六方會談表示歡迎，對朝鮮核問題取得進展感到非常高興。此外，參加六方會談另外三方韓國、俄羅斯、日本以及聯合國秘書長科菲·安南都對此作出積極回應。

### Chinese

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

