

Rosette 言語処理プラットフォーム

多言語検索アプリケーションの最先端

テキスト解析プラットフォーム

Rosette® 言語処理プラットフォームは、世界で幅広く使用されている多言語テキストの検索および解析用コンポーネント・ライブラリーです。非構造化テキストの、自動言語判別、テキストの正規化および固有表現（キーワード）の抽出機能を単一の統合フレームワークで提供します。

「テキスト解析は、今や学術的な研究分野を超え、検索やディスカバリー用ソフトウェアの必須条件になりました。製品販売、テロリスト追跡、ニュース配信、音楽再生といった用途から、世界中のユーザー間のコミュニケーションの向上に至るまで、あらゆるところで使われています。精度が向上した Basis Technology の新バージョンの Rosette 7 プラットフォームは業界水準をさらに高め、新しいタイプの情報解析アプリケーションとして期待できます。」

— IDC 社リサーチ部門長 Susan Feldman 氏

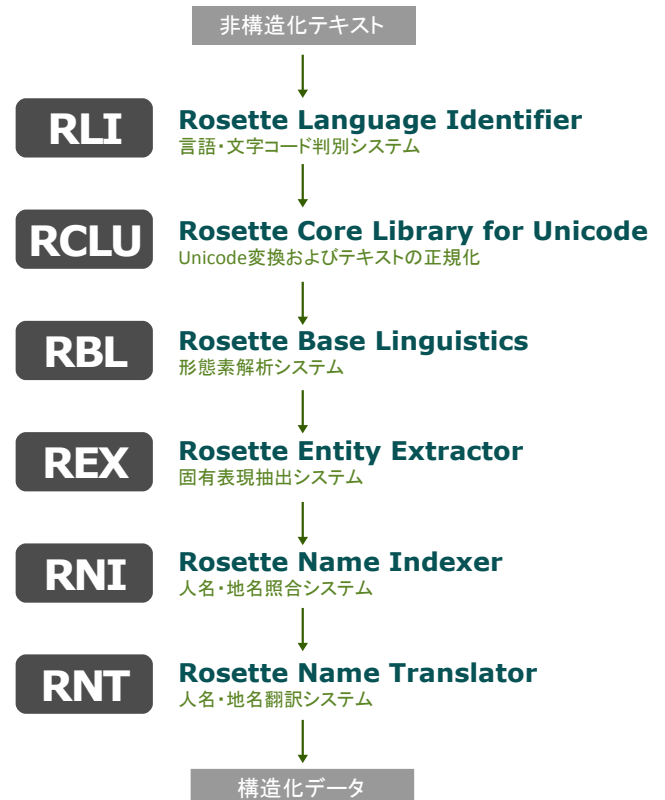
「Google は、日本語・中国語・韓国語版の最良サーチエンジンを作るべく、アジア言語処理技術を提供する Basis Technology を選びました。これは、世界中のインターネット利用者にもっとも好まれるサーチエンジンとしての Google の地位を築くための重要な一石を投ずることとなりました。」

— Google シニア・ヴァイス・プレジデント, Urs Hölzle 氏

柔軟な構成

アプリケーションで多言語の生データを解析、知的に処理・運用し、各ユーザーの要件に合った、コンポーネントを組み合わせて、柔軟なソリューションを構成することができます。新しい言語、文字セット、およびデータ・ソースを追加利用できるようにしながら、既存のワークフロー内でシームレスな処理を可能にします。Rosette は、大量の非構造化テキストから意味のあるインテリジェンスを正確、迅速、かつコスト効率の高い方法で抽出する必要がありますすべてのアプリケーションを強化します。

- 文書内に存在する単一または複数の言語を自動的に判別し、テキストを解析、フィルタリング、および検索
- 多言語の全文検索の適合率および再現率を向上
- 非構造化テキストから人名、地名、日付、ID番号などを抽出



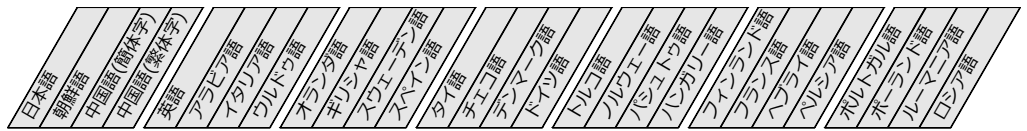
Rosette コンポーネントは、C、C++、Java、.NET をサポートする単一の API によってアプリケーションにプラグインされます。アプリケーションおよびワークフロー内で必要なモジュールを使用し、適宜新しい機能または言語を簡単に追加することができます。また、オープン・ソース検索エンジン Apache Lucene/Solr と使用することで、商用レベルの高度な言語処理機能を利用できるようになります。これにより多数の言語を索引付けし、検索できます。

主要顧客 (民間企業)

NTT レゾナント	Cisco	HP
富士通	Clearwell Systems	Mark Logic
日立製作所	EMC	Microsoft
日本電気 (NEC)	Endeca	Oracle
Attivio	Exalead/	Symantec
Autodesk	Dassault Sys	Yahoo!
	Google	

主要顧客 (政府関連機関)

Berico Technologies	Northrop Grumman
CACI International	SAIC
In-Q-Tel	米国国防総省
Lockheed Martin	米国司法省
MITRE	米国情報機関



Rosette 形態素解析システム

文の単語への分割	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y				
品詞の判別	Y	Y	Y	Y	Y	Y	Y		Y	Y	Y		Y	Y	Y		Y	Y	Y		Y	Y	Y		Y	Y	Y		Y	Y	Y		Y	Y	Y		Y	Y	Y	
文境界の検出	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y				
名詞句の抽出	Y		Y	Y	Y	Y	Y		Y		Y				Y				Y				Y				Y				Y				Y					
辞書形 (基本形) の派生	Y	Y	n/a	n/a	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y				
複合語の分割	Y	Y	Y	Y	n/a	n/a	n/a	n/a	Y	n/a	Y	n/a	n/a	Y	Y		Y	P	Y		n/a		n/a		n/a	n/a			n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a				

その他の対応言語
 アルバニア語, インドネシア語, ウクライナ語, エストニア語, カタロニア語, クロアチア語, スロバキア語, スロベニア語, セルビア語, ブルガリア語, マレー語, ラトビア語

Rosette 固有表現抽出システム

人名	S	S	S	S	S	S	S	S	S			S				S			S		S	P	S		S			S	S			S
場所	S	S	S	S	H	H	S	S	S			S				S			S		S	P	S		S			S	S			S
組織名	H	H	H	H	H	H	H	H	H	G		H		G		H			H	G	H	P	H		H	G		H	H	G		H
タイトル	S	S	S	S	S	S	S	G	G			S				S			S		S	G			S			S	S			S
国籍	G	G	G	G	G	G		S															G									G
宗教			G	G	G	G		S																								
クレジットカード番号	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
距離	R	R	R	R	R	R	R		R			R				R				R				R	R			R	R			R
緯度・経度 / UTM 図法	R	R	R	R	R	R	R	R	R	R	P	R		R	P	R		P	R	R		R	R		R	R	P	R	R	R	P	R
通貨	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
個人識別番号 / SSN	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
電話番号	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
Eメール・アドレス / URL	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
日付	R	R	R	R	R	R	R		R	R		R				R				R	R			R	R			R	R			R
時間	R	R	R	R	R	R	R		R			R				R				R	R			R	R			R	R			R

Rosette 人名・地名照合システム

ネイティブ文字体系の英語との照合	Y	Y	Y	Y	Y	Y		Y								Y				Y				Y				Y				Y
綴り字の差異の照合	Y	Y	Y	Y	Y	Y		Y								Y				Y				Y				Y				Y
欠落部分の照合	Y	Y	Y	Y	Y	Y		Y								Y				Y				Y				Y				Y
欠落スペースの照合	Y	P	P	P	Y	P																										Y

Rosette 人名・地名翻訳システム

ネイティブ言語名 → 英語	Y	Y	Y	Y	Y	Y		P								P				Y				Y				Y				Y
外国語名 → 英語	Y				Y	Y																										Y

- | | | |
|---------------|----------------------------|-------------------|
| 形態素解析 | 固有表現抽出 | 人名・地名照合／翻訳 |
| Y 完全サポート機能 | S 統計的モデル | Y 完全サポート機能 |
| n/a 当該言語では適用外 | G ガゼティア (各種用語辞書) 項目 (完全一致) | P 部分サポート機能 |
| | H ハイブリッド (統計およびガゼティア) | n/a 当該言語では適用外 |
| | R 正規表現 (パターン一致) | |
| | P 正規表現 (部分サポート) | |

詳細 www.basistech.jp お問い合わせ info@basistech.jp 電話 03-3511-2947

〒102-0084
東京都千代田区二番町9-6

One Alewife Center
Cambridge, MA 02140

2553 Dulles View Drive
Herndon, VA 20171

171 Second Street
San Francisco, CA 94105

