

# Rosette 言語・文字コード判別システム

## 言語と文字コードを正確に判別し、 文字化けを防止するツール

Internet の世界的普及につれ、世界各国から発信されるWebページがますます増加し、その情報収集や利用が進んでいますが、その一方で、それら各国語Webページの情報が文字化けして困ることがあります。それはそのWebページのデータの文字コード(\*注1)が、ユーザーの使用するブラウザで正しく認識されないときに生じます。

Rosette® 言語・文字コード判別システム (RLI: Rosette Language Identifier) は、言語と文字コードをすばやく、かつ正確に判別するので、文字コードを適切に処理して文字化けを防ぎます。また言語ごとにデータを仕分けできるため、その後のアプリケーションでの処理において、言語別の処理が効率的におこなえます。

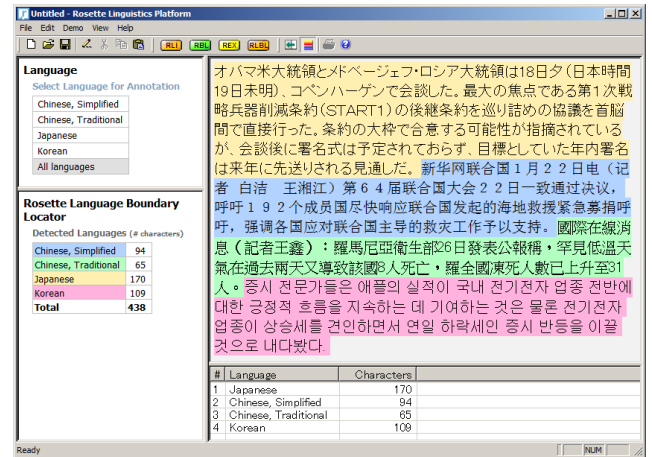
### 対象アプリケーション

- CRM のようなキーアプリケーション用の不特定なテキスト処理
- データウェアハウジング
- 以下のようなデータを通常使用する情報検索: Eメール、ホームページ、従来のシステムデータ、外部データ入力

RLI は、テキストの文字コードと言語を正確に判別します。様々なデータソースからテキストを操作、検索するあらゆるアプリケーションでテキスト操作を行います。

### 仕様

- APIでの提供 (ドキュメント完備)
- C, C++, Java対応
- コマンドラインインターフェースもご要望に応じて提供可能



### 特長

アプリケーションを問わず、あらゆる未知の言語や文字コードのテキストの完全自動処理を可能にします。

このテキスト処理を自動化することにより次のことが可能です。

- クライアントが要求する高速な稼働速度
- 最新のn-gramアルゴリズムを使用した、より精度の高い判別
- わずか10MBのメモリー消費

### お問合せ

さらに詳しい製品情報ならびに評価版のご利用をご希望の方は下記へご連絡ください。

[info@basistech.jp](mailto:info@basistech.jp)

[www.basistech.jp](http://www.basistech.jp)

電話 03-3511-2947

### \*注1:

文字コードとは、コンピュータで文字を扱うために、特定の文字集合を定めてその各文字に一意的な数値を割り当てたものです。たとえば、日本語テキストではISO-2022-JP, Shift-JIS, EUC-JPなどの文字コードが使用されますが、同じ文字でも文字コードによって割り当てられている数値が異なります。たとえば、Shift-JISでは「あ」のコードは「82A0x」ですが、EUC-JPでは「A4A2x」です。そのため、Shift-JIS データを誤ってEUC-JP, JIS-2004 (JISX0213) として処理してしまうと文字化けが生じるのです。

## 対応言語および文字コード一覧

Rosette 言語・文字コード判別システムは、世界の主要55言語、ならびに45種の文字コードと7種類の字訳をサポートしています。

日本語 — EUC-JP, ISO-2022-JP, Shift-JIS, Shift-JIS-2004 (JIS X 0213)  
朝鮮語 — EUC-KR, ISO-2022-KR  
中国語(簡体字) — GB-2312, GB-18030, HZ-GB-2312, ISO-2022-CN  
中国語(繁体字) — Big5, Big5-HKSCS  
英語 — ISO-8859-1, Windows-1252  
アイスランド語 — ISO-8859-1, Windows-1252  
アラビア語 — ISO-8859-6, Windows-720, Windows-1256  
アルバニア語 — ISO-8859-1, Windows-1252  
イタリア語 — ISO-8859-1, Windows-1252  
インドネシア語 — ISO-8859-1, Windows-1252  
ウクライナ語 — ISO-8859-5, Windows-1251, KOI8-R  
ウズベキスタン語 — ISO-8859-5, Windows-1251, KOI8-R  
ウルドゥ語 — ISO-8859-6, Windows-1256  
エストニア語 — ISO-8859-13, Windows-1257  
オランダ語 — ISO-8859-1, Windows-1252  
カタロニア語 — ISO-8859-1, Windows-1252  
カンナダ語 — ISCII-Kannada  
ギリシャ語 — ISO-8859-7, Windows-1253  
グジャラート語 — ISCII-Gujarati  
クルド語 — Windows-1256  
クロアチア語 — Windows-1250  
スウェーデン語 — ISO-8859-1, Windows-1252  
スペイン語 — ISO-8859-1, Windows-1252  
スロバキア語 — Windows-1250  
スロベニア語 — Windows-1250  
セルビア語 — ISO-8859-5, Windows-1251  
ソマリ語 — ISO-8859-1, Windows-1252  
タイ語 — Windows-874  
タガログ語 — ISO-8859-1, Windows-1252  
タミル語 — ISCII-Tamil  
チェコ語 — ISO-8859-2, Windows-1250

テルグ語 — ISCII-Telugu  
デンマーク語 — ISO-8859-1, Windows-1252  
ドイツ語 — ISO-8859-1, Windows-1252  
トルコ語 — ISO-8859-9, Windows-1254  
ノルウェー語 — ISO-8859-1, Windows-1252  
バシュトゥ語 — ISO-8859-6, Windows-1256  
ハンガリー語 — ISO-8859-2, Windows-1250  
ヒンディー語 — ISCII-Hindi  
フィンランド語 — ISO-8859-1, Windows-1252  
フランス語 — ISO-8859-1, Windows-1252  
ブルガリア語 — ISO-8859-5, Windows-1251, KOI8-R  
ベトナム語 — TCVN, VIQR, VISCII, VNI, VPS  
ヘブライ語 — ISO-8859-8, Windows-1255  
ペルシア語 — ISO-8859-6, Windows-1256  
ベンガル語 — ISCII-Bengali  
ポルトガル語 — ISO-8859-1, Windows-1252  
ポーランド語 — ISO-8859-2, Windows-1250  
マケドニア語 — ISO-8859-5, Windows-1251  
マラーヤラム語 — ISCII-Malayalam  
マレー語 — ISO-8859-1, Windows-1252  
ラトビア語 — ISO-8859-13, Windows-1257  
リトアニア語 — ISO-8859-13, Windows-1257  
ルーマニア語 — ISO-8859-2, Windows-1250  
ロシア語 — ISO-8859-5, Windows-1251, KOI8-R, IBM-866, Mac Cyrillic  
音訳アラビア語 — ISO-8859-1, Windows-1252, Windows-1256  
音訳ウズベキスタン語 — Windows-1251  
音訳ウルドゥ語 — ISO-8859-1, Windows-1252  
音訳クルド語 — ISO-8859-1, Windows-1252, Windows-1256  
音訳セルビア語 — ISO-8859-2, Windows-1250  
音訳バシュトゥ語 — ISO-8859-1, Windows-1252  
音訳ペルシア語 — ISO-8859-1, Windows-1252, Windows-1256

## 対応プラットフォーム

以下のプラットフォーム対応のSDKを提供します。その他のプラットフォームのサポートも、ご要望に応じ対応します。

AIX 6.1, PPC  
HP-UX 11i, IA64  
Linux CentOS 4.x/5.x, IA32/AMD64  
Linux Debian 5.x, IA32/AMD64

Linux Red Hat 4.x/5.x, IA32/AMD64  
Linux Ubuntu 10.x/11.x, IA32/AMD64  
MacOS  
Solaris 10, SPARC32/64, IA32/AMD64

Windows XP/Vista/7, IA32/AMD64  
Windows Server 2003, 2008

詳細 [www.basistech.jp](http://www.basistech.jp) お問い合わせ [info@basistech.jp](mailto: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

