

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

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

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

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

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ます。

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 39 種の文字コードと 6 種類の字訳をサポートしています。

アルバニア語 – ISO 8859-1, Windows-1252	リトアニア語 – ISO 8859-13, Windows-1257
アラビア語 – ISO 8859-6, Windows-720, Windows-1256	マレー語 – ISO 8859-1, Windows-1252
音訳アラビア語 – ISO 8859-1, Windows-1252, Windows-1256	マラーヤラム語 – ISCII-Malayalam
ベンガル語 – ISCII-Bengali	ノルウェー語 – ISO 8859-1, Windows-1252
ブルガリア語 – ISO 8859-5, Windows-1251, KOI8-R	パシュトゥ語 – ISO 8859-6, Windows-1256
カタロニア語 – ISO 8859-1, Windows-1252	音訳パシュトゥ語 – ISO 8859-1, Windows-1252
中国語(簡体字) – GB-2312, HZ-GB-2312, GB-18030, ISO 2022-CN	近代ペルシア語 – Windows-1256
中国語(繁体字) – Big5, Big5-HKSCS	音訳ペルシア語 – ISO 8859-1, Windows-1252, Windows-1256
クロアチア語 – Windows-1250	ポーランド語 – ISO 8859-2, Windows-1250
チェコ語 – ISO 8859-2, Windows-1250	ポルトガル語 – ISO 8859-1, Windows-1252
デンマーク語 – ISO 8859-1, Windows-1252	ルーマニア語 – ISO 8859-2, Windows-1250
オランダ語 – ISO 8859-1, Windows-1252	ロシア語 – ISO 8859-5, Windows-1251, IBM-866, Mac Cyrillic, CP20866 (KOI8-R)
英語 – ISO 8859-1, Windows-1252	セルビア語 – ISO 8859-5, Windows-1251
エストニア語 – ISO 8859-13, Windows-1257	音訳セルビア語 – ISO 8859-2, Windows-1250
フィンランド語 – ISO 8859-1, Windows-1252	スロバキア語 – Windows-1250
フランス語 – ISO 8859-1	スロベニア語 – Windows-1250
ドイツ語 – ISO 8859-1, Windows-1252	ソマリ語 – ISO 8859-1, Windows-1252
ギリシヤ語 – ISO 8859-7, Windows-1253	スペイン語 – ISO 8859-1, Windows-1252
グジャラート語 – ISCII-Gujarati	スウェーデン語 – ISO 8859-1, Windows-1252
ヘブライ語 – ISO 8859-8, Windows-1255	タガログ語 – ISO 8859-1, Windows-1252
ヒンディー語 – ISCII-Hindi	タミル語 – ISCII-Tamil
ハンガリー語 – ISO 8859-2, Windows-1250	テルグ語 – ISCII-Telugu
アイスランド語 – ISO 8859-1, Windows-1252	タイ語 – Windows-874
インドネシア語 – ISO 8859-1, Windows-1252	トルコ語 – ISO 8859-9, Windows-1254
イタリア語 – ISO 8859-1, Windows-1252	ウクライナ語 – ISO 8859-5, KOI8-R, Windows-1251
日本語 – EUC-JP, ISO-2022-JP, Shift-JIS, Shift-JIS-2004 (JIS X 0213)	ウルドゥ語 – ISO 8859-6, Windows-1256
カンナダ語 – ISCII-Kannada	音訳ウルドゥ語 – ISO 8859-1, Windows-1252
朝鮮語 – EUC-KR, ISO-2022-KR	ウズベキスタン語 – ISO 8859-5, KOI8-R, Windows-1251
クルド語 – Windows-1256	音訳ウズベキスタン語 – Windows-1251
音訳クルド語 – ISO 8859-1, Windows-1252, Windows-1256	ベトナム語 – TCVN, VIQR, VISCI, VNI, VPS
ラトビア語 – ISO 8859-13, Windows-1257	

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 5.2, PowerPC	Linux Fedora Core 4, IA32/AMD64	Solaris 8, SPARC32/64
FreeBSD 4.8, IA32	Linux Fedora Core 5, IA32/AMD64	Solaris 9, SPARC32/64, IA32
FreeBSD 6.0, IA32/AMD64	Linux Red Hat ES 2.1, IA32	Solaris 10, SPARC32/64, IA32/AMD64
HP-UX 11.0, PA-RISC	Linux Red Hat ES 3.0, IA32/AMD64	Windows NT/XP/2003, IA32/AMD64
HP-UX 11.22, IA64	Linux Red Hat ES 4.0, IA32/AMD64	Windows Vista/2008, IA32/AMD64
Linux Debian 3.1, IA32/AMD64	Linux Suse EL 10, IA32/AMD64	

© 2009 Basis Technology Corporation. All rights reserved. “Basis Technology”, “Rosette”, and “We Put the World in the World Wide Web” are registered trademarks. All other company and product names mentioned are trademarks or registered trademarks of their respective owners. (AD09.10.22)



詳細
www.basistech.jp

お問合せ
info@basistech.jp

電話番号
03-3511-2947

ベイシス・テクノロジー株式会社
〒102-0084 東京都千代田区二番町 9-6

U.S.本社 (ボストン)
One Alewife Center, Cambridge, MA 02140

ワシントン D.C. 支社
13800 Coppermine Road, Herndon, VA 20171