

## Rosette® 中国語形態素解析システム

### 応用例

- データベースおよび Web 検索
- 文書管理システム
- 機械翻訳システム
- コンテンツ・フィルタリング
- オンライン・リテイル
- 文書要約・文書分類システム
- 索引付け・キーワード抽出
- ドキュメント・オーサリング
- テキスト・マイニング
- CRM
- ERP

手書き文字・音声認識用  
ポストプロセッサ

Basis Technology の Rosette® 中国語形態素解析システム (RBL-ZH : Rosette® Base Linguistics for Chinese) はポータブルで高性能な中国語テキスト単語分節エンジンです。ウェブページ、電子メール、各種データベースの文書など、どのような種類の文書にも RBL-ZH は威力を発揮します。

中国語テキストは通常スペースなしで書かれているため、索引付け、キーワード検索といったテキスト処理の前にいくつかのステップが必要です。

Basis Technology の RBL-ZH はこれらの処理をおこなうためのアルゴリズムを持ち、さらに品詞情報およびピンインによる読みデータを収録した辞書に基づき、中国語テキストを正確に単語分割します。

このソフトウェア・ライブラリはポータブルで、デスクトップ PC から、1 分間に数百文書进行处理するような高速マルチ CPU ウェブサーバまで、多くのプラットフォームで動作可能です。

### 応用分野

RBL-ZH は中国市場ですでにご活躍の企業、また今後中国市場参入を目指す企業で幅広くご活躍いただけます。特にインターネット用ソフトウェア、企業用アプリケーション、また、消費者直結型の e ビジネスにおいても高い利用価値をもっています。

特に膨大な中国語文書の索引付けを必要とする情報検索、または分節処理や語幹抽出、品詞分析を必要とする自然言語処理にその威力を発揮します。

### 特長

- ・ 辞書形式はテキスト正規化に対応
- ・ ストップワードのユーザー定義が可能
- ・ 句読点、スペース、数字の識別およびフィルタリング
- ・ ISO/ANSI C++ポータブルサブセットによる実装
- ・ 完全な Unicode 内部アーキテクチャ
- ・ スレッドセーフ
- ・ ユーザー定義辞書の利用が可能

## 仕様

- ・ スタティックあるいはダイナミックライブラリ
- ・ 簡体字および繁体字をサポート
- ・ 内部アーキテクチャには Unicode を使用。入出力は Big5、GB-2312、HZ、CNS-11643、GB/K 等に対応。
- ・ 辞書をメモリーにマップし、必要なメモリー量を小さく抑え、素早い起動を実現

## 辞書データ

繁体字と簡体字あわせて約 50 万語を収録

\* さらに 170 語以上の新語が追加

- ・ 固有名詞（繁体字・簡体字）約 15 万語

## 対応プラットフォーム

SDK は以下のプラットフォームでご利用いただけます。

- ・ Microsoft Windows
- ・ Sun Solaris
- ・ Linux
- ・ MacOS

他のプラットフォームへの対応もご要望により承ります。

年間テクニカルサポート、アップグレード契約もあります。