

## Rosette® 英語形態素解析システム

### 応用例

ENLAは言語的な処理を必要とするさまざまなアプリケーションにご利用いただけます。

- 情報検索
- データ解析
- 索引/キーワード抽出
- CRM
- ドキュメント管理
- ヘルプ ディスカバリー

Basis Technology の Rosette® 英語形態素解析システム (ENLA: English Language Analyzer) は、英文テキストを処理するポータブルで高性能な言語解析エンジンです。ウェブページ、電子メール、データベースの文書などを対象とする情報検索およびテキスト・マイニングにおいて ENLA は威力を発揮します。

特異な綴りと使用法をもつ多くの外来語や新語など、膨大かつ多様な語彙が英語の自動解析を難しいものにしています。テキスト処理も同様に容易ではありません。動詞と前置詞で構成された熟語 (例: Pick up) は、別の単語が間に入っても意味は変わりません (例: John picked the newspaper up)。それに加え、“be” “have” “do” などの助動詞は、文章により助動詞とされる場合と、動詞と判断される場合があるため、解析作業をさらに複雑にします。

Basis Technology の ENLA は、基本化およびステミング、または品詞判別などを含む、非常に高性能な英語形態素解析によりこれらの問題を解決します。

### 特長

#### 正規化

テキストの正規化、短縮形のスペルアウト・分割、トークンの大文字から小文字への変更をおこないます。

- ・ 例) 短縮形のスペルアウト・分割  
aren' t → are not
- ・ 例) 大文字から小文字への書き換え  
Computer → computer

#### 基本化・ステミング

接辞を削除し、単語を原型または基本形へと変換します。

例) Walked → Walk、whiteness → white

例) Went → go、their → they

#### 品詞判別

名詞、固有名詞、動詞、形容詞等の品詞を正確に判別します。

## 仕様

- ・ 辞書
- ・ 未知語の学習的認識機能
- ・ C++/C/Java APIs
- ・ 内部アーキテクチャに Unicode を使用
- ・ スレッドセーフ

## 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 NT/2000/XP
- ・ Sun Solaris
- ・ Linux
- ・ Hewlett-Packard HP-UX
- ・ IBM AIX

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。

## その他の対象言語

Rosette® 形態素解析システムは、英語、アラビア語、アジア言語の他に下記の言語にも対応しています。

詳しくは下記までお問い合わせください：

[Tokyo.marketing@basistech.com](mailto:Tokyo.marketing@basistech.com)

その他の対象言語：

- ・ フランス語
- ・ イタリア語
- ・ ドイツ語
- ・ スペイン語
- ・ オランダ語
- ・ ポルトガル語
- ・ ロシア語
- ・ ギリシャ語
- ・ チェコ語
- ・ ハンガリー語
- ・ ポーランド語 他

\* そのほか現在開発中の言語についてもお問い合わせください。