

Rosette® 欧州語形態素解析システム

欧州言語のテキスト処理にはそれぞれ独特な課題があります。例えば、ドイツ語の文書には多くの複合名詞が含まれており、テキスト処理をする前にそれら複合名詞の分解が必要です。また、スペイン語、フランス語、ドイツ語、イタリア語などでは、複数の基本語から作られた正字法が存在します。例えば、フランス語では、“de”と“le”が隣りあった場合、“du”という1つの単語に変化します。英語の場合には、“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 Rosette® 欧州語言語処理システムは、このような欧州語のテキストを、検索エンジン、テキストマイニングツール、分類エンジンなどで効率的に扱うことができるように設計しています。

コアとなる Rosette 言語処理プラットフォームにより、複数の言語対応モジュールが一つの API から利用できるよう設計されており、それぞれの言語および機能の ON/OFF の切り替えが可能です。

本製品の特長：

- ・ 欧州語テキストをより正確に処理するための準備をおこなう。

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

そのほか現在開発中の言語に関するお問い合わせは下記まで：

Tokyo.marketing@basistech.com

Basis Technology の Rosette 欧州語形態素解析システムは、正規化、ステミング、品詞解析などの言語処理をおこないます。

特長

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。

入力 lu 出力 lire
入力 gezogen 出力 ziehen

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。

- éditeurs --> N_PL
- Heiße --> ADJ

名詞の分解

正確な情報検索を行うために、(ドイツ語などいくつかの言語にみられる) 複合名詞を基本語に分解します。

入力: Kontrollsysteme 出力: [Kontroll] [systeme]

入力: Jugendarbeit 出力: [Jugend] [arbeit]

文脈による解析

基本辞書あるいはカスタム辞書に登録されていない語を文脈から分析します。

応用例

- 情報検索
- データ解析
- 索引付け/キーワード抽出
- CRM
- 文書管理
- ナレッジ・ディスカバリー

仕様

- C/C++/Java API
- 内部アーキテクチャに Unicode を使用
- 複数のユーザー定義辞書の利用が可能

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/Server2003
- 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。