

# Rosette 朝鮮語形態素解析システム

## 検索システムのキーワード抽出などの精度を改善

朝鮮語は、音韻と字母を組み合わせる表記形式のため、形態素解析を行う上で解決すべき問題が複数存在します。特に、単語の境界識別をコンピュータ処理することが難しいですが、弊社のRosette® 朝鮮語形態素解析システム (RBL-KO: Rosette Base Linguistics for Korean) は、朝鮮語テキストの分節処理に非常に優れた力を発揮します。

朝鮮語テキストの正確な分節を素早くおこなうために、統計的な技法を利用し、さらに詳細な辞書をベースとしています。統計的な技法には、語の頻度に関する情報、語形変化の法則を反映し、辞書データは定期的に更新しています。

人名、地名、日時などの特定語句の検出や抽出が可能な、Rosette 朝鮮語固有表現抽出システムもご用意しています。

### 特長

- テキスト正規化のために基本形を返す
- ストップワードのユーザー定義が可能
- 複合語の識別、およびその分解(随意)
- 句読点、スペース、数字の識別およびフィルタリング

### 仕様

- ISO/ANSI C++ ポータブルサブセットによる実装
- スレッドセーフ
- ユーザー定義辞書の利用が可能
- Unicode UCS-2 による朝鮮語テキストを処理

부산 지역 한 복당과 조선 의원은 "청와대나 당으로부터 어떤 연락도 받지 못했다"면서 "당협위원장들이 나름대로 고생했으니가 배려 차원에서 한 것으로 이해는 하고 있지만, 양해라도 구했으면 될 일을 그러지 못한 것 같아 아쉽다"고 말했다.

다른 복당 중진 의원은 "영부인이 당협위원장 부인들과 여성당 직자를 부르는 일이야 당연하지만, 복당과를 배려하지 않는 것은 문제"라고 지적했다.

한 영남권 의원은 "친박 복당을 받았으면 우리 먼저 배려해야 되는 것 아니냐"면서 "이명박 대통령도 어쨌거나 복당이 됐으면 한 번은 우리를 청와대로 불러야 하는데, 아무도 안 부르고 이래저래 너무하는 것 아니냐"고 불만을 터뜨렸다.

### 辞書データ

- 汎用語:60,000語
- 助詞:1,200語
- 語尾形式:1,300語
- 数百以上の語形式を規則により処理
- 朝鮮文字、ハンジャ漢字、英数字を処理

\* 朝鮮語辞書と複合名詞辞書がひとつになり、編集およびコンパイル(または再コンパイル)が可能なユーザー定義辞書となりました。

### 応用例

- エンタープライズ用 検索エンジン
- Web検索
- 文書管理システム
- ドキュメント・オーサリング
- 機械翻訳システム
- テキスト・マイニング
- コンテンツ・フィルタリング

## 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	Linux Ubuntu 10.x/11.x, IA32/AMD64
HP-UX 11i, IA64	MacOS
Linux CentOS 4.x/5.x, IA32/AMD64	Solaris 10, SPARC32/64, IA32/AMD64
Linux Debian 5.x, IA32/AMD64	Windows XP/Vista/7, IA32/AMD64
Linux Red Hat 4.x/5.x, IA32/AMD64	Windows Server 2003, 2008

## 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。

[info2011@basistech.jp](mailto:info2011@basistech.jp)

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

電話 03-3511-2947

詳細 [www.basistech.jp](http://www.basistech.jp) お問合せ [info2011@basistech.jp](mailto:info2011@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

