

Rosette 朝鮮語形態素解析システム

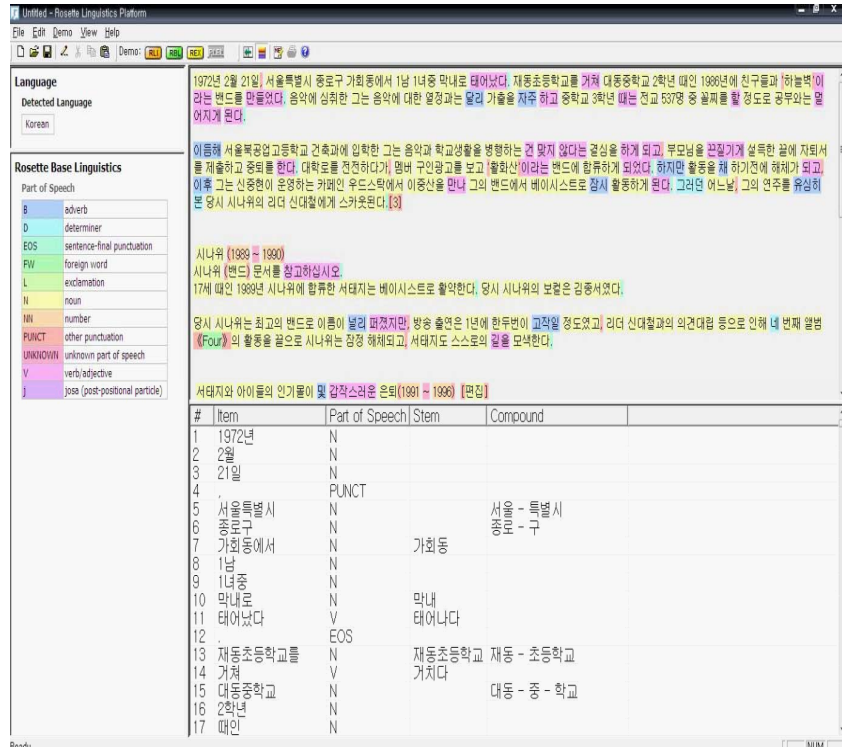
検索システムのキーワード抽出などの精度を改善

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-K0: 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。

出力例：



The screenshot shows the Rosette Base Linguistics Platform interface. The main window displays a Korean text snippet with various parts of speech highlighted in different colors. Below the text, a table lists the extracted items with their corresponding part of speech, stem, and compound information.

| # | Item | Part of Speech | Stem | Compound |
|----|---------|----------------|--------|-------------|
| 1 | 1972년 | N | | |
| 2 | 2월 | N | | |
| 3 | 21일 | N | | |
| 4 | . | PUNCT | | |
| 5 | 서울특별시 | N | | 서울 - 특별시 |
| 6 | 종로구 | N | | 종로 - 구 |
| 7 | 가회동에서 | | 가회동 | |
| 8 | 1남 | N | | |
| 9 | 1녀 | N | | |
| 10 | 막내로 | N | 막내 | |
| 11 | 태어났다 | V | 태어나다 | |
| 12 | . | EOS | | |
| 13 | 재동초등학교를 | N | 재동초등학교 | 재동 - 초등학교 |
| 14 | 가치 | V | 가치다 | |
| 15 | 대중중학교 | N | | 대동 - 중 - 학교 |
| 16 | 2학년 | N | | |
| 17 | 때인 | N | | |

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

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

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。

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。