

株式会社クレストック / アイビー・システム株式会社

本社住所	〒431-3105 静岡県浜松市東区笠井新田町 676 番地 株式会社クレストック
URL	http://www.crestec.co.jp
展示名	多角的な文書比較『歌詠』と カスタマイズ容易な日本語精査『鶯』
お問 合せ先	担当部署 アイビー・システム(株) やまと歌担当宛 TEL 011-271-7117 FAX 011-271-0550 E-mail yamatouta@ivysystem.co.jp http://www.ivysystem.co.jp/yamatoUta/index.html

■アピールポイント

歌詠は、任意の文書を多角的に比較し、差分を対比表形式で表示するツールです。段落や文ごとの類似度を指定して目的に合った比較ができます。

鶯は、オープンソース形態素解析エンジン MeCab の結果から、文章が正しいか、業界ルールに従っているか判断してコメントするツールです。精査ルールの一部は、カスタマイズが可能です。

【産業日本語との関連】

原文と産業日本語文章を文ごとに比較して結果を保存しておく、翻訳に適した文にするための参考情報になります。類似した文の並びから表記のゆれを発見できます。形態素情報を利用して用語の抽出や日本語の精査ができます。

【詳細】

<システムの形態>

1. デスクトップアプリケーションであるスタンドアロン版(Windows XP / Vista / 7 / 8 32/64bit)
2. セキュア・ウェブアプリケーションとしてのカスタマイズ版(Windows Server / Linux)

<比較機能『歌詠』の特徴>

差分の対比表示

Microsoft Word の比較と異なり、差分結果を分かり易い対比表形式で Web Browser に表示します。外部アプリケーション(Microsoft Word)は不要です。比較結果は、目的に適した表示へ動的に切り替え可能です。ウェブ版では基準文書と2つ以上の文書を同時に比較できます。

移動した要素の検知

比較する文書の間で内容の位置だけが変化していた場合、移動元と移動先を表示します。

差分計算

段落(改行)や文(日本語は句点「。」、英語はピリオド「.」)ごとに区切り、文字または形態素の単位で比較します。差異の大きさは、Levenshtein の編集距離を用いて算出します。

類似度

比較対象の編集距離が小さいほど、差異は小さくなり、類似度は高くなります。したがって一定の類似度閾値以上の場合は、変更された箇所と見なし、下回る場合は、別の内容の要素であると判断します。比較する前に類似度を指定し、比較対象文書に即した所望の結果を取得できます。

比較オプション

文書の先頭から順番に内容を比較する機能、または類似度が最大の要素を探してそれを比較する(文書の構成が異なる文書や行の入れ替えがある表の比較のための)機能を選択できます。

表記ゆれの検出

類似度が高い文すべてをリストアップし、文書内にある表記ゆれの可能性を検出します。

英文比較

和文と和文以外(欧米文)を区別した比較処理によって、類似度精度を向上しています。

<日本語精査機能『鶯』の特徴>

精査ルール

以下のように CSV 形式で記述しているため、ルールの削除、変更、追加が可能です。

文章中から用語として取り出す形態素パターンの一部 (*は任意の値)

○-含める ×-除く	着目する単語				直前の単語				直後の単語			
	単語	品詞	細分1	細分2	単語	品詞	細分1	細分2	単語	品詞	細分1	細分2
○	*	名詞	サ変接続	*	*	接頭詞	名詞接続	*	*	動詞	*	*
×	*	名詞	サ変接続	*	*	*	*	*	*	動詞	*	*
○	*	名詞	サ変接続	*	*	*	*	*	*	名詞	*	*

用語の抽出

以上のルールによって、文章中にある見逃し易い用語を抽出し、用語辞書にある登録済の用語と照合します。一般的な用語か、新しい用語か、用語にゆれがないか発見する支援をします。

文書の内容精査の一例

- 文章内にある誤り可能性(脱字、誤字、不要な文字)を指摘します。
- 誤りやすい慣用句、用字・用語のチェック、活用形のチェックをします。
- 特許請求の範囲において、「前記」や「該」に続く語句が初出の場合、指摘します。

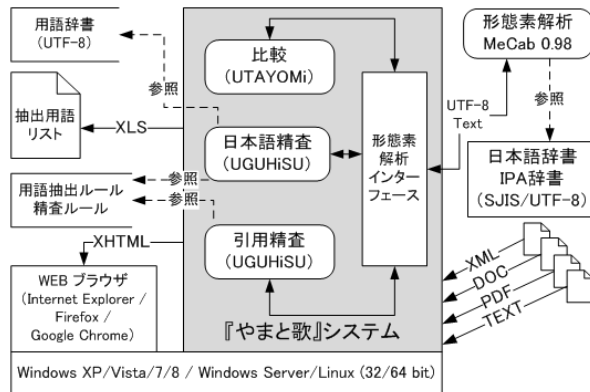
形態素解析の結果から誤り可能性を抽出するルールの一部 (*は任意の値)

直前の単語				直前の単語の例外				直後の単語				直後の単語の例外			
単語	品詞	細分1	活用型	細分1	活用型	活用	基本形	単語	品詞	細分1	活用型	細分1	活用型	活用	基本形
*	助詞	係助詞	*	*	*	*	*	*	助詞	係助詞	*	*	*	*	*
*	助動詞	*	*	*	*	*	*	*	動詞	*	*	自立	*	*	*

精査結果の表示

文書比較時、精査を同時に実行すると比較前後における精査コメントの変化も確認できます。

<モジュール構成概要>



<主な機能>

