



Code

Archive



cmecab-java

MeCabのJavaバインディング + Lucene/Solr用トークナイザ・フィルタ

<http://github.com/takscape/cmecab-java> に移転しました。

何ですかこれは

MeCabをJavaから利用するためのライブラリです。以下のコンポーネントから構成されます。

- MeCabのJavaバインディング
 - SWIG不使用、辞書のエンコーディングとUnicodeの間で自動変換を行います。
- MeCabを用いたLucene用Tokenizer、素性情報を用いてフィルタリングを行うTokenFilter、およびAnalyzer
- TinySegmenterのJava移植版およびTokenizer, TokenizerFactory
- サロゲートペアを正しく認識するCJKTokenizerクローン
- Solr用TokenizerFactory、TokenFilterFactory
- GroovyでSolrのTokenizer, TokenFilterを書くためのFactory類

ライセンス

- Public Domain (プロジェクト情報にはMIT licenseと書いてありますが、本当はPublic Domainです)
- ただし、TinySegmenterに関する部分のみは、修正BSDライセンスです。

ドキュメント類

- インストール方法
- 使用方法
- リリースノート

Project Information

- License: [MIT License](#)
- 13 stars
- svn-based source control

Labels:

[mecab](#) [java](#) [solr](#) [lucene](#) [tinysegmenter](#)