

中古和文 UniDic ユーザーズマニュアル

Ver. 1.2 2012/04/23 小木曾智信

1. 中古和文 UniDic とは

中古和文 UniDic は、現代語向けの UniDic をもとにして中古和文文を解析できるようにした形態素解析辞書です。現代語版と同様、次のような言語研究に適した特長を持っています。

- ・ 「短単位」という揺れがない齊一な単位で設計されています。
- ・ 語彙素・語形・書字形・発音形の階層構造を持ち、表記の揺れや語形の変異にかかわらず同一の見出しを与えることができます。

ただし、UniDic の特長である発音形・アクセント等の音声情報は、現代の読み方による限定的なものです。

中古和文 UniDic が主たる対象とするテキストは、近代の文語論説文（明治普通文）です。一般的な文語論説文のテキストであれば、概ね 95%～98%程度の精度で解析することができます。

中古和文 UniDic は現代語の UniDic をもとにして作成されており、見出しに互換性があります（見出し語についてみると、中古和文 UniDic はほぼ現代語 UniDic のスーパーセットになっています¹⁾）。このため、現代語版の UniDic で解析した結果と中古和文 UniDic で解析した結果を比較することができます。

利用条件

中古和文 UniDic の利用条件は次の通りです。現代語版 UniDic と同等の内容です。

【中古和文 UniDic 利用条件】

1. 中古和文 UniDic の著作権は、小木曾智信、小椋秀樹、近藤明日子 および The UniDic consortium（代表：伝康晴）が保持する。
2. 中古和文 UniDic を複製又は改変することは、個人的な利用に限り認める。
3. 中古和文 UniDic 及びこれを改変したものを再配布してはならない。

¹⁾顔文字など中古和文文には現れない一部の見出し語は削除されています。

4. 中古和文 UniDic を利用して行った研究等の成果を公表する場合は、中古和文 UniDic を利用したことを明記すること。
5. 営利を目的として、中古和文 UniDic を利用する場合は、事前に著作権者と協議すること。
6. 中古和文 UniDic を利用することによって、直接的・間接的に生じたいかなる損害についても、著作権者は賠償する責任を負わない。
7. 本文書に定めのない事項については、著作権者と協議すること。

ダウンロードサイト

中古和文 UniDic は次の Web ページからダウンロードすることができます。

<http://www2.ninjal.ac.jp/lrc/index.php?UniDic>

2. 利用方法

中古和文 UniDic は Windows 版パッケージとバイナリ辞書の二つの形態で配布されています。一般の方は前者をお使いください。MeCab 等のツールに詳しい方、Windows 以外の OS を利用している方はバイナリ辞書をお使いください。

Windows 版パッケージ

(Windows 上で簡単に利用できるようにした配布形態です。)

Windows パッケージは簡単にインストールして利用することができるようになっていました。ダウンロードしたファイルを実行するとインストーラが起動しますので、あとは画面上の指示に従ってください。ただし、中古和文 UniDic の他に、形態素解析器として ChaSen または MeCab をインストールする必要があります。詳しくは「インストールマニュアル」を御覧ください。

形態素解析は、同梱されている解析用のプログラム「近代茶まめ」を起動することで手軽に利用できます。インストール後に現れる「近代茶まめ」アイコンから起動してください。詳しい使い方は「茶まめマニュアル」を御覧ください。

バイナリ辞書

(詳しい方向けの配布形態です。辞書本体以外のツールを同梱していませんので、現代語版 UniDic のツールをご利用ください。)

ダウンロードしたファイルを展開し、ディスク上の任意の場所にコピーしてください。
辞書は UTF-8 版のみ公開しています（エンディアンが同じであれば 64bit 環境でも使えます）。解析のためには形態素解析器が必要ですので、ダウンロードした辞書にあわせ、ChaSen 2.4.3 以降または MeCab0.96 以降をインストールしてください。

形態素解析の実行は次のように行います。

MeCab の場合

```
mecab -d /usr/local/unidic/dic/unidic-mecab FILE
```

ChaSen の場合（UniDic 標準の XML タグ形式で出力する場合）

```
chasen -i w -r /usr/local/unidic/dic/unidic-chasen FILE
```

ChaSen で表形式テキストの出力を利用したい場合には、現代語版 UniDic パッケージ付属の chauni を利用してください。

出力結果の列名は、現代語版 UniDic のユーザズマニュアルを御覧ください。また、辞書に同梱されている chasenrc, dicrc を参考にしてください。

ソース辞書

中古和文 UniDic のソース辞書は現在一般公開を行っていません。研究目的で利用される場合にはデータをお送りしますので、必要な場合には連絡先までご相談ください。

3. 短単位規定

中古和文 UniDic の短単位の規定については、『中古和文 UniDic 短単位規程集』をご覧ください。上記ダウンロードサイトから入手できます。

4. 参考文献等

中古和文 UniDic に関する参考文献等は、『和文系資料を対象とした形態素解析辞書の開発・研究成果報告書』をご覧ください。上記ダウンロードサイトから入手できます。