

国立国語研究所学術情報リポジトリ

Improvement of “Web ChaMame” and experimental production of “Web ChaMame Web API”

メタデータ	言語: jpn 出版者: 公開日: 2018-03-20 キーワード (Ja): キーワード (En): 作成者: 川口, 寛治, 薦田, 龍輝, 堤, 智昭, KAWAGUCHI, Motoharu, KOMODA, Ryuki, TSUTSUMI, Tomoaki メールアドレス: 所属:
URL	https://doi.org/10.15084/00001481

形態素解析ソフトウェア『Web 茶まめ』の改良と Web API の試作

川口寛治(東京電機大学)

薦田龍輝(東京電機大学)

堤智昭(東京電機大学)

Improvement of “Web ChaMame” and experimental production of

“Web ChaMame Web API”

Motoharu Kawaguchi (Tokyo Denki University)

Ryuki Komoda (Tokyo Denki University)

Tomoaki Tsutsumi (Tokyo Denki University)

要旨

国立国語研究所では、様々な時代の日本語資料の分析に利用可能な形態素解析辞書である UniDic を公開している。この UniDic を利用した形態素解析支援アプリケーションである『Web 茶まめ』は Web 上で公開されており、インターネットを通じて誰でも利用できる。本稿では、『Web 茶まめ』について以下の二点を報告する。一点目は、『Web 茶まめ』を公開した以降にユーザから寄せられた意見や指摘をもとに行った改良についてである。二点目は、ブラウザを用いずにインターネットを通じて Web 茶まめの機能を利用するための、WebAPI の試作についてである。

1. はじめに

国立国語研究所では、UniDic[1] [2] [3]を用いた形態素解析の支援を目的とした Web アプリケーションである Web 茶まめ[4]を公開している。Web 茶まめは Web 上で公開されており、インターネットを通じて誰でもアクセス、利用が可能である。

本稿では、Web 茶まめの解析機能の改良と、WebAPI の試作について報告する。解析機能の改良は、ユーザの使い勝手向上を目的とし、公開後にユーザから寄せられた意見を元に行った。WebAPI の試作については、Web 茶まめの機能を、ブラウザを介さず、他のシステムから利用し易い形で提供することを目的に行った。WebAPI では HTTP 通信を用いてデータを送受信することで、Web 茶まめの機能を提供する。

2. Web 茶まめ概要

Web 茶まめは、図 1 に示すようにサーバ内の形態素解析エンジン MeCab[5]と、9種類の UniDic と 1種類の IPAdic を用いて形態素解析を行い、その結果を Web ブラウザ上に表示する形態素解析支援アプリケーションである。形態素解析処理はすべて Web 茶まめサーバ内で行われるため、特別なソフトウェアのインストールや特別な環境を用意することなく形態素解析を行う事ができる。また、Web 茶まめに新しい機能を実装したり動作の変更を行う場合には、ユーザの環境を変更する必要がなく、サーバにインストールされたソフトウェアを更新することで実現できる。

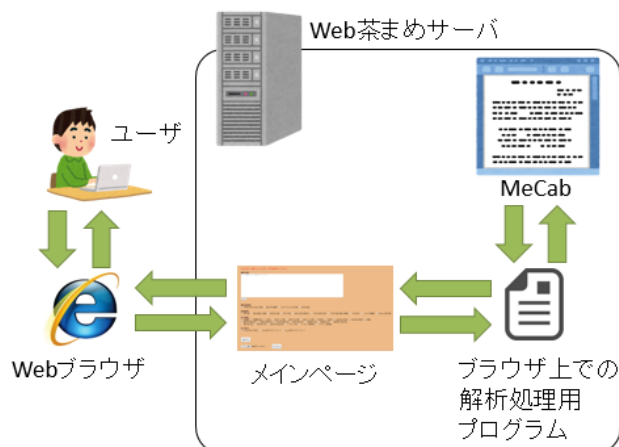


図 1 Web 茶まめ概要図

3. Web 茶まめの改良

3.1 品詞項目の細分化

従来の Web 茶まめにおいて、品詞は 1 つの出力項目として処理されていた。しかし、Web 茶まめが用いる形態素解析辞書である UniDic では、辞書内の項目として品詞の分類が、大分類、中分類、小分類、細分類の 4 つに分かれている。そこで、主によく使われる、大分類、中分類、小分類を出力項目として追加し、個別に出力するか否かを切り替えられるように変更した。

3.2 出力文字コードの選択機能の追加

Web 茶まめでは、形態素解析の結果を HTML、CSV、EXCEL の 3 形式で出力することが可能である。そのうち、CSV 出力時には、文字コードとして UTF-8 を用いていた。しかし、Windows に代表される一部の OS では ShiftJIS が使われることが多く、ユーザは出力データの文字コードを、必要に応じてその都度自前で変更する必要があった。そこで、CSV 出力を選択した場合には、出力データの文字コードを UTF-8 と ShiftJIS の 2 種類から選択可能とした。

3.3 IPAdic を用いた解析の改良

Web 茶まめでは UniDic の他に、IPAdic を用いた形態素解析が可能である。しかし、IPAdic は UniDic とは辞書内の項目数、及び項目名が異なる。そのため、UniDic と同様の項目を指定し形態素解析を行った場合、図 2 に示すように項目の違いに依存したエラーが発生してしまう。この問題を解決するためには、IPAdic の項目に準拠した出力項目の指定が必要である。

そこで、IPAdic 選択時に MeCab に送る命令文のフォーマットを IPAdic に準拠したものに变更し、IPAdic に対応した項目名と出力結果が表示されるように修正を行った。この時、IPAdic の出力項目は、IPAdic が持つ 9 つの項目を固定で出力することとした。修正後の Web 茶まめで IPAdic を用いた形態素解析を行った結果を図 3 に示す。

また、この修正に伴い、Web 茶まめの辞書・出力項目選択の GUI について、UniDic と IPAdic 利用時で出力項目が異なるために起きる、ユーザの混乱を避けるための変更を行った。具体的には図 4 に示すとおり、IPAdic を選択した場合には、出力項目チェックボックスを全て外した状態にし、チェックができないようにした。

辞書	文境界	書字形(=表層形)	語彙素	語彙素読み	品詞	活用型	活用形	発音形出現形
IPAdic(現代語)	1	given index is out of range						

図 2 不具合修正前の IPAdic 解析結果

辞書	文境界	書字形(=表層形)	読み	発音	原型	品詞	品詞細分類1	品詞細分類2	品詞細分類3	活用型	活用形
IPAdic(現代語)	B	庭	ニワ	ニワ	庭	名詞	一般				
IPAdic(現代語)	1	に	ニ	ニ	に	助詞	格助詞	一般			
IPAdic(現代語)	1	は	ハ	ワ	は	助詞	係助詞				
IPAdic(現代語)	1	二	ニ	ニ	二	名詞	数				
IPAdic(現代語)	1	羽	ワ	ワ	羽	名詞	接尾	助数詞			
IPAdic(現代語)	1	鶏	ニワトリ	ニワトリ	鶏	名詞	一般				
IPAdic(現代語)	1	が	ガ	ガ	が	助詞	格助詞	一般			
IPAdic(現代語)	1	いる	イル	イル	いる	動詞	自立			一段	基本形
IPAdic(現代語)	1	。	。	。	。	記号	句点				

図 3 不具合修正後の IPAdic 解析結果

辞書選択

現代語
 現代語話し言葉
 旧仮名口語
 近代文語
 近世口語(洒落本)
 中世口語(狂言)
 中世文語(説話・随筆)
 中古和文
 上代(万葉集)
 IPAdic(現代語)

出力項目※IPAdic選択時のみ項目は固定されます

語彙素
 語彙素読み
 品詞
 品詞-大分類
 品詞-中分類
 品詞-小分類
 活用型
 活用形
 発音形出現形
 仮名形出現形
 語種
 書字形(基本形)
 発音形(基本形)
 仮名形(基本形)
 語形(基本形)
 語頭変化型
 語頭変化形
 語頭変化結合型
 語末変化型
 語末変化形
 語末変化結合型
 アクセント型
 アクセント接続型
 アクセント修飾型

図 4 チェックボックス GUI の変更

3.4 2 辞書比較の解析の改良

Web 茶まめでは、辞書を 2 つまで選択し形態素解析を行い、解析結果を比較して見ることが出来る。この 2 辞書比較時の機能として 2 つの機能が存在する。1 つ目は、解析結果に差異が生じた行の文字を赤色に変更し発見しやすくする「色変更機能」である。2 つ目は、辞書毎に形態素が異なり出力結果の行がずれる場合に、行ずれを防ぐ為に空白行を挿入する「行ずれ修正機能」である。

しかし、これらの機能には、いくつかの不具合が発生していた。具体的な不具合は以下の 4 点である。実際に不具合が起こっている場合の例を図 5~7 に示す。

- (1)先頭行に必ず色変更機能が実行される。
- (2)差異が生じた行ではない場合でも、色変更機能が実行される場合がある。
- (3)英数字に対し、行ずれ修正機能が実行されない場合がある。
- (4)同一行で内容に差分があるときに色変更機能が実行されない場合がある。

Web 茶まめのプログラムを修正することにより、上記 4 点の不具合の修正を行った。その結果図 8~10 に示すように、正常に 2 辞書比較時の機能が実行されるようになった。

辞書	文境界	書字形(=表層形)	語彙素	語彙素読み	品詞	活用型	活用形	発音形出現形	仮名形出現形	語種	書字形(基本形)	発音形(基本形)	仮名形(基本形)	語形(基本形)	辞書	文境界	書字形(=表層形)	語彙素	語彙素読み	品詞	活用型	活用形	発音形出現形	仮名形出現形	語種	書字形(基本形)	発音形(基本形)	仮名形(基本形)	語形(基本形)
現代語	B	庭	庭	ニワ	名詞-普通名詞-一般			ニワ	ニワ	和	庭	ニワ	ニワ	ニワ	現代話し言葉	B	庭	庭	ニワ	名詞-普通名詞-一般			ニワ	ニワ	和	庭	ニワ	ニワ	ニワ
現代語	1	に	に	ニ	助詞-格助詞			ニ	ニ	和	に	ニ	ニ	ニ	現代話し言葉	1	に	に	ニ	助詞-格助詞			ニ	ニ	和	に	ニ	ニ	

図 5 (1)の不具合例

辞書	文境界	書字形(=表層形)	語彙素	語彙素読み	品詞	←→	辞書	文境界	書字形(=表層形)	語彙素	語彙素読み	品詞
現代語	B	「	「		補助記号-括弧閉	←→	近世口語(洒落本)	B	「	「		補助記号-括弧閉
現代語	I	ワールド	ワールド-world	ワールド	名詞-普通名詞-一般	←→	近世口語(洒落本)	I	ワールドカップ			感動詞
現代語	I	カップ	カップ-cup	カップ	名詞-普通名詞-一般	←→	近世口語(洒落本)	I				
現代語	I	」	」		補助記号-括弧閉	←→	近世口語(洒落本)	I	」	」		補助記号-括弧閉

図 6 (2)の不具合例

辞書	文境界	書字形(=表層形)	語彙素	語彙素読み	品詞	活用型	活用形	発音出現形	仮名出現形	語種	書字形(基本形)	発音(基本形)	仮名(基本形)	語形(基本形)	←→	辞書	文境界	書字形(=表層形)	語彙素	語彙素読み	品詞	活用型	活用形	発音出現形	仮名出現形	語種	書字形(基本形)	発音(基本形)	仮名(基本形)	語形(基本形)	
現代語	B	2	二	二	名詞-数詞					漢	二	ニ	ニ	ニ	←→	近世口語(洒落本)	B	2017			名詞										
現代語	I	0	ゼロ-zero	ゼロ	名詞-数詞		ゼロ	ゼロ	外0		ゼロ	ゼロ	ゼロ	←→	近世口語(洒落本)	I	年	年	ネン	名詞-普通名詞-助数詞可能			ネン	ネン	漢	年	ネン	ネン	ネン		
現代語	I	1	一	イチ	名詞-数詞		イチ	イチ	漢1		イチ	イチ	イチ	←→	近世口語(洒落本)	I	カレンダー			名詞											
現代語	I	7	七	ナナ	名詞-数詞		ナナ	ナナ	和7		ナナ	ナナ	ナナ	←→	近世口語(洒落本)	I	.	.		補助記号-句点									記号		
現代語	I	年	年	ネン	名詞-普通名詞可能		ネン	ネン	漢年		ネン	ネン	ネン	←→	近世口語(洒落本)	I	EOS														
現代語	I	カレンダー	カレンダー-calendar	カレンダー	名詞-普通名詞一般		カレンダー	カレンダー	外	カレンダー	カレンダー	カレンダー	カレンダー	←→	近世口語(洒落本)	I															

図 7 (3)(4)の不具合例

辞書	文境界	書字形(=表層形)	語彙素	語彙素読み	品詞	活用型	活用形	発音出現形	仮名出現形	語種	書字形(基本形)	発音(基本形)	仮名(基本形)	語形(基本形)	←→	辞書	文境界	書字形(=表層形)	語彙素	語彙素読み	品詞	活用型	活用形	発音出現形	仮名出現形	語種	書字形(基本形)	発音(基本形)	仮名(基本形)	語形(基本形)
現代語	B	底	底	ニワ	名詞-普通名詞一般		ニワ	ニワ	和底		ニワ	ニワ	ニワ	ニワ	←→	現代語	B	底	底	ニワ	名詞-普通名詞一般		ニワ	ニワ	和底	ニワ	ニワ	ニワ	ニワ	
現代語	I	に	に	ニ	助詞-格助詞		ニ	ニ	和に		ニ	ニ	ニ	ニ	←→	現代語	I	に	に	ニ	助詞-格助詞		ニ	ニ	和に	ニ	ニ	ニ	ニ	

図 8 (1)の不具合修正後の例

辞書	文境界	書字形(=表層形)	語彙素	語彙素読み	品詞	←→	辞書	文境界	書字形(=表層形)	語彙素	語彙素読み	品詞
現代語	B	「	「		補助記号-括弧閉	←→	近世口語(洒落本)	B	「	「		補助記号-括弧閉
現代語	I	ワールド	ワールド-world	ワールド	名詞-普通名詞-一般	←→	近世口語(洒落本)	I	ワールドカップ			感動詞
現代語	I	カップ	カップ-cup	カップ	名詞-普通名詞-一般	←→	近世口語(洒落本)	I				
現代語	I	」	」		補助記号-括弧閉	←→	近世口語(洒落本)	I	」	」		補助記号-括弧閉

図 9 (2) 不具合修正後の例

辞書	文境界	書字形(=表層形)	語彙素	語彙素読み	品詞	活用型	活用形	発音出現形	仮名出現形	語種	書字形(基本形)	発音(基本形)	仮名(基本形)	語形(基本形)	←→	辞書	文境界	書字形(=表層形)	語彙素	語彙素読み	品詞	活用型	活用形	発音出現形	仮名出現形	語種	書字形(基本形)	発音(基本形)	仮名(基本形)	語形(基本形)
現代語	B	2	二	二	名詞-数詞					漢	二	ニ	ニ	ニ	←→	近世口語(洒落本)	B	2017			名詞									
現代語	I	0	ゼロ-zero	ゼロ	名詞-数詞		ゼロ	ゼロ	外0		ゼロ	ゼロ	ゼロ	←→	近世口語(洒落本)	I	年	年	ネン	名詞-普通名詞-助数詞可能			ネン	ネン	漢	年	ネン	ネン	ネン	
現代語	I	1	一	イチ	名詞-数詞		イチ	イチ	漢1		イチ	イチ	イチ	←→	近世口語(洒落本)	I	カレンダー			名詞										
現代語	I	7	七	ナナ	名詞-数詞		ナナ	ナナ	和7		ナナ	ナナ	ナナ	←→	近世口語(洒落本)	I	.	.		補助記号-句点									記号	
現代語	I	年	年	ネン	名詞-普通名詞可能		ネン	ネン	漢年		ネン	ネン	ネン	←→	近世口語(洒落本)	I	年	年	ネン	名詞-普通名詞-助数詞可能			ネン	ネン	漢	年	ネン	ネン	ネン	
現代語	I	カレンダー	カレンダー-calendar	カレンダー	名詞-普通名詞一般		カレンダー	カレンダー	外	カレンダー	カレンダー	カレンダー	カレンダー	←→	近世口語(洒落本)	I	カレンダー				感動詞									

図 10 (3)(4) 不具合修正後の例

4. WebAPI の試作

4.1 目的

Web 茶まめは、煩雑な形態素解析環境を用意せずとも利用可能であるという特徴から、形態素解析を処理として含むシステムの一部として利用されつつある [6]。しかし、Web ブラウザ上での動作を前提とした形態素解析支援アプリケーションであり、他の Web サイトやアプリケーションが Web 茶まめの機能を利用・連携して動作するのは難しい。また、利用するシステムに適した GUI に変更することはできない、といった問題がある。

そこで、Web 茶まめの機能のみを簡単に利用可能にするための WebAPI を試作した。WebAPI の利用者は、HTTP 通信を用いて、Web 茶まめサーバに解析を行うテキストデータや出力項目、使用する辞書の情報等を送信し、形態素解析結果を受信する。これにより Web サイトやアプリケーションに Web 茶まめを用いた形態素解析機能を容易に組み込むことができ、形態素解析に必要なテキストデータの入力画面や出力方法などは、WebAPI を利用するシステムが自由に組み換え可能となる。

4.2 Web 茶まめの WebAPI 概要

今回、我々が作成した WebAPI は Web 茶まめの機能の 1 つである解析前処理を行う API (以下、解析前処理 API) と形態素解析を行う API (以下、形態素解析 API) の 2 種類である。解析前処理 API は、形態素解析を行う前に実行されるため形態素解析 API とは分ける形で作成した。

作成した 2 つの WebAPI はその名の通り、Web ブラウザ版 Web 茶まめの持つ解析前処理機能と形態素解析処理を、それぞれ利用するための WebAPI である。Web ブラウザ版 Web 茶まめとの相違点としては、以下の 3 点が挙げられる。

- (1) IPAdic を利用した形態素解析には非対応である。
- (2) 2 辞書を同時に選択し、解析結果を比較することはできない。
- (3) 出力形式は XML、及び JSON の 2 種類である。

(1) については、Web 茶まめでは UniDic を用いた解析がメインであるため、今回は IPAdic を用いた形態素解析は非対応とした。(2) については、結果の比較や表示は WebAPI を利用するシステム側が自由に設定する事項であるため、WebAPI 側では極力処理をしない方針としたため、実装を行わなかった。(3) については、HTTP 通信を用いたデータをやり取りする場合には XML と JSON が一般的によく使われ、適切であると言われるため、これらの形式を採用することとした。

これらの WebAPI は図 11 に示すように、外部のアプリケーション等から Web 茶まめサーバ内にある解析前処理 API または、形態素解析 API にそれぞれ対応した URL にアクセスすることで処理を実行することができる。実行する際には、解析対象のテキストデータや、利用する辞書の情報、実行する解析前処理の種類等を Web サーバに伝えるために、URL にクエリパラメータを付加した URL を用いる。

4.3 解析前処理 API

4.3.1 解析前処理 API の使用方法

Web 茶まめサーバの WebAPI を利用するための URL に、解析に必要な情報を付与しアクセスすることで利用可能である。試作した WebAPI では、解析前処理 API のリクエスト URL を「<http://lcm.tsu-lab.sie.dendai.ac.jp/V1/processings>」とした。この URL 以降に、WebAPI の実行に必要な情報をリクエストパラメータとして付加する。解析前処理 API のパラメータは

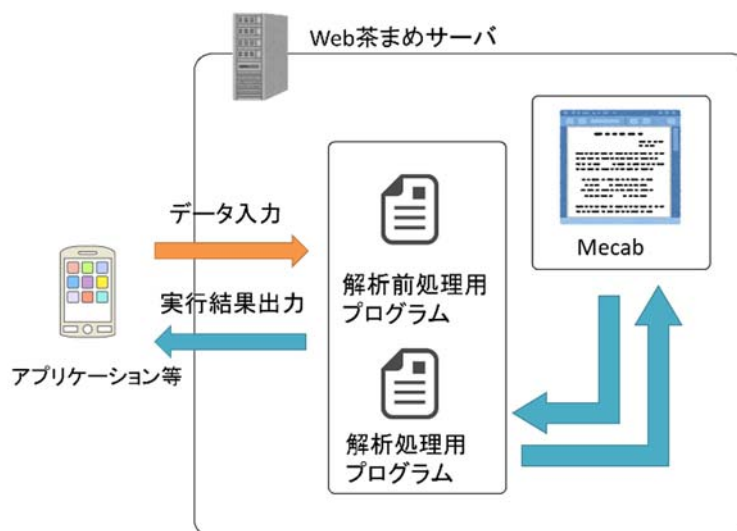


図 11 Web茶まめのWebAPI概要図

表 1 に示すように設定している。パラメータ名 `sentence` は、処理を行うテキストデータを入力する。ここでは URL で日本語を送信するためにパーセントエンコーディングを用いて入力する必要がある。パラメータ名 `processings` は、解析対象文章に行う前処理を選択する。前処理は複数同時に選択可能である。パラメータ名 `types` は出力結果を XML または JSON を入力し、出力形式を選択する。

表 1 解析前処理 API のパラメータ

パラメータ名	説明
<code>sentence</code>	解析前処理対象テキストを入力する。
<code>processings</code>	使用する出力項目を選択する。
<code>types</code>	出力形式を選択する。

4.3.2 解析前処理 API の出力結果

図 12 に、本 API を実行した場合の例を示す。この例では、日本語文「ゼンカク すゞ はんてん 1 2 3 4 5」に対して、解析前処理、「半角文字を全角に変換」「踊り字を展開」「カタカナひらがな反転」「数字処理」の 4 つを選択し、出力形式は XML とした。出力結果の XML タグがもつ意味は表 2 のようになっている。

```
<?xml version="1.0" standalone="true"?>
- <results>
  <zenkaku>ゼンカク すゞ はんてん12345</zenkaku>
  <odorizi>ゼンカク すゞ はんてん12345</odorizi>
  <hanten>ぜんかく スヅ ハンテン12345</hanten>
  <suuzi>ゼンカク すゞ はんてん一万二千三百四十五</suuzi>
</results>
```

図 12 解析前処理 API の XML 出力結果

表 2 解析前処理 API のタグ

フィールド名	説明
results	レスポンス結果
zenkaku	半角文字を全角に変換した文
odorizi	踊り字を展開した文
hanten	ひらがなカタカナを反転した文
suuzi	数字処理をした文
error	エラー結果
message	エラーメッセージ

4.4 形態素解析 API

4.4.1 形態素解析 API の使用方法

形態素解析 API も、解析前処理 API と同様に Web 茶まめサーバの WebAPI を利用するための URL に、解析に必要な情報を付与しアクセスすることで、利用可能である。試作した WebAPI では、解析前処理 API のリクエスト URL を「<http://lcm.tsu-lab.sie.dendai.ac.jp/V1/analyses>」とした。

形態素解析 API のリクエストパラメータは表 3 に示すように設定している。パラメータ名 sentence は、解析前処理 API と同様にパーセントエンコーディングを用いて入力する。パラメータ名 dics は、解析で使用する辞書を 1 つ選択する。選択は半角数字 1~9 のいずれかを入力する。それぞれに数字は、Web 茶まめが利用できる 9 つの辞書に対応しており、その順番は、Web ブラウザ版に表示されている順番である。例えば、「現代語」辞書は 1 に、「旧仮名口語」辞書は 2 に、「上代(万葉集)」辞書は 9 に対応している。パラメータ名 options_id は、出力項目を選択するためのパラメータである。選択は半角数字 1~24 の数字を、半角カンマ (,) 区切りで入力する。こちらも「語彙素」を表示したい場合は 1、「語彙素読み」を表示したい場合は 2 といったように、数字と項目は Web ブラウザ版に表示されている順番に対応している。例えば、「語彙素」「語彙素読み」「活用形」の 3 つを出力したい場合は「options_id=1,2,8」となる。パラメータ名 types は解析前処理 API と同様に XML または JSON を入力し、出力形式を選択する。

表 3 形態素解析 API のリクエストパラメータ

パラメータ名	説明
sentence	解析対象テキストを入力する。
dics	使用する辞書を 1 つ選択する。
options_id	出力項目を選択する。
types	出力形式を選択する。

4.4.2 形態素解析 API の出力結果

図 13 は、源氏物語の冒頭「いづれの御時にか、女御・更衣あまたさぶらひたまひける中に、いとやむごとなき際にはあらぬが、すぐれて時めきたまふありけり。」を解析テキストとして入力し、使用辞書に中古和文を選択、出力項目に語彙素・語彙素読み・品詞を選択、出力形式に JSON を選択した場合の出力結果である。

出力結果の JSON タグは、文境界と表層形、品詞-大分類、品詞-中分類、品詞-小分類、レ

スポンズ結果, エラー結果とエラーメッセージ以外の出力項目名は UniDic 内部の項目名に準拠している. 準拠していないタグについては, 文境界は `boundary`, 表層形は `surface`, 品詞の大分類, 中分類, 小分類はそれぞれ `pos1`, `pos2`, `pos3` とした. レスポンス結果, エラー結果, エラーメッセージは, 解析前処理 API の XML タグと同様のタグ名である.

5. おわりに

本稿では, 形態素解析ソフトウェア Web 茶まめの改良について報告を行った. ユーザから寄せられた情報を元に, 不具合修正や機能の改善を行い, Web 茶まめの利便性を向上させた. また, Web 茶まめの機能を他のアプリケーションから容易に利用可能とするために, WebAPI を試作した. 試作した WebAPI を用いることで, 他の教育システムや形態素解析を用いた研究システムに Web 茶まめの形態素解析機能を組みこむことが可能となった.

今後の課題として, 今後も寄せられるブラウザ版 Web 茶まめのユーザからの要望に答えるためにシステム改良を実施していく. また, WebAPI におけるパーセントエンコーディングを用いないデータ入力機構の実装や, WebAPI を用いた外部アプリケーションを試作し, WebAPI の機能を検証する必要があると考えている.

```
{
  "results": {
    "dictionary": "中古和文",
    "word_list": {
      "word0": {
        "boundary": "B",
        "surface": "いづれ",
        "lexeme": "何れ",
        "lForm": "イズレ",
        "pos": "代名詞"
      },
      "word1": {
        "boundary": "I",
        "surface": "の",
        "lexeme": "の",
        "lForm": "ノ",
        "pos": "助詞-格助詞"
      }
    }
  }
}
```

図 13 形態素解析 API の JSON 出力結果

参考文献

- [1] 小木曾 智信, 小町 守, 松本 裕治 (2013) 「歴史的日本語資料を対象とした形態素解析」『自然言語処理』20(5), pp. 727-748.
- [2] 国立国語研究所: 近代文 UniDic,(オンライン), <<http://www2.ninjal.ac.jp/lrc/index.php?UniDic%2F%B6%E1%C2%E5%CA%B8%B8%ECUniDic>> (参照 2015-09-15).
- [3] 国立国語研究所: 中古和文 UniDic,(オンライン), <<http://www2.ninjal.ac.jp/lrc/index.php?UniDic%2F%C3%E6%B8%C5%CF%C2%CA%B8UniDic>> .
- [4] 堤 智昭, 小木曾 智信: 歴史的資料を対象とした複数の UniDic 辞書による形態素解析支援ツール『Web 茶まめ』, じんもんこん 2015 論文集, Vol.2015, NO.1, pp.179-184,(2015).
- [5] 工藤 拓: MeCab: Yet Another Part-of-Speech and Morphological Analyzer, (online), 入手先 <<http://mecab.sourceforge.net/>> (参照 2015-09-15)
- [6] 土山 玄: 絵入源氏物語のテキストデータに対する統計解析 web アプリケーションの設計, じんもんこん 2016 論文集, Vol. 2016-CH-112, NO.1, pp.1-4, (2016).