

WordMiner Recipe 構成要素変数の生成と編集

ここでは、構成要素変数の生成や編集に関わる操作のヒントをレシピ (WordMiner Recipe) としてまとめています。

詳細についてはマニュアルを参照してください。

内容は断りなく改変いたします。

新規追加・改変内容は、WordMinerのホームページに適宜アップしております。

<http://wordminer.comquest.co.jp/>

区分	Recipe ID	タイトル
構成要素変数の生成	RD101	原始変数を分かち書きし、新たな構成要素変数を生成する
	RD102	分かち書きされた原始変数から新たな構成要素変数を生成する
	RD103	構成要素変数を併合して、あらたな構成要素変数を生成する
	RD104	原始変数を併合して、あらたな構成要素変数を生成する
	RD105	分かち書き回避記号により、分かち書きを抑制する
	RD106	構成要素変数をエクスポートし、ファイル編集後、再生成する
編集辞書の作成	RD201	「構成要素の編集辞書の管理」から新規辞書を作成する
	RD202	「構成要素変数の編集」から新規辞書を作成する
	RD203	構成要素変数の編集に用いた辞書を修正・再適用する
	RD204	「構成要素変数の編集」で作成した辞書を別の編集で利用する
構成要素変数の編集	RD301	「構成要素一覧と検索」から構成要素の編集辞書(削除)を作る
	RD302	「構成要素一覧と検索」から構成要素の編集辞書(置換)を作る
	RD303	構成要素の編集辞書を別のプロジェクトに複写(コピー)する
	RD304	ファイルから構成要素の編集辞書(削除)を作る
	RD305	ファイルから構成要素の編集辞書(置換)を作る
	RD306	構成要素の編集辞書(置換)をエクスポート(ファイル出力)する
構成要素の編集	RD401	分かち書き結果を編集辞書を用いて編集(連結)する
	RD402	分かち書き結果を編集辞書を用いて編集(切断)する
構成要素編集処理	RD501	構成要素変数の編集の適用順序
	RD502	構成要素の編集辞書の適用順序
	RD503	構成要素の編集辞書の適用結果の確認
構成要素変数の応用	RD601	固有の構成要素からなる構成要素変数を作成する
	RD602	共通の構成要素からなる構成要素変数を作成する
	RD603	特定の構成要素からなる構成要素変数を作成する

原始変数を分かち書きし、新たな構成要素変数を生成する

Step1.
「変数情報の管理」>「変数の生成」から「構成要素変数を生成」の「原始変数を分かち書きし、新しい変数を生成」をチェック(選択)します。

変数管理番号	変数名	種類	文字種	有効サンプル数	無記
[V0003]	1番目に大切なもの	原始変数	その他	1036	
[V0004]	次に大切なもの	原始変数	その他	1002	
[V0005]	あなたにとって大切なもの	原始変数	その他	1036	
[V0006]	性別	原始変数	その他	1038	
[V0007]	性別・年齢区分	原始変数	その他	1038	

Step2.
構成要素を生成する
原始変数を選択します。

Step3.
分かち書き画面 (Happiness/AiBASE) で、「実行」を選択し、分かち書き・キーワード抽出を行います。

Step4.
「ファイル」から「分かち書き終了」を選択し、構成要素変数を生成します。

No.	SEQ	変数名
1	[00000001]	あなたにとって大切なもの 自分自身 (生活、趣味)
2	[00000002]	環境です。色々ありますが、交通が便利だと言っている人は少ない
3	[00000003]	自分を心配してくれる人がいること...一緒に歳をとって人間がいることが一番大切で、今まで考えたこともなかったもの(たとえば、車とか)と思っ
4	[00000004]	環境です。色々ありますが、空気がきれいとか、飲み水が安心して飲める
5	[00000005]	自己と配偶者の健康・経済的
6	[00000006]	情報(いわゆる情報、個人間の知り合い等々)

[狙い]

原始変数を分かち書きし、新たな構成要素変数を生成する。

[ポイント]

- (1)分かち書き処理を終了すると、分かち書き「<原始変数名>-分かち書き」とキーワード抽出「<原始変数名>-キーワード」の2つの構成要素変数が生成される。
- (2)分かち書き及びキーワード抽出方法を設定するオプションがある。
切断単位やキーワードの組み合わせ、切断回避記号設定を設定することができる。
- (3)分かち書き画面の「ファイル」から「プロパティ」を選択すると、サンプル数、総処理文字数、分かち書き数、キーワード数などの処理結果情報を得ることができる。

分かち書きされた原始変数から新たな構成要素変数を生成する

Step1.
「変数情報の管理」>「変数の生成」から「構成要素変数を生成」の「変数の種類を変更し、新しい変数を生成」をチェック(選択)します。

Step2.
構成要素を生成する原始変数を選択します。

変数管理番号	変数名	種類	文字種	有効サンプル数	無記入/空白数
[V0001]	F	原始変数	英数字	5	0
[V0002]	好きな本	原始変数	その他	5	0

[狙い]

既に分かち書きされた原始変数(構成要素の候補が空白で区切られている)から、新たな構成要素変数を生成する。

[ポイント]

- (1)区切り記号の空白は半角でも全角でも、どちらでもよい。
- (2)半角と全角による表記の異なりは、異なる構成要素となる。
(分かち書き処理により構成要素を生成する場合は全角に統一される。)
- (3)分かち書きを行わずに構成要素を生成するので、WordMinerのエクスポートデータを再編集した結果、あるいは他の手段により得た分かち書き結果を読み込み、構成要素変数を生成することができる。

[例]

生成された構成要素変数
(構成要素の一覧と検索)

構成要素番号	構成要素	文字列長	構成要素数	サンプル度数
1	F	1	2	2
2	スタイリッシュ・キッズ	12	1	1
3	さいはての二人	7	1	1
4	ばら色の人生	6	3	3
5	スタイリッシュ・キッズ	11	1	1
6	バイバイ	4	2	2
7	帰れぬ人びと	6	2	2
8	駆け回る少年	5	1	1
9	川べりの道	5	2	2
10	大統領のクリスマスツリー	12	1	1
11	葉桜の日	4	1	1

原始変数(入力データ)

構成要素変数を併合して、あらたな構成要素変数を生成する

Step1.
「変数情報の管理」>「変数の生成」から「構成要素変数を生成」の「構成要素変数同士を併合し新しい変数を生成する」をチェック(選択)します。

Step2.
併合する元の構成要素変数を選択(追加)します。

Step3.
構成要素変数の選択(追加)を確認して、「完了」します。
(新たに構成要素変数が生成されます。)

変数管理番号	変数名	種類	文字種	有効サンプル...
[V0008]	● 一番好きな料理-分かつ書き	構成要素変数	その他	1110
[V0019]	● 次に好きな料理-分かつ書き	構成要素変数	その他	1093

実行事項: 併合する構成要素変数の追加後、[完了]します。現在の該当数②
注意事項: 生成後の変数名は、「変数情報の確認(変数名変更・削除など)」を確認して下さい。

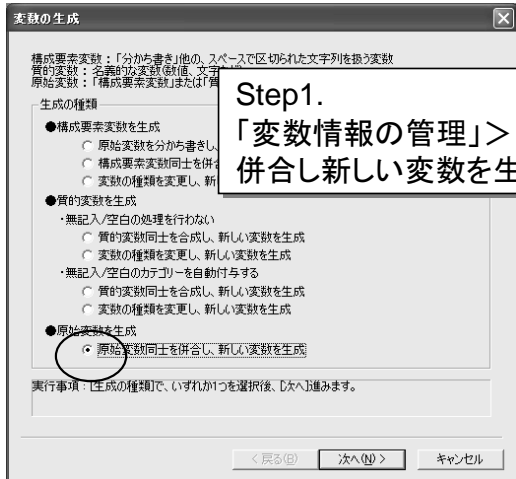
[狙い]

いくつかの構成要素変数を併合して、新たな構成要素変数を得る。

[ポイント]

- (1)例えば、「1番大切なもの(FA)」と「その他に大切なもの(FA)」の構成要素変数から、両構成要素変数を併合した新たな構成要素変数を得る。
- (2)併合された構成要素変数のデフォルトの名称は、併合した「構成要素変数名(1)+構成要素変数名(2)+・・・+構成要素変数名(n)」となる。
- (3)構成要素変数の併合により得られた構成要素変数は、即座に構成要素変数として利用できる。分かつ書き等、新たに構成要素変数の生成手続きを行う必要はない。

原始変数を併合して、あらたな構成要素変数を生成する



Step1.
「変数情報の管理」>「変数の生成」から「原始変数同士を併合し新しい変数を生成する」をチェック(選択)します。



Step2.
併合する元の原始変数を選択(追加)します。



Step3.
原始変数の選択(追加)を確認して、「完了」します。
(新たに原始変数が生成されます。)

Step4.
併合された(新たに生成された)原始変数を分ち書きし、構成要素変数を生成します。

[狙い]
いくつかの原始変数を併合して、新たな原始変数を得る。
この新たな原始変数から構成要素変数を生成する。

[ポイント]

- (1)例えば、「1番大切なもの(FA)」と「その他に大切なもの(FA)」の原始変数から、両原始変数を併合した新たな原始変数を得る。
- (2)併合された原始変数のデフォルトの名称は、併合した「原始変数名(1)+原始変数名(2)+・・・+原始変数名(n)」となる。
- (3)先に、それぞれの原始変数から構成要素を生成し、それらの構成要素変数を併合して新たな構成要素変数を得ることもできる。

分かち書き回避記号により、分かち書きを抑制する

Step1.

分かち書き画面 (Happiness/AiBASE)で、「オプション」選択します。

Step2.
「キーワード設定」(タブ)を表示し、「分かち書き回避記号」を設定します。

Step3.

「実行」を選択し、分かち書き・キーワード抽出を行います。

Step4.
「ファイル」から「分かち書き終了」を選択し、構成要素変数を生成します。

[狙い]

分かち書き回避記号を設定し、分かち書きを抑制し、任意の文字列を構成要素とする。

[ポイント]

- (1)回避記号で括られた文字列中の空白(全角・半角)は削除され、空白を詰めた構成要素が生成される。例)「A B C D」→「ABCD」
- (2)キーワードによる構成要素変数生成の場合、同時に組み合わせ設定が指定されていると、組み合わせ設定の最小語基が優先される。
- (3)回避記号が対になっていない場合、エラーとなり、そのサンプル(回答)から構成要素は生成しない。以降のサンプル(回答)の処理は継続して行うことができる。

構成要素変数をエクスポートし、ファイル編集後、再生成する

Step1.

データビューアを表示し、エクスポートする変数をチェックします。

The screenshot shows the 'データビューア' (Data Viewer) window. On the left, a list of variables is shown with checkboxes. Several variables are checked, including '管理番号', 'あなたにとって大切なもの', '自分にとって大切なもの', '性別', '生活余裕', 'お金の安心', '生活の安定', '家族音楽', '健康', 'お金の人間関係', '家族仕事', '福祉関係の仕事', and '趣味の時間'. A callout box points to the 'すべてエクスポート' (Export All) button. Below the list, a text box says 'この画面で右クリック' (Right-click on this screen). To the right, the 'エクスポート' (Export) dialog box is open, showing the save location as 'マイドキュメント' (My Documents) and the filename as 'データビューア.csv'. A callout box points to the '保存' (Save) button. A third callout box says 'Step3. 出力ファイル名を設定して、保存します。' (Step 3. Set the output filename and save).

Step4.

出力したファイルを適当なエディタで開き、構成要素変数を編集します。

The screenshot shows the Microsoft Excel application with the exported 'データビューア.csv' file open. The data is displayed in a table with columns A through E. The first few rows are:

管理番号	あなたにとって大切なもの	自分にとって大切なもの	あなたにとって大切なもの-キーワード
130	9978	自分の時間 健康	自分の時間 健康
803	9962	健康 家族 有意義な暮らし	健康 家族 有意義な暮らし
129	9950	生活の余裕 お金と心の安定	生活の余裕 お金と心の安定
475	9937	家族 音楽	家族 音楽
28	9936	健康に暮らせること、お金の人間関係	健康に暮らせること、お金の人間関係
802	9932	家族仕事 = 福祉関係の仕事なので、利用者満足してもらえるサービスを用意すること。趣味の時間	家族仕事 = 福祉関係の仕事なので、利用者満足してもらえるサービスを用意すること。趣味の時間
474	9876	家族 お金、平和、自由、雇用	家族 お金、平和、自由、雇用

 A callout box points to the table with the text 'Step5. WordMinerで、新たにファイルを読み込み、「分かち書きされた原始変数から新たな構成要素変数を生成する」方法により、構成要素変数を生成します。' (Step 5. Use WordMiner to load the file and generate new constituent variables from the original variables that have been split into words, using the method of generating constituent variables from original variables that have been split into words, you can generate constituent variables.)

[狙い]

WordMinerで生成した構成要素変数をエクスポートし、外部のファイルエディタ等で修正した後に、分かち書き済のデータとして読み込み、構成要素変数を生成する。

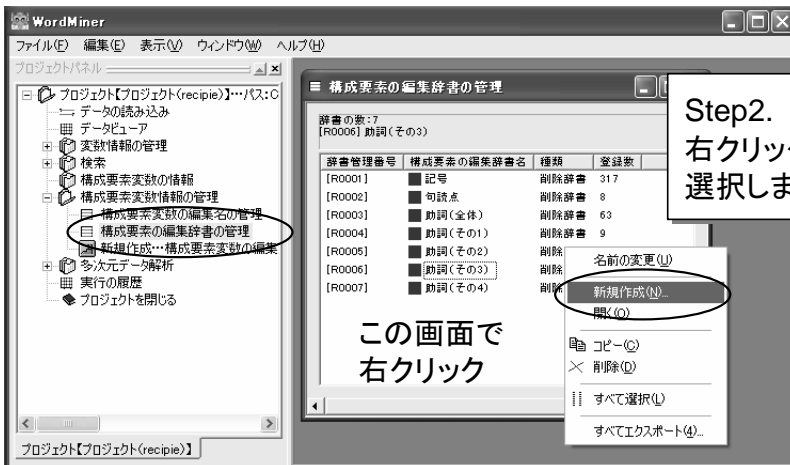
[ポイント]

- (1) データビューアから変数を選択し、エクスポートする。サンプル数かつ順序が保証される場合、変数の追加読み込みも可能だが、新たに全てのデータ(原始変数)を読み込むことを前提に必要な変数をすべて選択し、エクスポートすることが望ましい。
- (2) 出力した構成要素変数は構成要素が半角空白で区切られている。
- (3) 空白(半角・全角)で区切られた原始変数は、「分かち書きされた原始変数から新たな構成要素変数を生成する」方法により、構成要素変数を生成することができる。

「構成要素の編集辞書の管理」から新規辞書を作成する

Step1.
プロジェクトパネルから「構成要素の編集辞書の管理」を開きます。

構成要素変数情報の管理 > 構成要素の編集辞書の管理



Step3.
画面に従い、辞書の種類(置換辞書、または、削除辞書)と辞書名を設定します。



Step4.
辞書の内容を登録します。

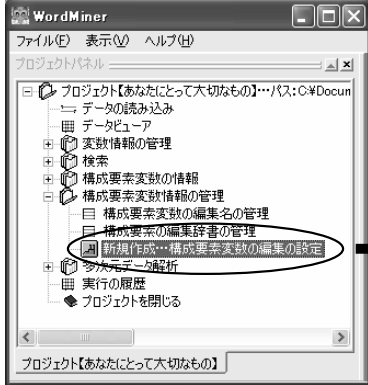
[狙い]
プロジェクト内で共通して利用できる編集辞書を作成する。

[ポイント]
(1)「構成要素の編集辞書の管理」から作成した辞書は、プロジェクトにおける構成要素の編集処理で共通して利用することができる。
(2) ただし、この辞書作成以前に編集した変数は、編集時の状態で専用保存された辞書を利用するため、編集処理を再実行する際に、ここで作成した新しい辞書は選択できない。
(3)辞書内容の登録方法については「構成要素変数の編集」についての説明を参照。

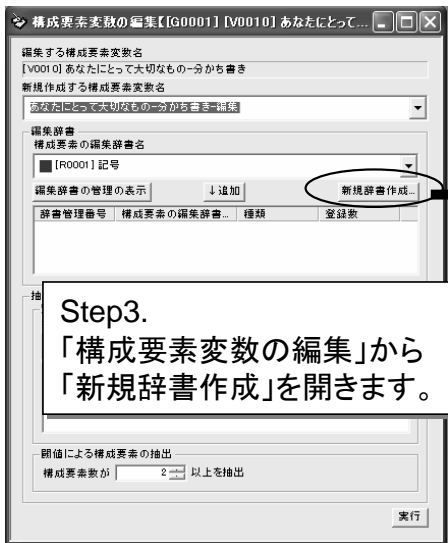
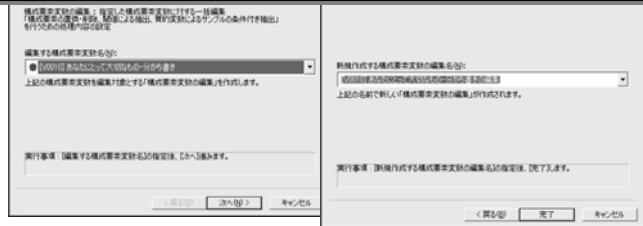
「構成要素変数の編集」から新規辞書を作成する

Step1.
プロジェクトパネルから「新規作成・・・構成要素変数の編集の設定」を開きます。

構成要素変数情報の管理 > 新規作成・・・構成要素変数の編集の設定

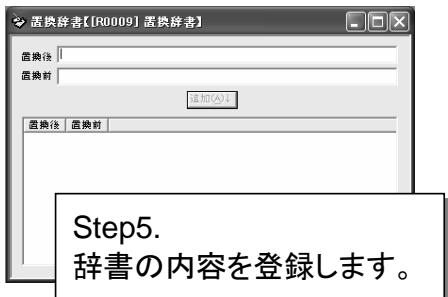


Step2.
画面に従い、「編集する構成要素変数名」と「新規作成する構成要素変数の編集名」を設定します。



Step3.
「構成要素変数の編集」から「新規辞書作成」を開きます。

Step4.
画面に従い、辞書の種類(置換辞書、または、削除辞書)と辞書名を設定します。



Step5.
辞書の内容を登録します。

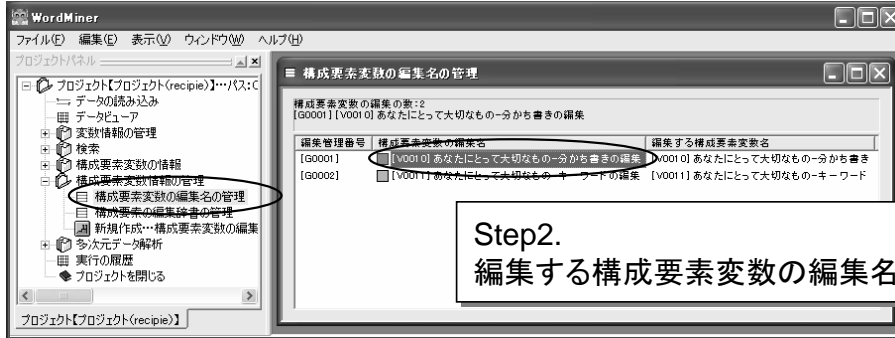
[狙い]
特定の構成要素変数の編集処理でのみ利用できる辞書を作成する。
[ポイント]
(1)「構成要素変数の編集」から作成した辞書は、その編集処理専用の辞書として保存される。
(2)プロジェクトパネルや他の編集名から「構成要素の編集辞書の管理」を開いても辞書は表示されないため、特定の編集手順以外の要因で修正されることがない。
(3)編集した変数を再編集する場合は、プロジェクトパネルから「構成要素変数の編集名の管理」を選択し、ここで(Step2.)設定した「構成要素の編集名」を選択する。
(4)辞書内容の登録方法については別途参照。

構成要素変数の編集に用いた辞書を修正・再適用する

Step1.

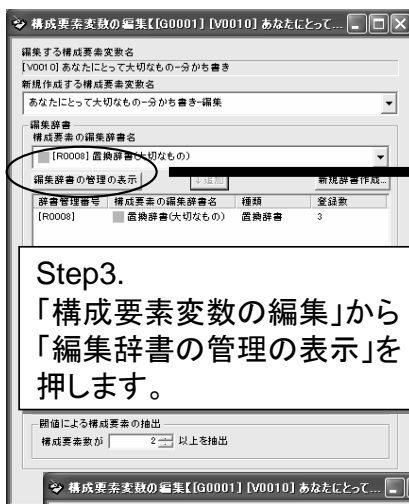
プロジェクトパネルから「構成要素変数の編集名の管理」を開きます。

構成要素変数情報の管理 > 構成要素変数の編集名の管理



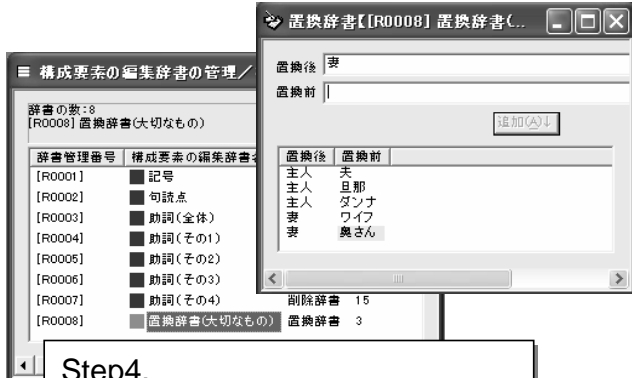
Step2.

編集する構成要素変数の編集名を選択します。



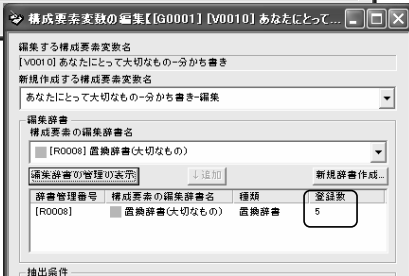
Step3.

「構成要素変数の編集」から「編集辞書の管理の表示」を押します。



Step4.

修正したい辞書を開き、内容を修正して閉じます。



Step5.

修正した辞書の登録数が変わっていることを確認し、「実行」で編集処理を実行します。

[狙い]
辞書を編集して、構成要素変数の編集処理を再度実行する。

[ポイント]

- (1)「構成要素変数の編集」から辞書を修正し、修正した内容で編集処理を再実行できる。
- (2)構成要素変数の編集時に専用保存された辞書を修正するため、修正内容は他の編集処理で選択する辞書には反映されない。
- (3)これらの構成要素変数の編集内容は「編集名」で管理されている。

作成した辞書を別の編集で利用する

ここでは、「構成要素の編集辞書の管理」で作成した辞書を、既に1度は編集した構成要素変数の編集に利用する例を示します。

Step1.

「構成要素変数の編集」から「編集辞書の管理の表示」を押します。

The screenshot shows the WordMiner interface with several windows and callouts:

- Step1:** A callout points to the '編集辞書の管理の表示' button in the '構成要素変数の編集【G0001】' window.
- Step2:** A callout points to the '構成要素の編集辞書の管理' window, which displays a table of dictionaries.
- Step3:** A callout points to the 'コピー' option in the context menu of the dictionary table.
- Step4:** A callout points to the '貼り付け' option in the context menu of the '構成要素の編集辞書の管理' window.

辞書管理番号	構成要素の編集辞書名	登録
[R0001]	記号	817
[R0002]	句読点	8
[R0003]	助詞(全体)	63
[R0004]	助詞(その1)	9
[005]	助詞(その2)	20
[006]	助詞(その3)	27
[007]	助詞(その4)	15
[008]	置換辞書(共通)	3

[狙い]

異なる編集処理で作成・利用した辞書を、他の編集処理で利用する。作成した編集辞書を他の編集処理に利用するためにコピーする。

[ポイント]

- (1)別の「構成要素変数の編集」に、辞書をコピーすることができる。
- (2)特定の変数の編集処理において作成、修正した辞書を、他の変数の編集処理に流用することができる。
- (3)「構成要素の編集辞書の管理」画面間において、辞書のコピー・貼り付けが行える。