



先輩～、急に呼び出さないでくださいよ～！また部長から商品企画のためのデータ分析を任せちゃって忙しいんですから！！



ごめんごめん。テキストマイニングの件で伝えておきたい最新情報があったんだけど・・・、じゃあ失礼しようかな。



最新情報？！そ、それなら早く教えてください！！



では、分析のお邪魔をしてはいけないので手短に。  
テキストマイニング入門 KH Coder と Excel でわかるデータ分析(以下、「本書」)は、執筆時点における KH Coder の最新版「**Version 3. Alpha. 13**」の環境で解説されている。ただ、ツールは進化・改良されていくものなので、KH Coder のバージョン更新に伴って、機能や操作に変更が生じる場合があるんだ。



まあそうですね。では、そういう場合の更新情報は、どうやって確認すればいいですか？



KH Coder のバージョンアップによる変更内容については、公式 web サイトの「更新履歴」(<http://kncoder.net/versions.html>)に掲載されているので、KH Coder の最新版をインストールする場合は内容を確認しておいてほしい。以上！では失礼。



ええっ？！それを伝えるためだけに私を呼び出したんですか？！せめて、現時点でのバージョン更新の内容だけでも教えてくださいよ～～！



では、2019/3/20 現在のバージョンアップ情報の中で、本書の内容と関係のある変更点について解説しておこうかな。



ほっ。さすが先輩！よろしくお願いします！

## 「Excel ファイルの保存について」

- ・前処理後に Excel ファイルを更新(上書き保存)し、KH Coder の「既存」プロジェクトを更新する

**Before** 「Version 3. Alpha. 13」の場合、

Excel や CSV ファイルを分析対象とした場合、「前処理」を実行した後に元データを編集しても、それらの変更内容を KH Coder に反映させることはできませんでした。

また、前処理を既におこなったファイルと同じ名前のもを分析対象ファイルとして新規プロジェクトを作成しようとするとき注意のメッセージが表示されるため、前処理を実行した後に Excel ファイルを編集した場合は別の名前をつけて保存し、再度 KH Coder で新規プロジェクトを作成する必要がありました。

- ・本書該当ページ:p.185 (下から 5 行目～)

※KH Coder で一度“前処理の実行”をおこなったデータに対して修正や追加をした場合は、“上書き保存”ではなく、別の名前をつけて保存してください。その後、新規に名前をつけて保存したファイルを使って KH Coder の新規プロジェクトを作成します(Excel ファイルを上書き保存しても KH Coder の分析結果には反映されませんし、前処理を既におこなったファイルと同じ名前のファイル名を KH Coder では重複して使うことができません)

- ・本書該当ページ:p.189 (下から 10 行目～)

先輩:「KH Coder で“前処理”が済んだあとに Excel のデータの内容を変更して上書き保存しても、KH Coder には反映されないんだ。前処理を一度おこなった後の Excel データを修正した場合は、上書き保存ではなく別の名前をつけて保存し、再び KH Coder で新規プロジェクトを作成する必要がある。」

舞:「一旦プロジェクトを作成した後に Excel の内容を変更しても KH Coder には反映されないってこと、しっかり覚えておきます！」



**After** 「Version 3. Alpha. 14」以降の場合、

KH Coder での前処理を終えた後に Excel や CSV の元ファイルを編集して上書き保存した場合、「既存プロジェクト」で前処理を再度実行することにより変更内容の読み込みが可能となりました。

## 操作例:

- ① KH Coder での前処理を終えた後、元の Excel ファイルに何らかの変更を加えたら、  
(ここでは、見た目で見やすいように、テキストの内容ではなく見出しの背景色を変えます(本質ではありませんが...))。

Before

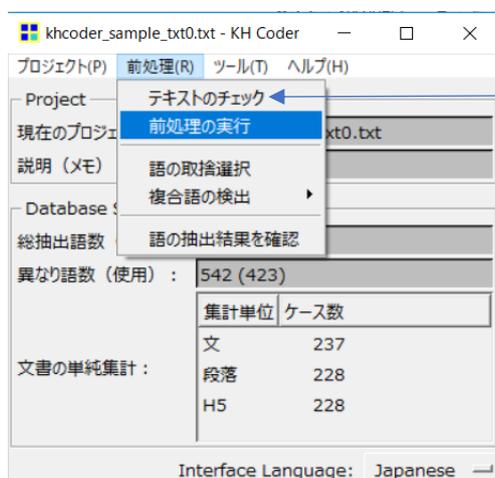
	A	B	C	D	E
1	評価理由	性別	年代	ブランド	総合満足度
2	健康管理に役立ち、コスパが良いので	女性	30代	A社	4
3	小型で軽量なので、時計として付けやすい。	女性	40代	A社	4
4	こまめに充電しなくてはならない点が不便	女性	40代	A社	2
5	まずは健康管理の機能が欲しかったので、そこが満たされているので満足	女性	50代	G社	5
6	スマホと連携してサクサクと快適に動作している	男性	30代	S社	4
7	機能が豊富なので、生活や仕事で役立っている	男性	30代	A社	4
8	使いやすい点は満足しているが、色のバリエーションが少ない	女性	30代	G社	2
9	直観的に選んだので、使いこなせていない	女性	30代	C社	3

- ② 変更後のファイルを上書き保存します。

After

	A	B	C	D	E
1	評価理由	性別	年代	ブランド	総合満足度
2	健康管理に役立ち、コスパが良いので	女性	30代	A社	4
3	小型で軽量なので、時計として付けやすい。	女性	40代	A社	4
4	こまめに充電しなくてはならない点が不便	女性	40代	A社	2
5	まずは健康管理の機能が欲しかったので、そこが満たされているので満足	女性	50代	G社	5
6	スマホと連携してサクサクと快適に動作している	男性	30代	S社	4
7	機能が豊富なので、生活や仕事で役立っている	男性	30代	A社	4
8	使いやすい点は満足しているが、色のバリエーションが少ない	女性	30代	G社	2
9	直観的に選んだので、使いこなせていない	女性	30代	C社	3

- ③ KH Coder で既存プロジェクトを開き(既にプロジェクトが開かれている場合はそのまま OK)、メニューの「前処理」→「前処理の実行」をクリックします。



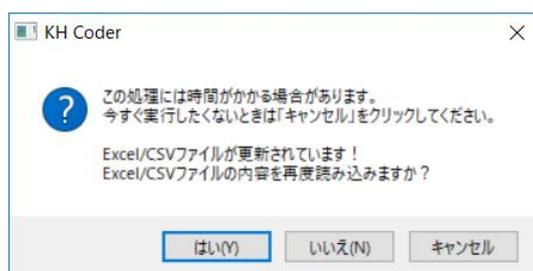
Version 3. Alpha. 15 以降のメニューでは、「分析対象ファイルのチェック」が「テキストのチェック」に表記変更されています。機能そのものに違いはなく、表記のみの変更です。

- ④ 元のファイルが更新されている場合、Excel/CSV ファイルの内容を再度読み込むかどうかを選択するダイアログが表示されますので、

更新後のファイルを読み込んで前処理をおこなう場合→「はい」

更新前の内容で分析を進める場合→「いいえ」

何も実行しない場合→「キャンセル」をクリックします。



“はい”を選択する(更新後のファイルを再読みして前処理をおこなう)と、最初の前処理で Excel/CSV ファイルを元に作られていたテキストデータ(“〇〇〇\_txt0.txt”)と外部変数データ(“〇〇〇\_var0.txt”)の内容も更新(上書き保存)されたうえで前処理が実行されるというわけですね。



そういうこと。“前処理”を再度実行することによって Excel の更新内容が KH Coder に反映されるようになったのは便利なことだが、元々の Excel データも、それを元に作られた KH Coder 読み込み用テキストデータも、全て上書き保存されるということなので気をつけてね。



?? 何に気をつければ良いですかね？



上書き保存で元データが無くなってしまった！！な～んてことがないよう、ご注意を。データのクレンジングをして分析の精度を上げていくことは必要だけれど、クレンジングを繰り返すたびにデータを“上書き保存”するということは、データ更新の“履歴”が残らず、元データの内容が再現できなくなってしまう可能性があるのです。



たしかに。元々のデータに立ち返ることも時々ありますものね。試行錯誤で段階的にデータ整形をしていると、「あ、やっぱり2つ前のクレンジング時のデータのほうが分析しやすかったかも・・・」と、気付くこともありますし。

Excel データを更新する時は、上書き保存ではなく“名前をつけて保存”で別の名前をつけてデータ更新の履歴が管理できるようにしていたほうが安心ということですね。



そうだね。例えば、ファイル名に連番を付けて Excel の分析対象データを保存しておけば、どの段階のデータなのかを番号で管理できるよね。

いずれにせよ、前処理をした後に更新した Excel データの内容が KH Coder の既存プロジェクトに反映できるという新機能は便利ではあるけれど、Excel の元データを更新する際は、そのまま Excel も KH Coder も上書きしても良いものかどうか(元データを再現したり、後から内容を確認したりする可能性の有無)を考慮してから実行するよう心がけておくといいね。



まあ、Excel を上書き保存してしまっても、また「新規」に KH Coder のプロジェクトを作成すれば、それが履歴管理になったりもするけどね。



えっ？ どういうことか具体的に教えてください！

#### ・前処理後に Excel ファイルを更新(上書き保存)し、KH Coder の「新規」プロジェクトを作成する



KH Coder の前処理を終えた後に Excel や CSV の元ファイルを編集して上書き保存しても、そのデータを使って新規に KH Coder でプロジェクトを作成することもできるよ。



さっき教えていただいた、Excel の更新に伴って KH Coder の**既存**プロジェクトを更新する方法と、新規プロジェクトを作成する方法との違いは何でしょうか？



**既存**プロジェクトの更新の場合は、前処理後に自動生成される分析用テキストファイルも更新(上書き)されるけれど、新規にプロジェクトを作ると、当然ながら分析用テキストファイルも新たに作成される。なので、“Excel データを上書き保存してしまった＝更新前のデータは再現できない”という状態であっても、KH Coder で生成される“テキストデータ”と“外部変数データ”は更新前後いずれのファイルも存在するということ。

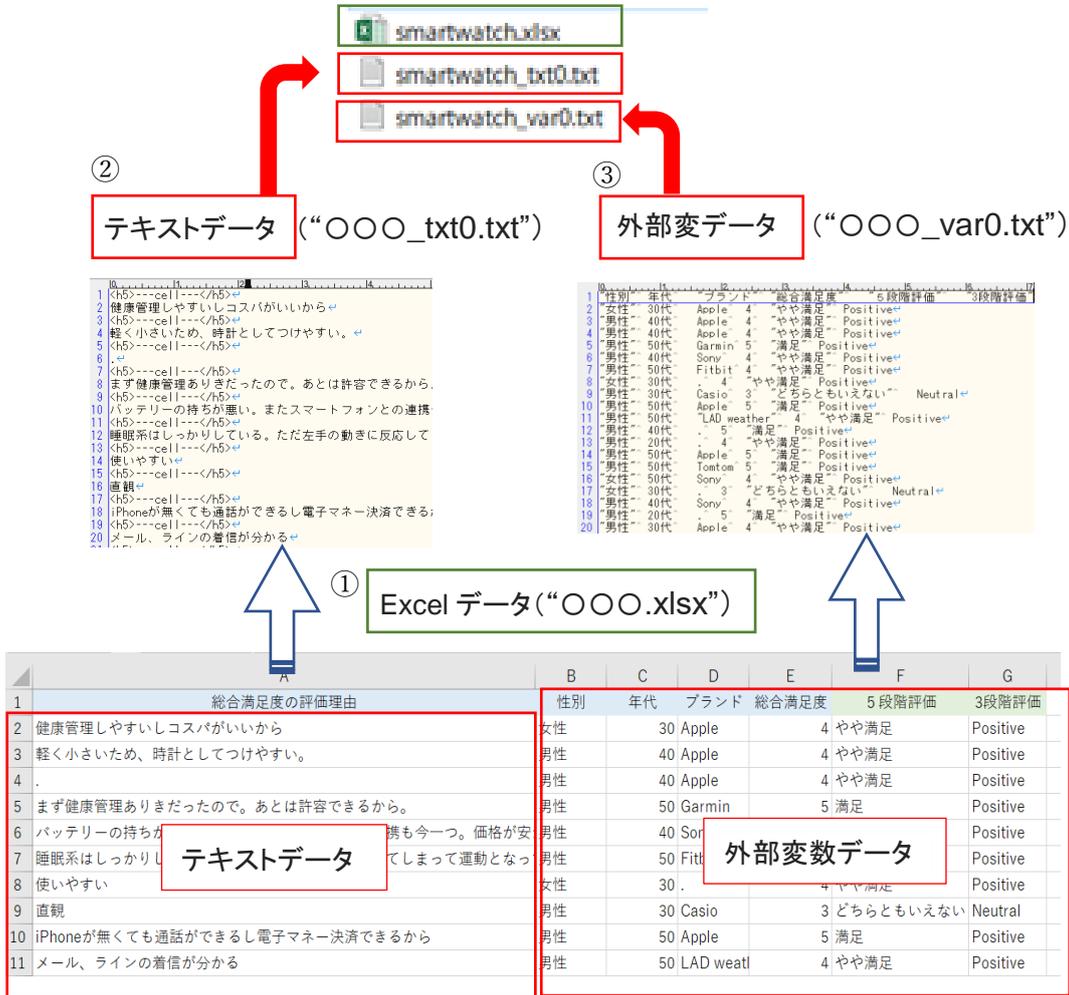


念のため確認ですが、“テキストデータ”と“外部変数データ”って、元データである Excel ファイルの保存場所の近くに自動生成される“〇〇〇\_txt0.txt”と“〇〇〇\_var0.txt”のことですよね？



そうだよ。本書 p.190 の以下のような図で言うと、①の Excel ファイルは更新(上書き保存)されたとしても、②と③のテキストファイルは上書きされず、KH Coder で**新規**にプロジェクトを作成する度に新たに自動生成される。そのとき、それぞれのファイル名の末尾の数字が自動的に連番で付けられるんだ。最初の前処理で作成されるテキストファイル

が”〇〇〇\_txt0.txt“だったとすると、2回目以降の前処理では”〇〇〇\_txt1.txt “、”〇〇〇\_txt2.txt “というように。



なるほど！Excel のファイル名が同一であっても、テキストファイルのほうは新たに連番が付けられた状態で作られるので、履歴管理ができるってわけですね。



そう。KH Coder でプロジェクトを新規作成するときに、「説明(メモ)」に更新内容やデータの状態などを書いてデータの管理をすれば間違いないね。

プロジェクト・マネージャ

対象ファイル	言語	説明 (メモ)	ディレクトリ
chosaku_seirei.txt	日本語		C:/Users/mikis/On
chosaku_shorei.txt	日本語		C:/Users/mikis/On
chosaku_ketsugo.txt	日本語		C:/Users/mikis/On
ketsugo_chosaku.txt	日本語		C:/Users/mikis/On
khcoder_sample_txt0.txt	日本語	原文データ	C:/Users/mikis/Dr
khcoder_sample_txt1.txt	日本語	クレンジング済み	C:/Users/mikis/Dr
khcoder_sample_txt2.txt	日本語	表記揺れ統合データ	C:/Users/mikis/Dr
sample_khcoder_txt0.txt	日本語		C:/Users/mikis/Dr

削除      新規      編集      開く      キャンセル