

0. Introduction

The whole purpose of this utility is to collect unknown words or phrases from various input sources in the first step, and then process the collected items up to the point where they are ready to be either added to the Dragon vocabulary or exported to an external TXT file, for further reference.

To that end, the utility has been designed as a fully integrated solution, which means that the whole functionality which is required for the task is available directly in the UI.

However, to understand the workings of the utility let's proceed by defining some terms which are fundamental in this context first.

An *unknown word*, by its definition, is any written form, or any combination of a written and a spoken form, which is not already in the active vocabulary of a given Dragon user profile.

Important Note: Words deleted from the Dragon vocabulary are discerned as known words when searching for unknown words to add to the vocabulary.

A *written form* is any combination of letters, digits or symbols as a single entity, or any combination of multiple such entities.

A *spoken form* can consist of a single word or multiple words. Single letters, separated by spaces, are also allowed in the spoken form. According to a specification which was introduced in DNS 12, however, the spoken form may only contain alphabetic characters or space characters, therefore digits and symbols are not allowed in the spoken form any longer. A spoken form which is not compliant with this particular specification will be referred to as "illegal" further into the text.

The new specification also included a limit to the total amount of characters both in the written and the spoken form which may not exceed 115.

Therefore, when collecting unknown words according to the definitions and specifications outlined above, the utility has to compare potential candidates (a) against the currently loaded vocabulary and (b) against its own collection which has been already established during a session, at any given point whenever a new match is added to the collection.

As a consequence, the inner workflow of the application is to create its contents incrementally as long as you go on adding input to it unless you choose to start over within the session by resetting the UI via a button provided for this scenario.

This way, de-duplication is guaranteed throughout the entire process which means that every item added to the collection is (a) an unknown word when compared to the Dragon vocabulary, but is also (b) unique within the collection.

In fact, catering for ongoing, fully-automated de-duplication has been one of the main concepts in creating this utility, as well as keeping in line with the underlying specifications. This way, collecting new words and phrases should be managed with minimal effort.

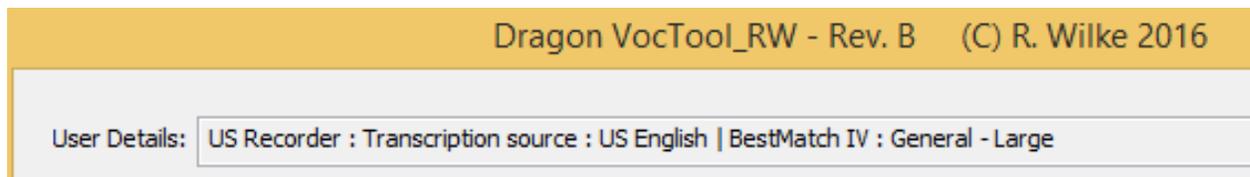
Technically, this has been achieved by turning the word collections into manageable objects based on the Dragon VocTools interfaces which provide all the means to do the hard work involved in it.

1. Getting Started

In order to meet the requirements specified above, it is mandatory that Dragon be running and a user profile be loaded in order to run the application. Therefore, when starting the application, the first thing it does is to check whether Dragon is running and a user profile is loaded. If these criteria are not met, a notification box will pop up and the application will terminate itself.

On the other hand, when closing Dragon first and leaving the application running, it will also terminate gracefully when shutting down, and not leave a stray natspeak.exe process dangling in the task manager, unlike many other third-party applications (sorry, I couldn't resist that).

So when starting the application successfully, for convenience, the name of the current user profile, along with additional information as to the current dictation source and the vocabulary, are displayed at the top of the UI.



With all of the above being said, it shouldn't be necessary to mention that any action taken within the UI from then on, and any results at end of processing the input to the UI, will be relative and relevant only with regards to the vocabulary in the currently loaded Dragon user profile.

Using the utility is executed in three steps basically:

Step 1: Collecting input from external sources to analyze for unknown words

Step 2: Reviewing and editing the input as well as choosing which items to add eventually.

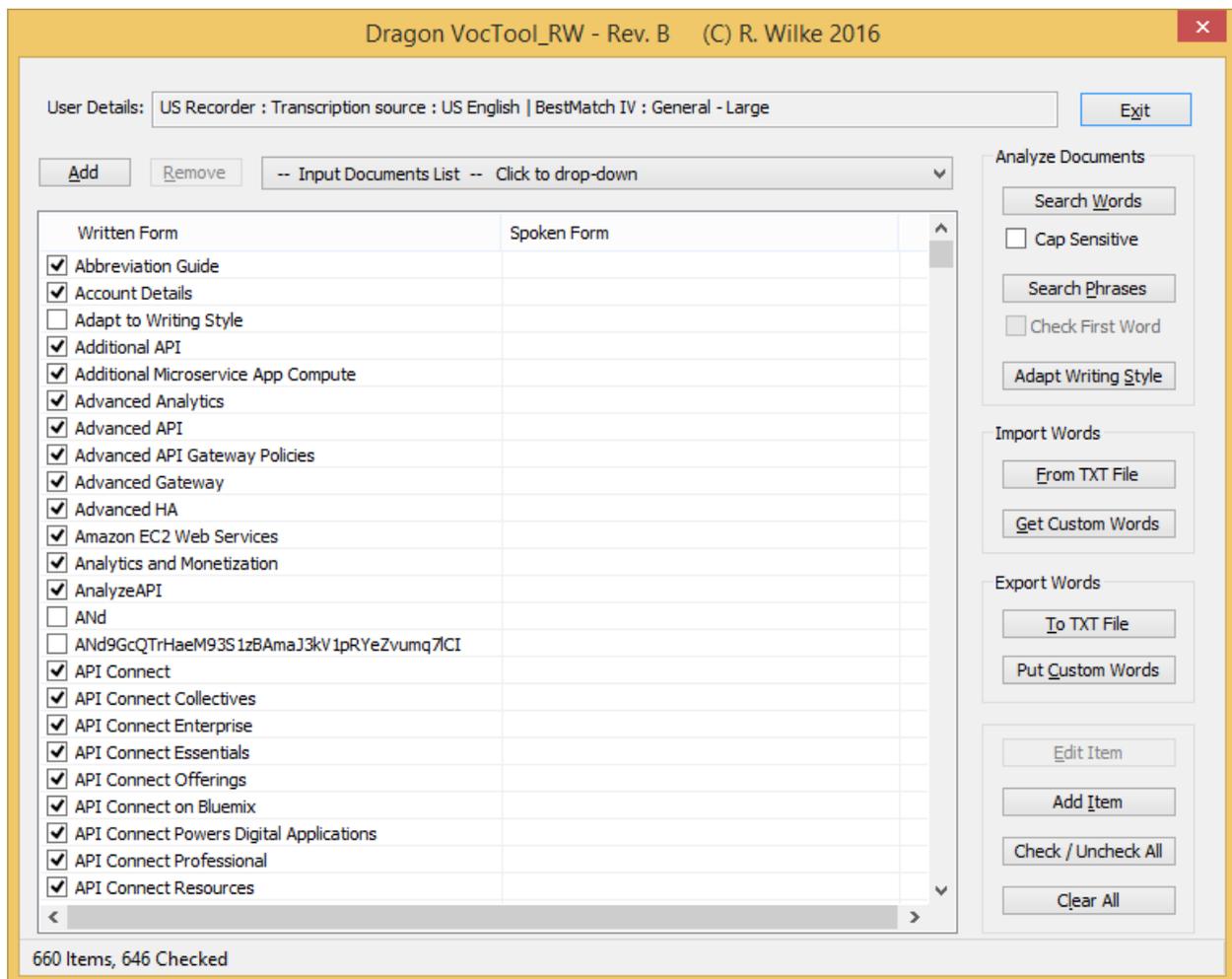
Step 3: Exporting the output so as to make it permanent, which the utility doesn't do for you.

It is very important to understand that running the utility does not interfere with the Dragon vocabulary in any way, and all the data collected during a session will be lost unless step three is executed one way or the other, please refer to the last section in this documentation for all the details.

2. Overview of the UI

As explained above, you might also think of the application as a machine which needs some input from you first to provide you some output at the end. There are various ways to get input into the UI, and to operate on the input, with a particular set of buttons for each, which I will explain below.

Throughout the process, the current results are displayed in the large list box in the center of the UI which will be populated automatically as you go along.



As we are working with words which have a written form and spoken form optionally, the list box has two columns to display them next to each other for each item in an individual row.

Additionally, after the list has been populated as in the screenshot above, there is a checkbox to the left of each row which can be used to select or unselect the particular item to either go into the output or be discarded at the end. By default, every item added to the list is checked automatically.

Additionally, whenever an item is added to the list box, the list will be sorted automatically, in alphabetical order.

Above and to the right of the list box there are numerous buttons which are used to handle getting the input, operating on it and providing the output which will be discussed below.

One thing to notice here, when exiting the utility, all the contents created during the session will be discarded and removed from memory, so make sure to save or export any critical data first before closing out.

3. Getting Input

Basically, there are two ways of getting input to operate on, via *document analysis*, or via *importing words*.

(a) Document Analysis

The method of document analysis has been long around within the Dragon UI, and this utility basically replicates the functionality provided within Dragon. To that end, you can add any number of document sources to a file list and then start analyzing them for unknown words.

As in Dragon, currently supported file formats are: TXT, RTF, DOC, DOCX, and HTML.

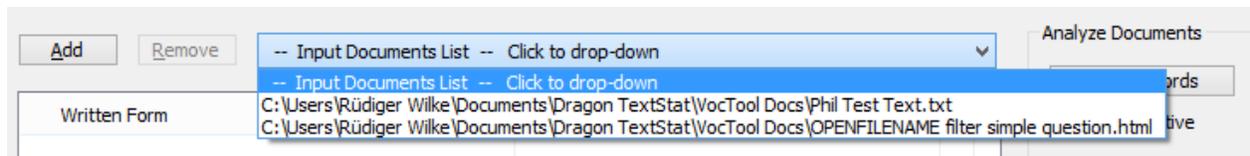
However, one of the goals in creating this utility was to enhance the capability of the Dragon document analysis by also searching for unknown phrases. To achieve this, regular expressions techniques have been implemented, but the implementation hasn't been completed yet, at least not at the time of writing this.

So, in order to prepare and start the document analysis functions, first go to the **Add** button on the left near the top of the UI, next to the drop-down list in the middle above the list box.

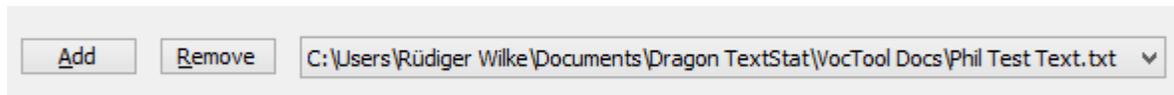


Pressing the button will start the typical open file dialog which lets you choose the documents you wish to analyze, and you can add them one by one. The routine takes care of documents not being added twice to the list itself.

The documents will be added to the drop-down list referred to above which you need to open to display them. Within the list, the fully qualified file paths are shown, and the list box has been designed to expand to the right automatically adjusted to the largest element in the list, if it exceeds the initial width.

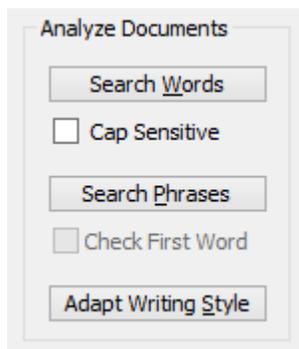


If you want to remove an item from the list, you first need to select it in the drop-down list which will also activate the **Remove** button, and then press the button. The item will be removed from the list without being prompted.



Once you are finished adding documents to the drop-down list for analysis, you can either search the documents for words or for phrases in two individual steps which may also be combined at a later stage of the development. Additionally, you can use the documents added to the drop-down list to adapt to writing style.

All the actions corresponding to and available for the documents currently in the drop-down list are in an individual section in the right hand panel labeled “Analyze Documents”.



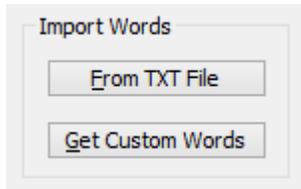
Pressing the **Search Words** or **Search Phrases** buttons will start analyzing the documents for unknown words or phrases, and searching the documents will take a few seconds to a few minutes depending on the amount of data to be searched through, during which the UI will be locked.

As explained above, you can always proceed and add to the input by analyzing further documents which will add to the input previously collected, incrementally. If you’re not satisfied with the results and want to start over, you can always press the **Clear All** button at the bottom right hand side of the UI which will delete all the current contents in the UI as well as all the buffers which have been created so far; please refer to the section “Overview of the UI” above.

At any time during the session, you can press the **Adapt Writing Style** button which will do as advertised, based on the currently selected documents in the drop-down list. However, as this will immediately modify the language model of the current user profile, a warning will be given when executing this function, and you will be prompted to confirm proceeding.

(b) Importing Words

Another way to get input into the UI is to import words from external files in TXT format which is typically used in Dragon to export or import custom words in. To do this, refer to the “Import Words” section in the middle of the right hand panel of the UI, and let’s start with **From TXT File**.



When editing an existing work list all when creating your own word lists and saving them in TXT files, you are well advised to take care of the integrity of the file encoding and format, for which the rules are:

- The encoding of the file should be ANSI, this is what the default encoding is when words are exported from the Dragon menu. UTF-8 may work also, but Unicode will be problematic, so better make sure that ANSI is being used.
- There are two viable formats for the content itself. In the previous format, written forms and spoken forms were separated by a single backslash, in a more recent format, they are separated by double backslashes, but then the file must be preceded by a header: @Version=Plato-UTF8.

Therefore, viable formats are, for instance:

```
@Version=Plato-UTF8
GatewayScripts10\\ gateway scripts ten
Advanced Developer Portal\\advanced developer portal
API Designer\\API designer
...
```

```
GatewayScripts10\gateway scripts ten
Advanced Developer Portal\advanced developer portal
API Designer\API designer
...
```

And make sure to be consistent and not to mix both formats as that will yield unwanted results. Also take care not to have spaces before and after the backslashes for the same reason.

Additionally, when editing spoken forms, pay attention to the specifications for the spoken form to avoid illegal characters as explained on page 1.

So when pressing the “From TXT File” button on the UI, and open file dialog will display with the father figure to said to TXT for obvious reasons, so you can choose the file to input from, one file at a time. The file will be analyzed and its contents will be added to the contents of the UI, again incrementally, as has been explained above.

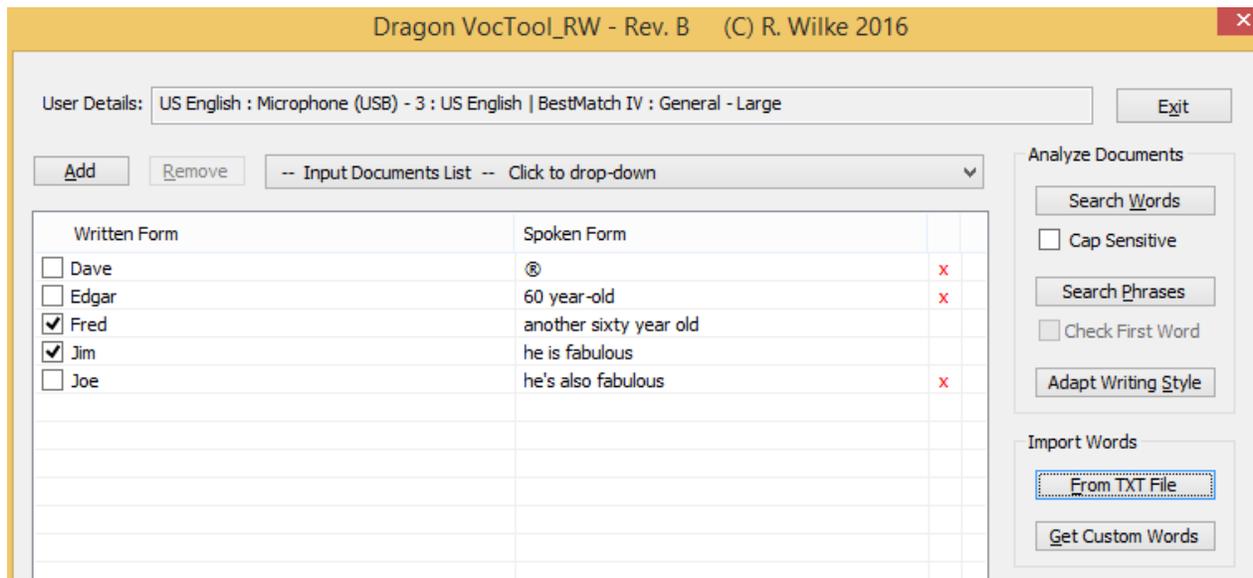
(c) Get Custom Words

Yet another way to get input into the UI is to press the **Get Custom Words** button in the same section on the right hand panel. This function has been added specifically per user request to facilitate retrieving the existing custom words in the current user profile for reviewing them.

The idea emanated from discussing the fact that the spoken forms for custom words may well have illegal characters if they have been imported from a list and had not been added manually. Obviously, the mechanisms inherent in the Dragon functions are not consistent here, or it may be due to allowing for illegal characters during a profile upgrade.

At first, the decision was to suppress such spoken forms in the UI, so as to make sure that the UI only contains fully valid items. But then the request was to keep them, also to facilitate reviewing and possibly editing them, if and when need arises.

Therefore another change went into the design which allows for illegal spoken forms on import, however, such items are not checked by default and they additionally receive a red cross at the right end of the row, as can be seen here:



Note that this routine has also been added to importing words from a TXT file, which has been executed in the screenshot above.

It is always possible to edit and legal spoken form to something legal using the methods which will be explained in a following section, and this will put a checkmark into the box and remove the red cross automatically.

However, given the very exceptional role of the “Get Custom Words” function within the scenario of the application as a whole, and as already existing custom words are not candidates to collect when searching for unknown words anyway, this is the only function where, when executed, any previous contents in the this box will be cleared first before displaying the custom words.

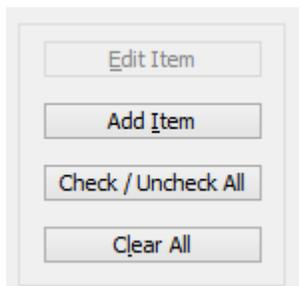
But a warning will be given accordingly.

4. Operating on Input

Anytime you have collected some input in this box you may want to review and edit it, and there are basically two functions available to do this:

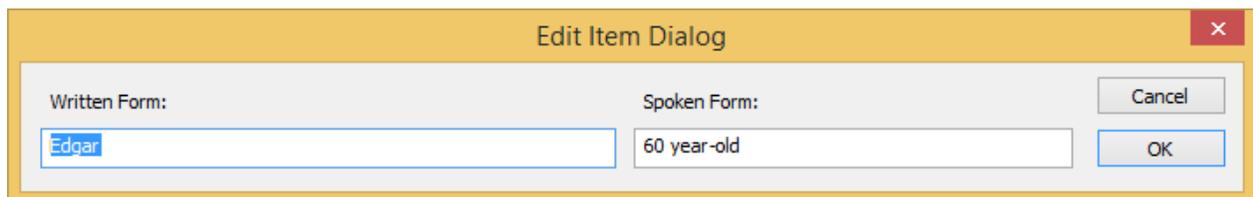
- the **Edit Item** button
- the **Add Item** button

both of which are available in the bottom section on the right hand panel in the UI.



The “Edit Item” button becomes active whenever a particular item in the list box is selected, and note how it isn’t activated in the screenshot extract above which was taken when no item was selected. It goes without saying that the list box only allows for single selection so it will always be clear which particular item to edit when the button is pressed. And please don’t confuse the “selected” with the “checked” state, the selected state of an item will be indicated by highlighting the entire row.

Now, since the list box UI element doesn’t allow for direct manipulating, such as editing into it directly, a sub dialog needs to be invoked in order to edit an item, which looks like this:



It is set up as a modal dialog, and the dialog box is positioned right beneath the main dialog, so as to keep it out of its view. In the box, there are fields for the written form and the spoken form, and the fields are automatically populated with the contents of the item you have chosen for editing.

You can override the contents in the fields and press the OK button to update the list box. Keep in mind though, the spoken form will be parsed for illegal characters and if such are found, updating the list box will be rejected.

When updating the contents in the list box, the previous item will be deleted first and the edited item will be inserted.

Alternatively, you can always cancel out of the dialog using the appropriate button.

The “Add Item” dialog looks much the same, except for not being populated, and the button to invoke it is always active of course.

Again, any potential input in the spoken form will be parsed for illegal characters first and rejected if such are found. If successful, the new item will be added to the list box and sorted in alphabetical order automatically.

Finally, there are two more buttons in this section which are there for convenience rather, namely the button to check or uncheck all items at once as well as the previously mentioned button to clear the entire contents collected so far.

As part of the editing phase during the session, you should of course also decide on which items to keep and which not to keep when saving or exporting the data eventually, and you do this by making sure that every item you want to keep does have a checkmark in the box.

5. Exporting the Output

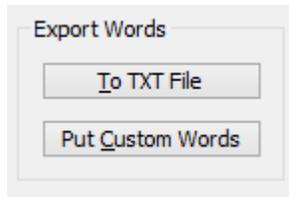
Once you’re finished preparing your output, by carefully reviewing and editing the items in the list box as well as checking or unchecking the items given which ones you want to keep and which ones you don’t want to keep you may probably want to make it permanent by saving it one way or the other.

Keep in mind that unless this is done, all the new words collected during the session will be lost as they only exist somewhere in the memory of your computer up to this stage.

Basically, there are two methods available in the UI to save the new words:

- Exporting the output to a TXT file, in the same way and format this is done inside Dragon
- Adding the new words as custom words straight to the Dragon vocabulary which is currently loaded

Both of these methods are under the “Export Words” section in the right hand panel of the UI.



Both of these methods are implemented like this. First, for every item in the list box which has a checkmark the spoken form is validated and if it doesn't contain any illegal characters, the item will be added to a collection for further processing. This is done to make sure that no invalid items are either saved or added to the vocabulary.

At the end of this step, the “To TXT File” function dumps the collected items to a file for which you are prompted to specify a location and a name under which it will be saved. The file encoding and format are compliant with the Dragon specifications.

This file can be used to import from into Dragon at a later stage, or even for further reviewing in the tool if you want to go back to it later on.

The “Put Custom Words” however will, just once more to double check, first compare the new words with the vocabulary before they are added to the vocabulary and immediately available in the user profile therefore.

Because of that, and since it will have an immediate impact on your user profile, you will be prompted to confirm that this action should be taken.

If you are done exporting or saving the new words in one of the two ways as described, you may safely exit the application, or just any other time, even if Dragon has been closed before.

Happy collecting!