Spellchecking in Perfion

Document version	1.2
Document date	2016-04-28
Created by	Zoran Kos Michael Engesgaard
Introduced in	Perfion 4.5.5966

Table of Contents

1. Spellchecking in Perfion – A short introduction

The spellcheck-feature in Perfion provides spellchecking capability inside the Item Editor. Spellchecking in Perfion works very much as it does in other modern word processors, like Microsoft Word, by highlighting misspelled words and underlining them with a red line, and supports choices like "Ignore", "Add to custom dictionary" and of course, it corrects misspellings.

In Perfion, since the Item Editor often is showing texts in multiple languages, the spellchecking works for all active user languages at the same time, that is, all languages enabled for the user and hereby available in the Item Editor.

The spellchecking feature in Perfion uses Open Office dictionaries to provide its service. In order to enable spellchecking, administrators and/or super users must upload one or more such dictionaries to Perfion.

Perfion stores dictionaries centrally in the database. This has the advantage, that you can centrally administrate which dictionaries are available. This even goes for custom dictionaries. Therefore, if your business are using some certain professional terms that are not included in a general dictionary for the language you are using, you can supply them as a list of words that Perfion also reckons as being correct. Users may, using the "Add to custom dictionary"-functionality added to the Item Editor, contribute to these centrally placed custom dictionaries, by adding words to them.

This document will first walk you through how you upload dictionaries and make them available for your users. Then it will show you how the end users will benefit from having spellchecking available.
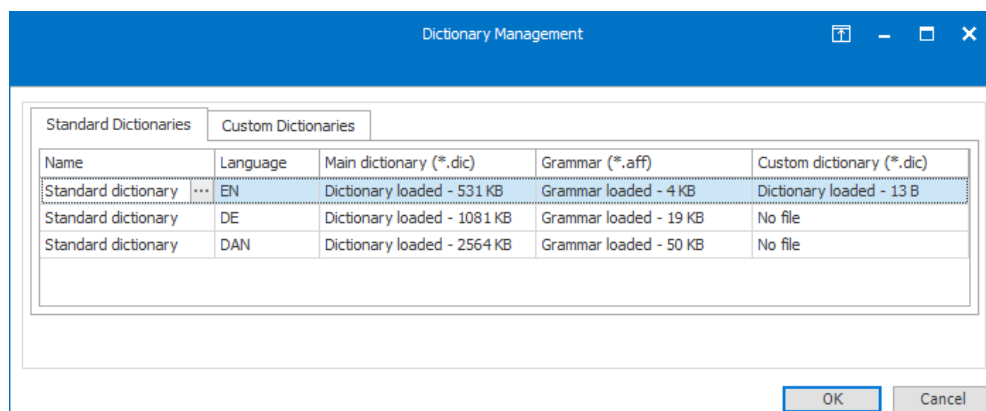
2. Managing dictionaries

To enable the spellcheck-feature you must first upload dictionaries for at least one language. As mentioned in the introduction Perfion keeps the dictionaries, including custom dictionaries, stored centrally in the database.

To administrate these dictionaries start the Perfion client and go to the Dictionaries-menu located in the Administration-tab in the Perfion Main Window in the Proofing-section:
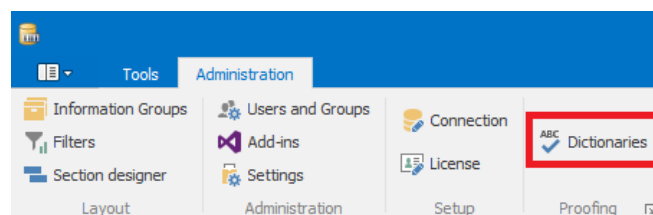
Note: This option is only available to Administrators and Super users.

Clicking this menu will take you to the "Dictionary Management"-dialog shown below:

This dialog has two tabs:

> **Standard Dictionaries:** Here you manage the Standard dictionaries. This is where you need to upload the Open Office-dictionaries containing a language-specific dictionary (English, German, Danish etc.).

> **Custom Dictionaries:** Here you manage the custom dictionaries, that is, dictionaries that allow you to upload specific sets of terms suiting the specific needs of your business.

Before diving into managing the dictionaries found in these tabs, it is important to understand the terms "Standard dictionary" and "Custom Dictionary".

2.1.   Standard- and Custom-dictionaries

A **Standard dictionary** in Perfion always has an Open Office-dictionary associated with it. This Open Office dictionary always for a single language (never no language, never multiple languages). Such a dictionary can be perceived as a whitelist of known, correctly spelled words, for that language. There is a bit more to it than just a basic list of words, since such a dictionary also takes the "rules of the language" (also called the grammar) into account. An Open Office dictionary consists of two files: The dictionary file that simply speaking contains the list of words in the language (file extension .dic) and the grammar-file containing the rules for the language (file extension .aff). Both these files are considered read only in Perfion, that is, they can be uploaded and they can be replaced by uploading another file, but the end user cannot modify them. The two Open Office dictionaries can be seen in columns "Main dictionary" and "Grammar" in the above dialog. More about uploading these dictionary-files in section 2.4.1.

As previously stated, the Open Office-dictionaries are considered read only, so to allow the user to add words, which are not in these Open Office-dictionaries, an additional "Custom dictionary" is created automatically. This dictionary is initially blank, but will be filled up by users adding words in the Item Editor by picking the "Add to dictionary"-menu available there. More about adding words to dictionaries in section 3. The custom dictionary being a part of the standard dictionaries is also called the **Default custom dictionary** for that language.

A (non-default) **Custom dictionary** in Perfion is a dictionary that may or may not belong to a language, but typically contains other words than those you would expect from a normal dictionary. Again it is basically a simple list of words, which are considered correctly spelled. While a non-default custom dictionary *can* be associated with a language, it does not need to. If the words it contains make sense in all languages, there should not be assigned a language to it. A custom dictionary serves three purposes:

1) As the default custom dictionary, it allows the user to add words to it (using "Add to custom dictionary" on the Item Editor).

2) It allows you to upload specific sets of terms suiting the specific needs of your business. As an example a custom dictionary containing all car brands would make sense in our Perfion demo database, so that brand names like Audi, BMW, Volvo and Peugeot are not considered misspellings. This is also an example of a custom dictionary making sense in all languages.

3) When creating more custom dictionaries each serving a special purpose you can handle these dictionaries as a whole. Consider a dictionary with brand names making sense in your business that you may want to edit in its entirety in a text editor and re-upload to Perfion when done. How this is done is described in section 2.6.

Notice the two important differences between the Standard-dictionaries (1st tab) and the non-default Custom dictionaries (2nd tab):

1. **Standard dictionaries** consist of two Open Office dictionary-files and a default custom dictionary. All these files describe words *in that particular language* (English, German, Danish etc.). The default custom dictionary is just one file and it simply contains a list of words (spelled correctly, of course) all of which should be normal words *in that language*.

2. While Standard dictionaries always represents a language, **Custom dictionaries** may or may not relate to a specific language. For example, "Legal terms" often relate to a specific language, so using a custom dictionary with a language would make sense. On the other hand, you can imagine a custom dictionary consisting of car brands. Such a dictionary is language independent, meaning that it will make sense for all languages.

   Consider again the custom dictionary for the Perfion standard database with car brand names. Such brand names are not likely to show up in Open Office-dictionaries, but it is annoying that they turn up as spelling mistakes for all users in all languages. While you could add these names to the default custom dictionary for all languages, you would not want to do that. First of all because you indeed would have to add them to all languages having them develop independently as different users may add different new brand names to each of them. By instead having a "Car brand names"-dictionary you have them collected in one dictionary that applies to all languages. And this dictionary can even be handled separately (for example edited in its entirety or even deleted for whatever reason that should be needed).

While the Open Office-dictionaries cannot be modified from within Perfion, custom dictionaries can indeed be updated by the end user (both the default custom dictionary and the non-default ones). When a user on the Item Editor in Perfion adds a word to dictionary, it will be added to exactly one

custom dictionary picked by the user. More on this in section 3.

Any custom dictionary can be created from scratch by using any text editor and start writing one word per line and finally uploaded to Perfion. More on this in section 2.5.1.

2.1.1. Central administration

As opposed to spellchecking features in most other text processors, Perfion stores its dictionaries centrally in the Perfion database as opposed to locally on each machine. This has the advantage that an administrator or super user can administer dictionaries centrally. Additionally it has the benefit that words added to a custom dictionary by one user will benefit all other users.

2.2. Standard Dictionaries tab

Going back to the "Dictionary Management"-dialog. When the "Standard dictionaries"-tab is shown for the first time, it contains a table with exactly one row for each language the logged-in user has access to.
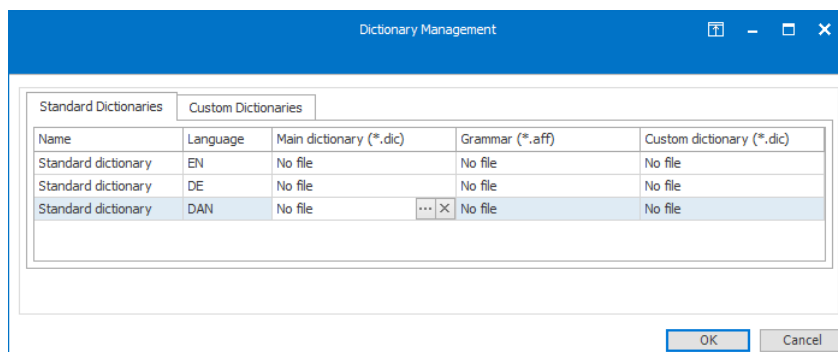
Each row tells you, for each language, whether a dictionary is uploaded or not. The table has 5 columns:

Name: All rows will contain the words "Standard Dictionary". This column makes more sense for custom dictionaries, since these you are allowed to name yourself. For Standard dictionaries the name is fixed.

Language: This column tells you which language the row represents. As previously described, all Open Office-dictionaries must have a language associated to it.

Main dictionary: This column contains the dictionary file being part of the Open Office-dictionary package. It tells you whether the main dictionary file has been uploaded for that respective language or not. When a file has been uploaded, Perfion will show you the text "Dictionary loaded" together with the size of the dictionary (in bytes). If no file has been uploaded, the cell will show "No File" as can be seen in the above screenshot.

Grammar: This column represents the other file being part of an Open Office-dictionary package, namely the Grammar-file (.aff). The file is uploaded the exact same way as the main dictionary.

Custom Dictionary: This column represents the default custom dictionary for that language. It often starts out empty, leaving it up to the users to add words to it. You can however, upload a dictionary file to it as a starting point.

Note that users are allowed to add words to a custom dictionary regardless of whether it starts out empty or not.

2.3. Custom Dictionaries tab

The content of the "Custom Dictionaries"-tab is very much like the content of the "Standard dictionaries"-tab: It is simply a list of dictionaries connected with languages. There are 2 fewer columns, since there is no Open Office-dictionary files associated with a custom dictionary.
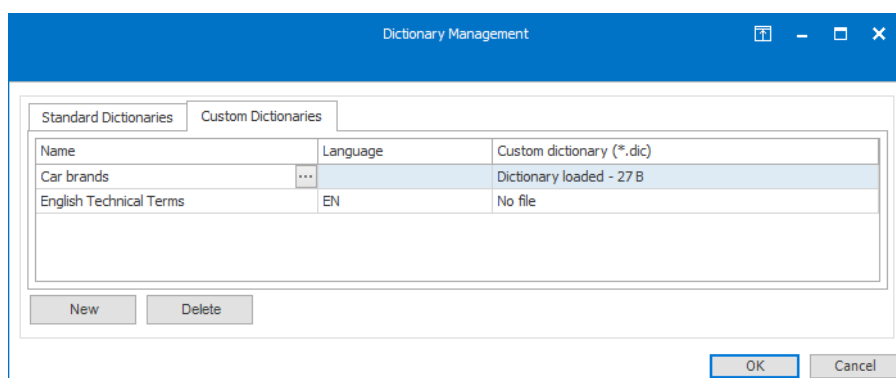
Name: As opposed to standard dictionaries, you are allowed to change the name of custom dictionaries. It is not allowed to use the same name twice to avoid confusing the user adding words to the custom dictionaries in the Item Editor. More about adding words to a custom dictionary in section 3). A dictionary name must contain at least 3 letters.

Language: This column tells you whether the custom dictionary is associated with a language. If it is, then the language is shown here, otherwise the cell is blank.
When a dictionary is associated with a language, the words in the dictionary are considered correctly spelled words for that language. If a dictionary is not associated with a language, the words in the dictionary are considered correctly spelled for all languages. Just as the example with car brand names previously described, where "Audi", "BMW" etc. should be considered correctly spelled in any language.

You cannot associate a dictionary with 2 or more languages (for example both British English and US English). If you have that need, you will have to upload the same dictionary-file twice.

Custom dictionary: This column holds the dictionary content. As with default custom dictionaries, you can choose to add a file to start with or you can simply leave it blank, making it up to the users to add words to it and hereby building it up from scratch.

2.4.   Managing standard dictionaries

Knowing now what a Standard and Custom dictionary is, we are now ready to describe how to add first a standard dictionary.

2.4.1. Adding a standard dictionary

Due to the licensing rules on most of these dictionaries, Perfion is not allowed to embed them as part of the Perfion-product (i.e. add them to the client installer or similar). We are however, allowed to distribute them as long as we do it publicly. What we at Perfion have done is, is making a collection of Open Office-dictionaries available on our support-site, which you can find on the address: support.perfion.com/dictionaries.html. There are, however, numerous places on the internet where Open Office dictionaries may be found, so feel free to look elsewhere if the dictionary-files you are after is not on the support-site.

Note that Open Office dictionaries are often delivered in an archive file (for example a .zip or .rar-file) since they consist of many files. These files can usually be divided into 3 types:

a. The dictionary (a .dif-file). Perfion as described later needs this file.

b. The grammar (a .aff-file). Perfion as described later needs this file.

c. The license information documents (usually text files describing the license). Make sure to read the license information, so that you do not violate the license.
    Many Open Office-dictionaries are indeed licensed under GNU GPL or Mozilla MPL in some version and can hereby freely be distributed as long as you keep the licensing information with them. Perfion does not need these files.

To add a standard dictionary carry out the following steps:

1) First, find the appropriate Open Office-dictionary for the language you want to spellcheck. Most (if not all) Open Office dictionaries are Open Source meaning that they are allowed to be used freely. As previously mentioned Perfion has collected a number of dictionaries and made them available on our support-site at support.perfion.com/dictionaries.html, but numerous other
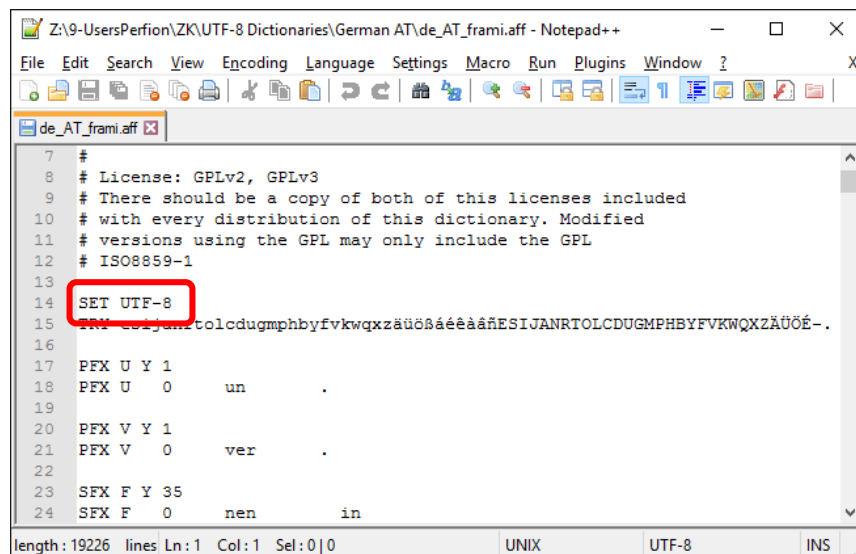
sources can be found on the internet.

2) If you have downloaded your dictionary from Perfions support-site, you can skip this step.

If not, you will have to check that the dictionary- and grammar-file are both encoded using UTF-8 encoding, since Perfion will only accept dictionaries in UTF-8. Often dictionaries found on the internet are *not* UTF-8 encoded, but are using some 8-bit encoding (typically ANSI).

In case they are not UTF-8 encoded, you would have to convert both the dictionary- (.dic) and grammar (.aff) to UTF-8 encoding.

The tasks of checking the encoding of a file and converting it to UTF-8 encoding can both be done using a text-editing tool like Notepad++. Notepad++ can freely be downloaded here: https://notepad-plus-plus.org/.

Below screenshot shows a German aff-file opened in Notepad++. Note that the encoding detected by Notepad++ can be seen in the bottom-right position. Here it is UTF-8. If it was not UTF-8, you should go to the Encoding-menu, select the "Convert to UTF8"-menu item and save the converted dictionary.

In addition to converting the encoding of the files, you will need to edit the aff-file finding the line beginning with "SET" followed by the original encoding. The original encoding needs to be changed into UTF-8 as showed in red rectangle above, where it has been done.

Note: Perfion will show an error message if uploaded files are not UTF-8 encoded or if the aff-file does not have a proper SET-statement.

Tip: The original SET statement can actually help you figure out how the file originally was encoded. That information might be helpful, in order to successfully convert the file. To check whether you converted the file correctly, it is often revealed by scrolling down the dictionary and checking that the words "look right". "Weird characters" are an indication that something went wrong.

3) Upload Main dictionary: Clicking the "triple dot"-button in the Main dictionary-cell opens the Upload Dictionary-dialog:
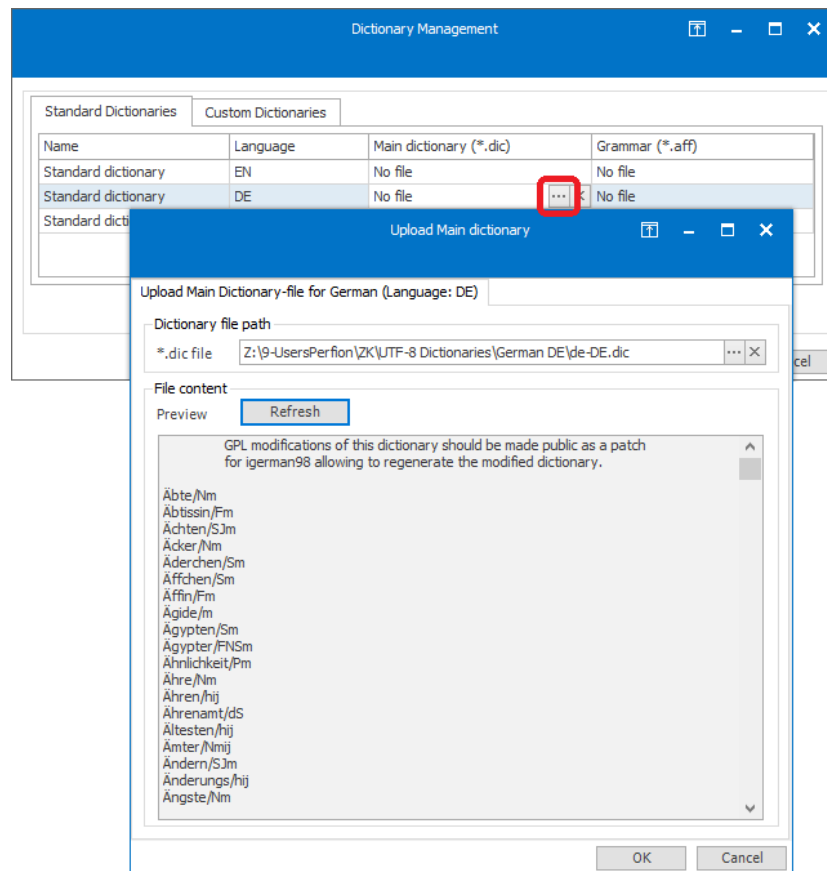
To upload a dictionary, please select a dictionary file at the top and click OK. If you want to check the content before you upload, you can click "Refresh" to load the content of the file into the preview textbox. This is seen in the above screenshot where Perfion is showing us the first words in a German-dictionary. Note that the words look OK: We indeed see the German letters like Ä (A-umlaut). Feel free to scroll down to check more words in your dictionary before clicking OK.

4) Upload grammar: This works the same way as for the main dictionary. Now you should just use the "triple dot"-button in the Grammar-cell and pick the corresponding aff-file in the Open Office-package.

5) Upload default custom dictionary. This part is usually skipped, so users will start of from scratch. If you want to upload a default custom dictionary you can see in section 2.5.1 how this is done.

When done the "Standard dictionaries"-tab should look like the following:
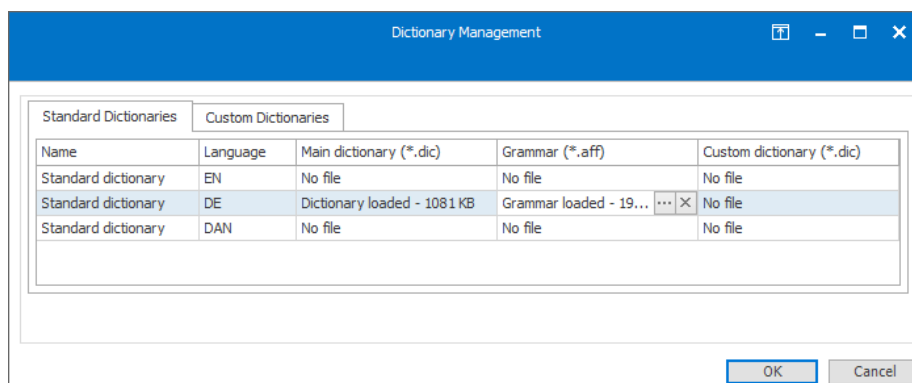
As can be seen our German main dictionary takes up 1081 KB of space while the grammar is 19 KB. We have chosen not to upload any custom dictionary, so that cell says "No file".

Note: In order to spellcheck in some languages, Perfion needs *both* the main dictionary- and its corresponding grammar-file for that language (i.e. both Open Office-dictionary files). Without both these files, it cannot do spellchecking in that language. Perfion will, when OK is clicked in the dialog above, warn you about all languages that only has one of the files uploaded.

6) When done uploading dictionaries click OK to save your changes. A click on cancel will drop all modifications and go back to the Perfion main form (you need to confirm).

2.4.2. Deleting a standard dictionary
If you, for some reason, want to delete a standard dictionary, you need to carry out the following steps:

1) Consider storing a backup of all the dictionary-files part of the standard dictionary. The Open Office-files can often just be found again on the internet, but the default custom dictionary will be lost for good when deleted. See how to make a backup of a dictionary in section 2.4.3.

2) In the "Main dictionary"-column click the X-button. Confirm that you indeed want to delete the main dictionary clicked.

3) Do the same for the Grammar-column.

4) Do the same for the Custom dictionary-column.

5) When done click OK to save your changes. Note that clicking cancel will about all modifications.

2.4.3. Viewing a standard dictionary

If you at any point want to check the content of a standard dictionary, you can do that by clicking the "triple dot"-button in the "Name"-column. This will take you to the "View Dictionary"-dialog shown here:
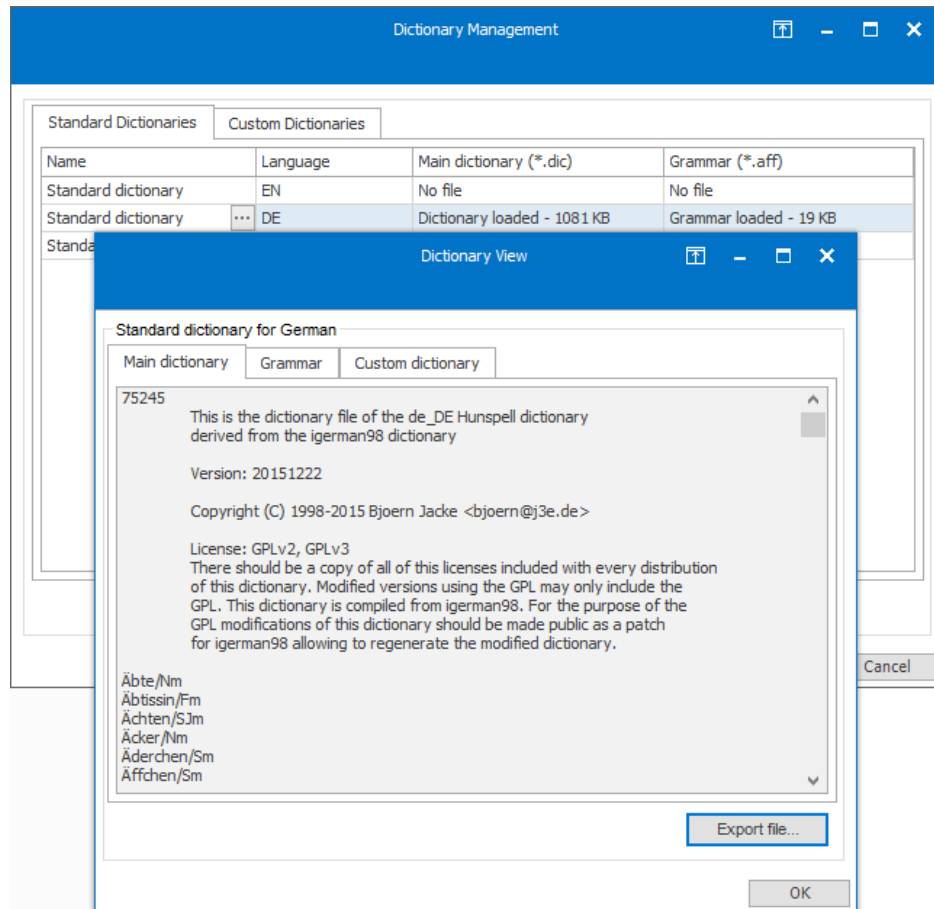
As expected, you can see both the "Main dictionary"-file, the "Grammar"-file and the default custom dictionary for that language.

Notice the "Export file…"-button in the dialog. This button allows you to export any dictionary- and grammar-file in Perfion to a text-file. This is a simple way of backing up individual dictionary files if needed.

2.5. Managing custom dictionaries

The purpose of Custom dictionaries is to allow you to add "special words" to Perfion, so that, for example, Perfion does not consider your brand names, or the technical terms you are using,

misspellings. They should not be used for dictionaries that could be considered containing normal words for some language. Such words should go into the default custom dictionary.

2.5.1. Adding a custom dictionary

When adding a custom dictionary, Perfion allows you to add a blank dictionary having the users built it up as they start editing products in the Item Editor. Simply by adding words to that dictionary. Remember that not only standard dictionaries are stored centrally and shared by the users, but also custom dictionaries. So whatever User A adds, will benefit User B.

In case you have some "custom words" already in some form, you can get a head start by uploading these words to Perfion as the initial content of that custom dictionary.

The file format for a custom dictionary is very simple: One word per line in a UTF-8 encoded text file with a .dif file extension. Again, a good text-editor like Notepad++ can be helpful to create a custom dictionary.

Here we, as an example, have the very beginning of a "Car brands"-custom dictionary:
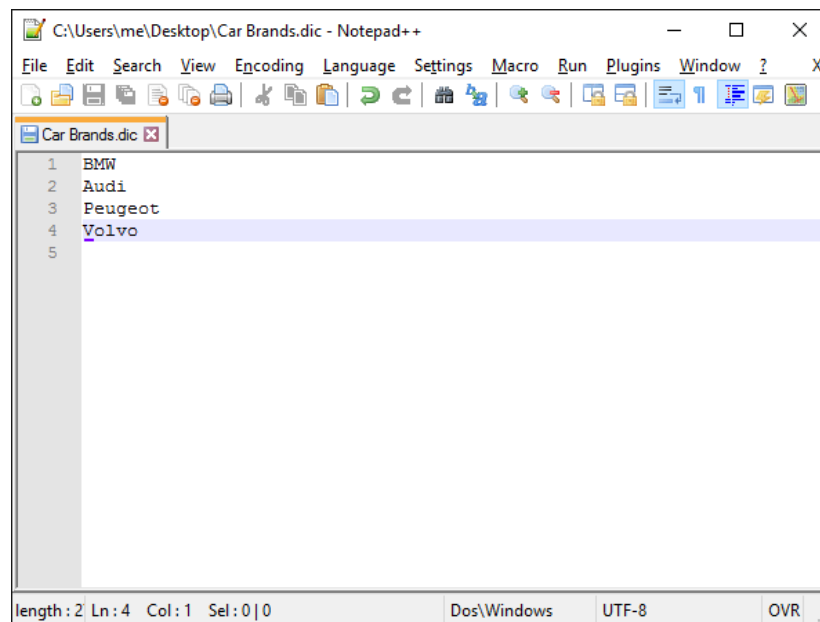
Notice the file extension is indeed .dic and that the order of words is unimportant.

Since all words in any custom dictionary are considered correctly spelled by Perfion, it means that you should be careful not to put spelling errors in here.

To add a custom dictionary to Perfion follow these steps:
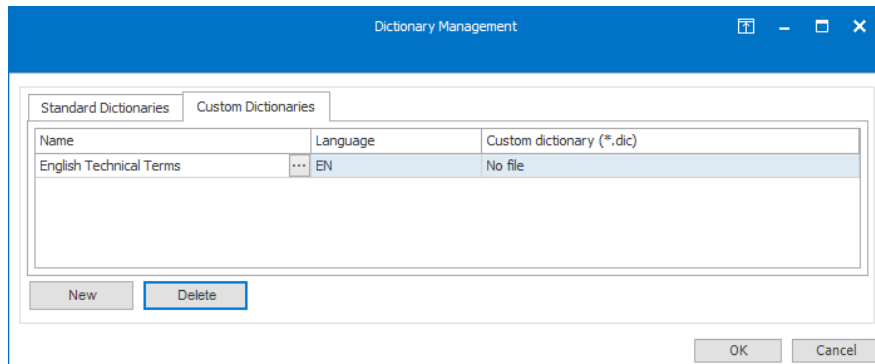
1) Click "New"-button

2) Type in a describing name in the "Name"-column. Make sure the name is unique. Perfion requires uniqueness of names in order for the user to be able to distinguish between multiple custom dictionaries.

   Note: The names "Dictionary", "Dictionary+" and "Grammar" are reserved by Perfion and cannot be used as names.

3) Pick a language if the dictionary is associated with a language, i.e. its words should only be considered correctly spelled for that particular language.

   If you erroneously pick a language for a custom dictionary, you can remove it by clicking the X-button that appears in the cell when it is selected.

4) If you have a .dic-file as described above containing the (initial) words for that dictionary you can upload it by clicking the "triple dot"-button in the "Custom dictionary"-cell. This will allow you to pick the file and upload it.

   If you instead want to start from scratch, you can leave this cell saying "No file".

5) Clicking "OK" will save your changes to Perfion.

If you carry out these steps for the "Car brands.dic"-file previously shown you will end up seeing this:
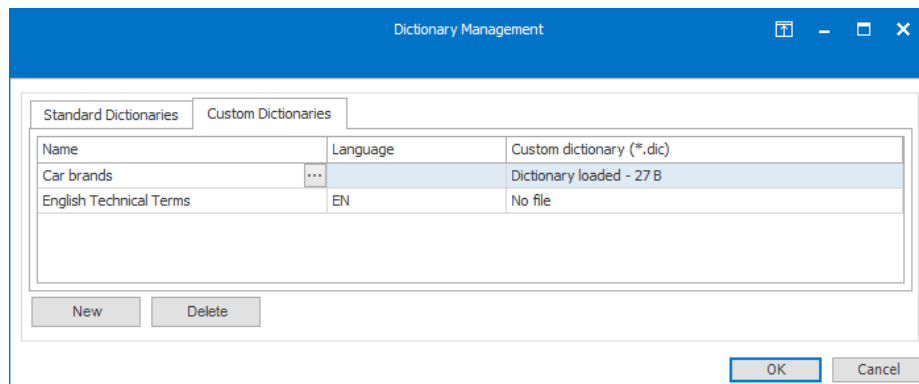
Notice that no language is selected, since car brands are considered universal to all languages.

2.5.2. Deleting a custom dictionary

In order to delete a custom dictionary, you need to following these steps:

1) Consider saving a backup of the dictionary before deleting it. See section 2.5.3 for details on how to back up a custom dictionary.

2) Select the dictionary to be deleted in the table.

3) Click "Delete"-button

4) Confirm that you indeed want to delete the dictionary.

5) Click OK to save your changes to the database.

2.5.3. Viewing a custom dictionary

To view the content of a custom dictionary click the "triple dot"-button in the "Name"-cell for that dictionary. This will bring up the "View dictionary"-dialog:
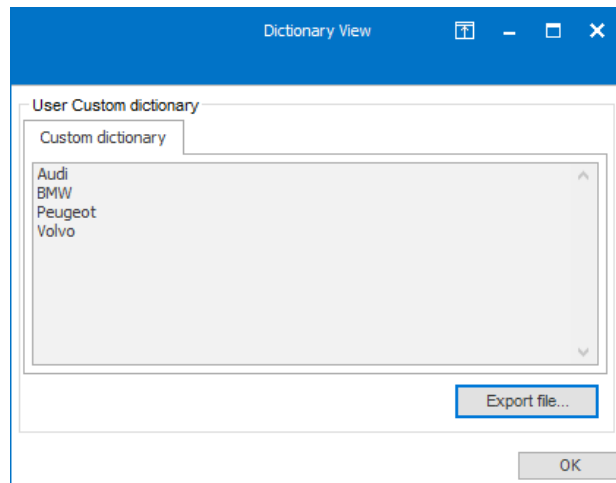
As for Standard dictionaries, you have the option of exporting the content of the dictionary to a file for example for backup purposes.

2.6.   Centrally editing a dictionary

Standard dictionaries, that is Open Office dictionaries, are not supposed to be editable in Perfion. The idea is that these dictionaries are maintained by various organizations. When they are updated, you can upload the new files to Perfion without worrying, that your local changes goes away.

As opposed to Standard dictionaries, we have custom dictionaries where it, at least to some extent, make sense to edit them centrally.

You cannot directly edit a dictionary in Perfion. What you can do, however, is to export a custom dictionary, modify it in some editor, and upload it again.

3.   Spellchecking functionality in Item editor

When at least one language has gotten an Open Office-dictionary the spellchecking feature is available for all users that has access to that language, that is, they can see it in the Item Editor.

The feature will be evident the first time the Item Editor is opened:
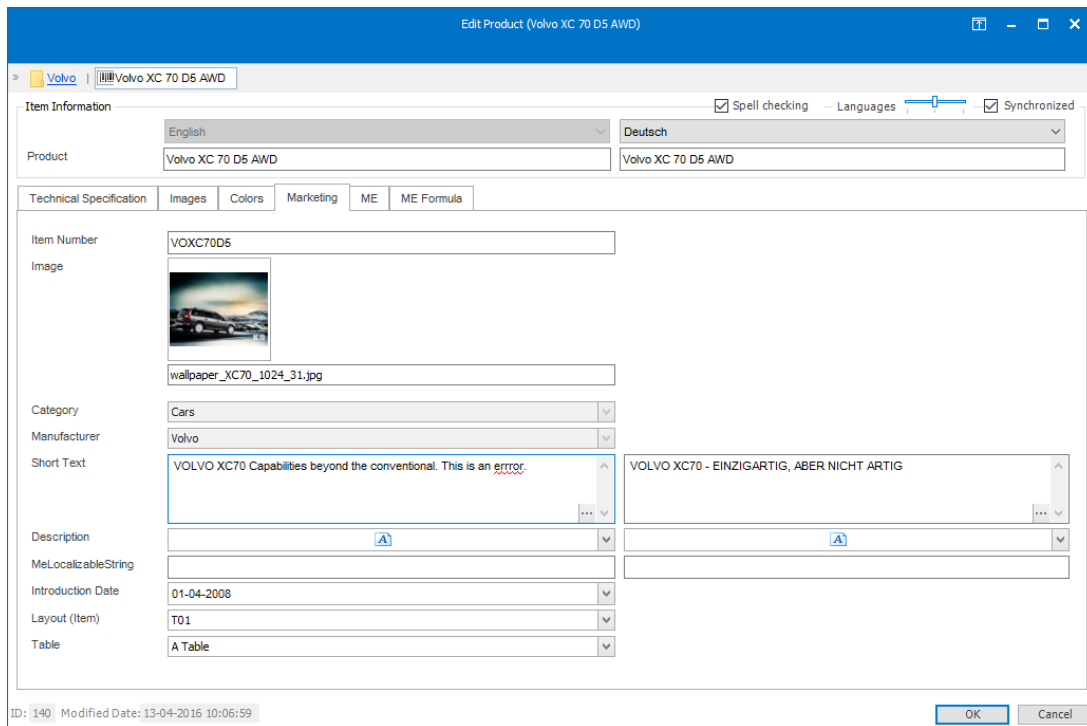
As you see a spelling error has been added to the English "Short text" for Volvo and it immediately is underlined with red.

Which controls that are spellchecked are governed by the following rules:

1) Spellchecking needs to be enabled on the Item Editor in order for any spellchecking to happen. It is enabled by default, but can be switched off by the user. This is controlled by the "Spell checking"-checkbox in the top right corner of the Item Editor. This check box is only available if there are dictionaries available for at least one of the languages shown.

2) Spellchecking is only carried out for localizable, explicit string- and text- features. In the above example "Short Text" and "Product"-features are spellchecked, but "Item Number" is not, since it is not a localizable feature.

3) Spellchecking is only carried in the currently selected control. As an example: If there was a misspelling in the German "Short text" it will not show unless you activated that control. By the same rule, a misspelling in the English product-feature will not be shown.

4) There is a limitation in Perfion at the time of writing (Perfion version 4.5.6) and that is that spellchecking is not carried out in the Rich Text-editor due to a technical limitation in that editor.

When an error is highlighted, in addition to simply allowing the user to correct the error by retyping the word, he can also right click the misspelled word to get some suggestions if Perfion can find any:
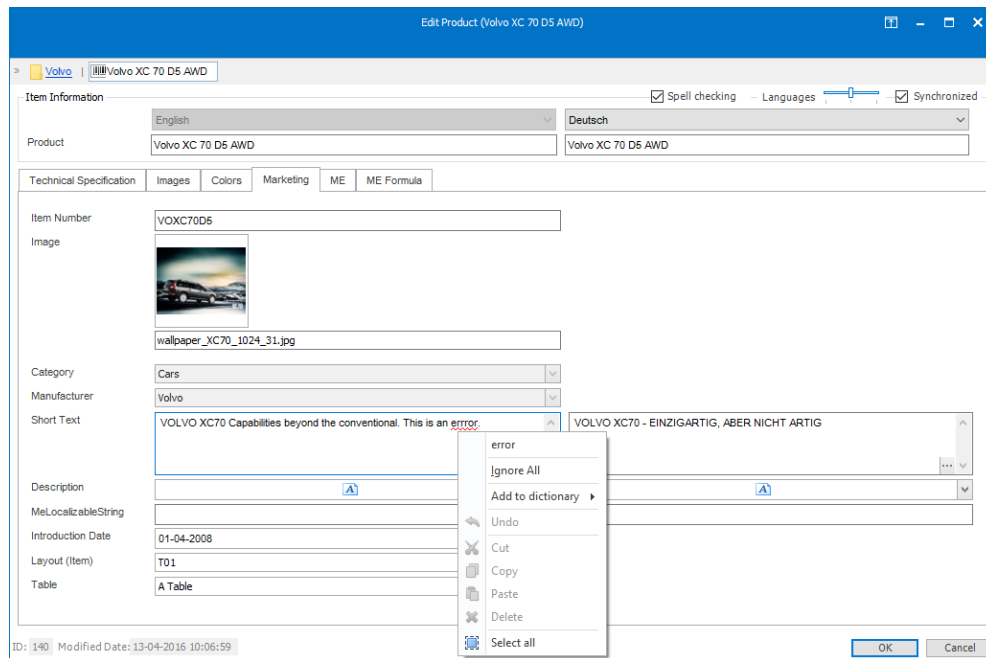
As can be seen in the above example the word "error" is suggested as the only one instead of the misspelled word "error".

Furthermore you have the option of adding what is right now considered a misspelling to a custom dictionary. This is shown in the following screenshot:
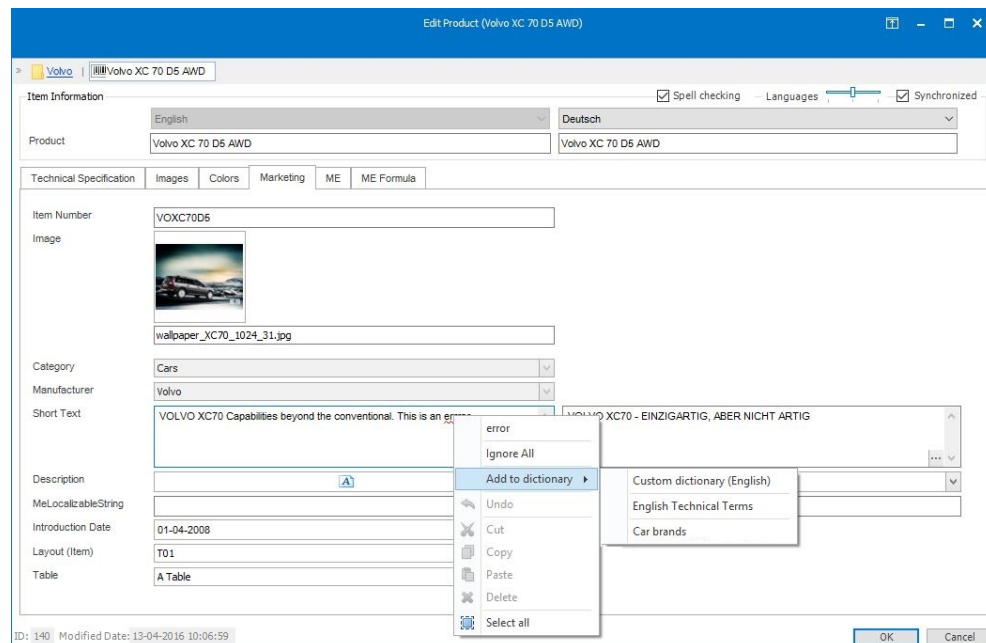
As can be seen you have the following options (dictionaries to pick from):

1) Custom dictionary (English): Picking this would mean adding the word to the default custom dictionary for the language (See more in section 2.4.3). Words that are considered correct words in the current language should be added here.

2) "English Technical Terms": This is a language specific custom dictionary. If the word is considered an "English Technical Term" this is the custom dictionary to pick.

3) "Car brands": Finally, you have the language independent dictionaries; here you only have "Car brands". If the word is a car brand it should go in here.

Words added by the user are stored in the database when the user leaves the Item Editor, regardless of whether the user saves the item edited or not.

3.1. Increased time to open and close Item Editor

Dictionaries can contain quite a lot of data and if used they need to be loaded and stored by the Perfion client.

Perfion loads and saves dictionaries in their entirety. Dictionaries are loaded the first time a user opens the Item Editor, but only for the languages, the user has access to. In case some user has access to many languages with dictionaries associated with them, the Item Editor may take a few extra seconds to load.

The same applies to doing modifications to custom dictionaries: In case a user makes a modification to a huge custom dictionary it may take a few seconds before Perfion closes the Item Editor, since changed dictionaries are saved upon closing it.

Perfion, of course, only saves custom dictionaries when modified and only reloads dictionaries if some user has modified them since last time it was loaded.