MetaAnn_ User Guide

Conteúdo

1. Intro

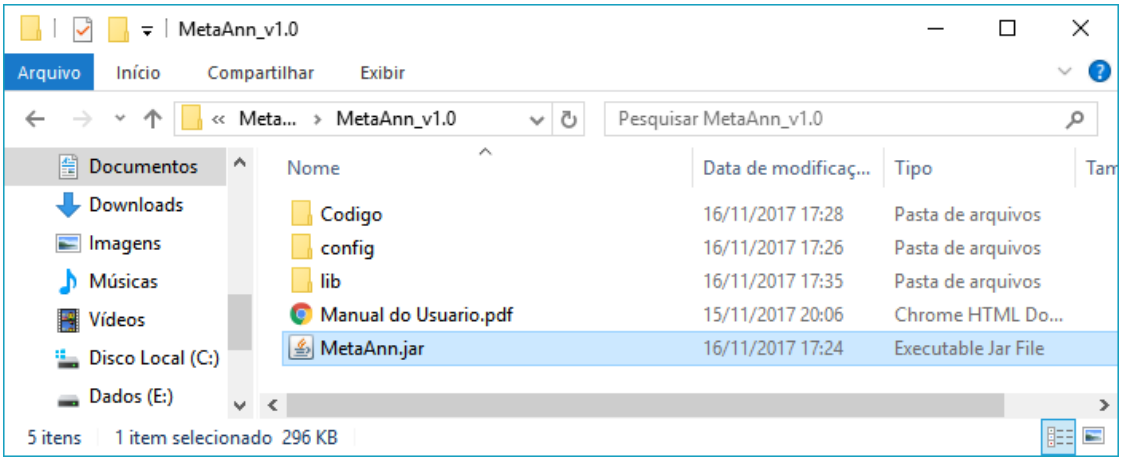
1.1 About MetaAnn

MetaAnn is a tool of annotation tools building. On MetaAnn, the user builds text annotation tools, with the possibility of annotate each basic unity of a *corpus* of interest.

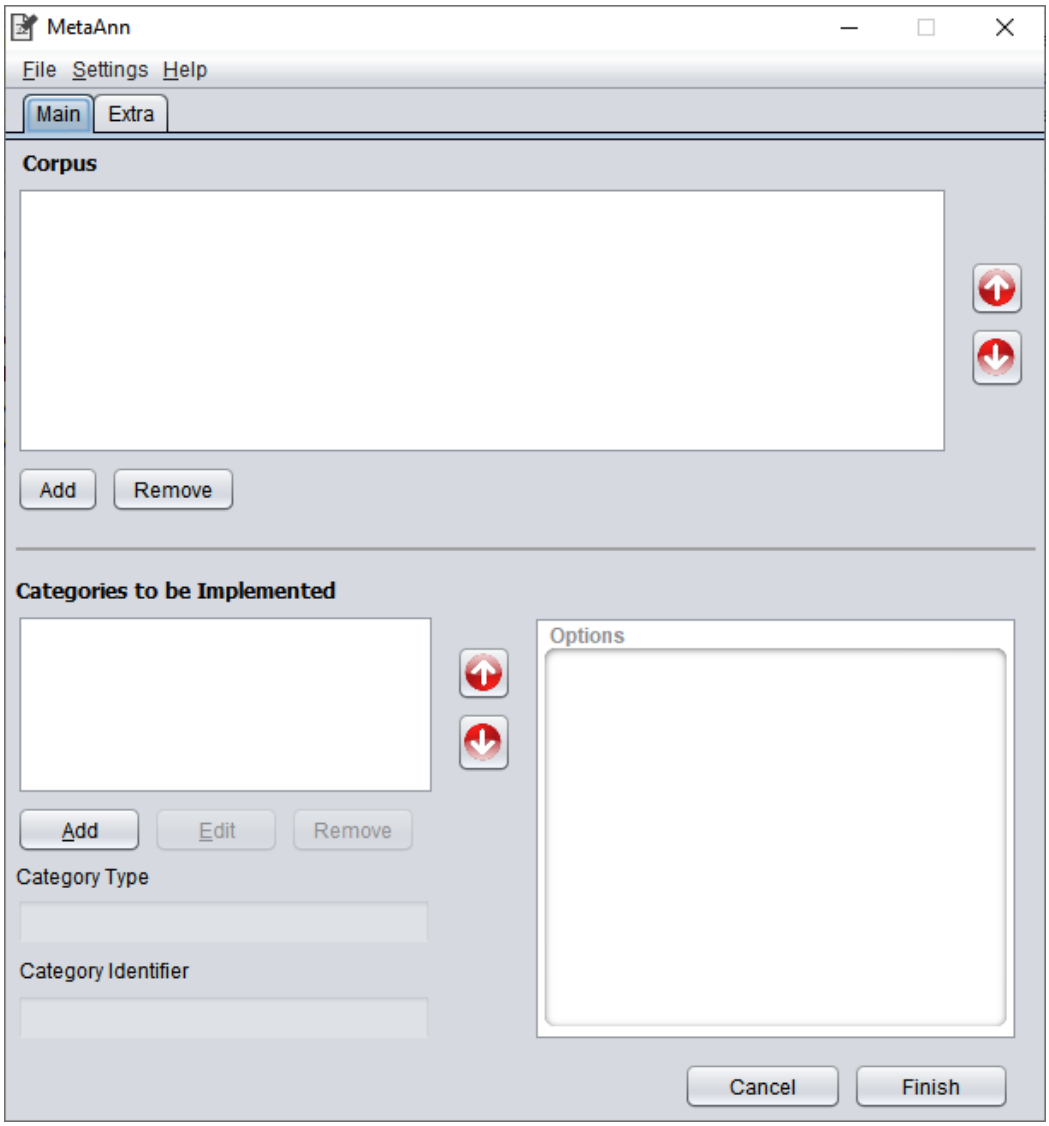
Both MetaAnn and the tools generated by it was built in Java and, so, it is necessary install on computer the JRE, in version 7 or newer.

1.2 Opening the program

To open MetaAnn, just double-click the file “MetaAnn.jar”, which is inside the program directory:



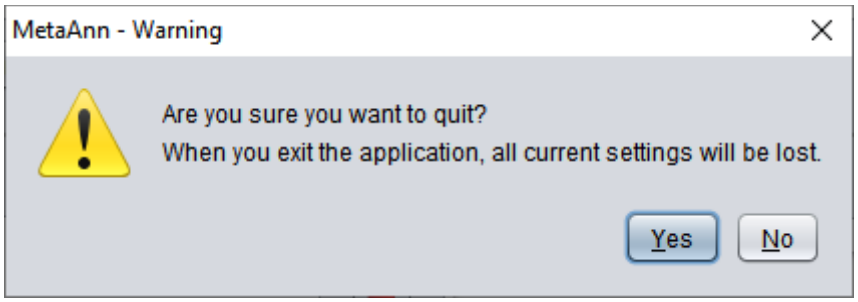
After opening the application, the user will face the initial screen. The MetaAnn window contains two tabs: The “Main” tab defines the main characteristics of the annotation tool: information regarding the corpus under study and the structure of categories of the tool. At “Extra” tab, the user can configure extra information about the corpus he wants to view in the generated tool.



1.3 Exit MetaAnn

To exit MetaAnn, the user can click the “Cancel” button in the main window, he can also click the close button (X) in the title bar, the user can also click the “Exit” item in the “File” menu, or simply use the keyboard shortcut “ALT + F4”.

Upon exiting the system, the following confirmation message will appear:

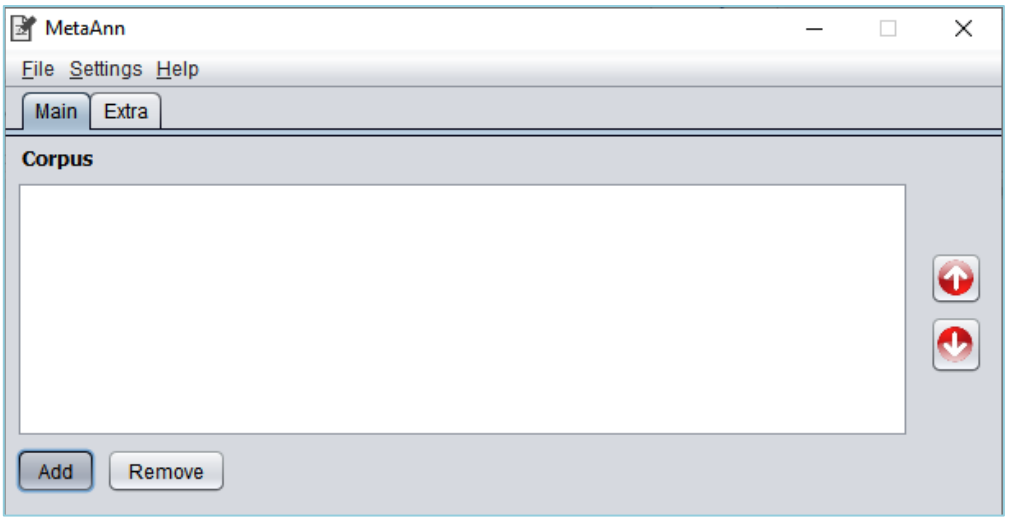


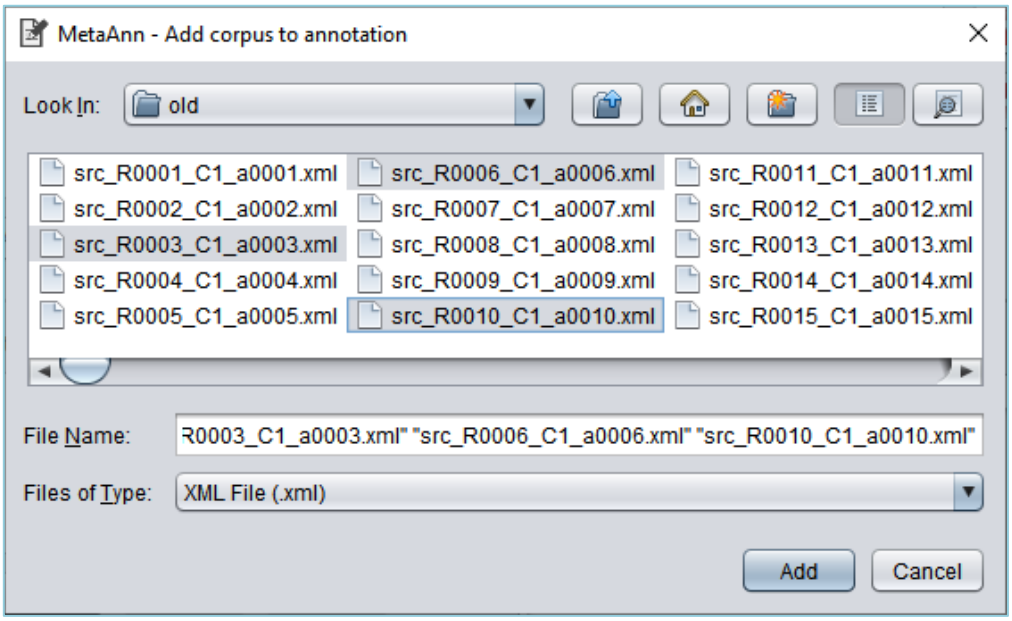
2. The “Main” tab

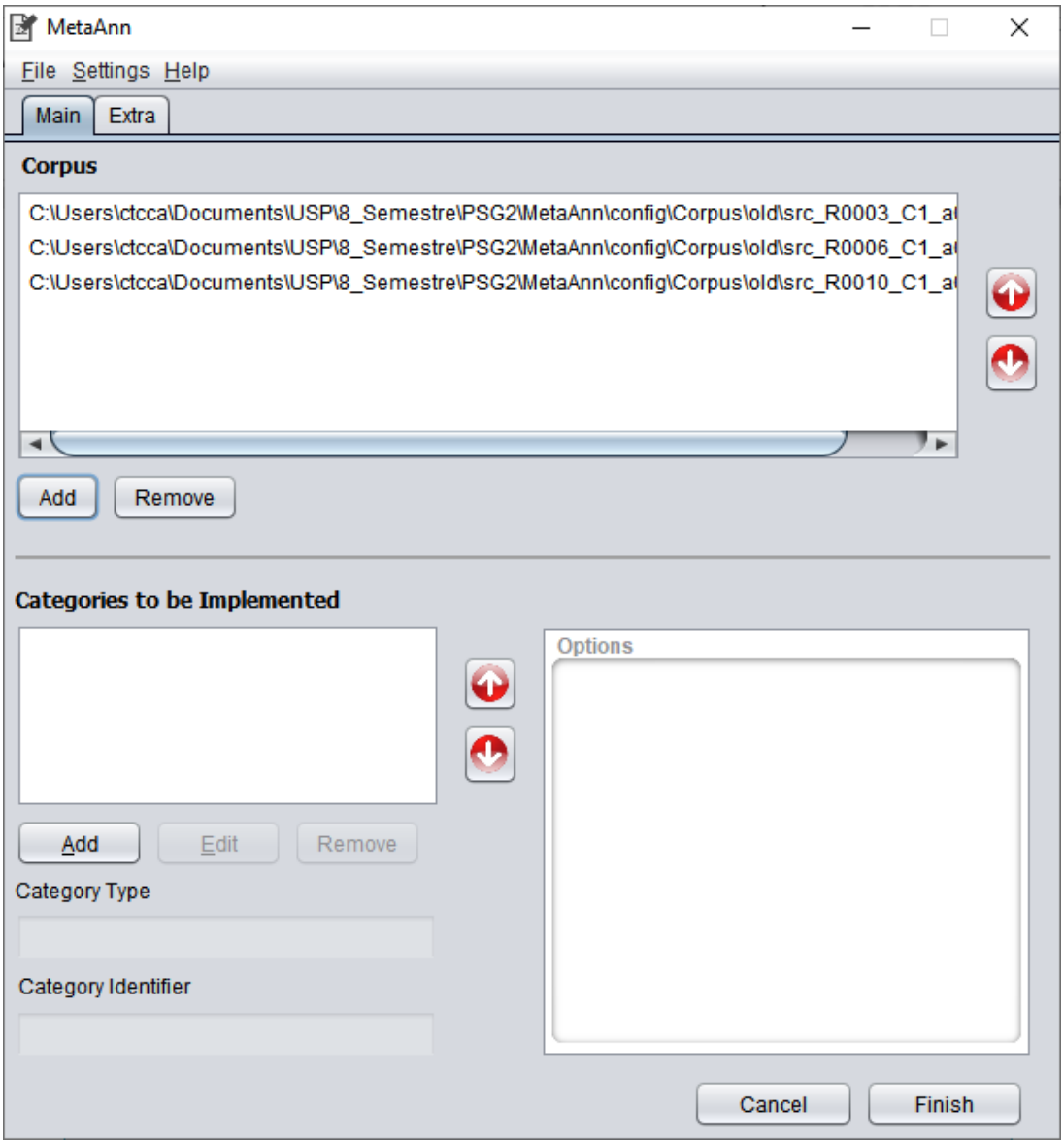
There are two sections at “Main”: the first section is for selecting the corpus to annotate, and the second section contains the categories that the generated tool should contain.

2.1 Defining the corpus to be annotated

To define the corpus, the user must include the text files properly segmented in basic annotation units, according to the standard described in ROMAN (2012). To include the files, just click on the "Add" button in the "Corpus" section of the main tab, select the files (with xml extension) and include:







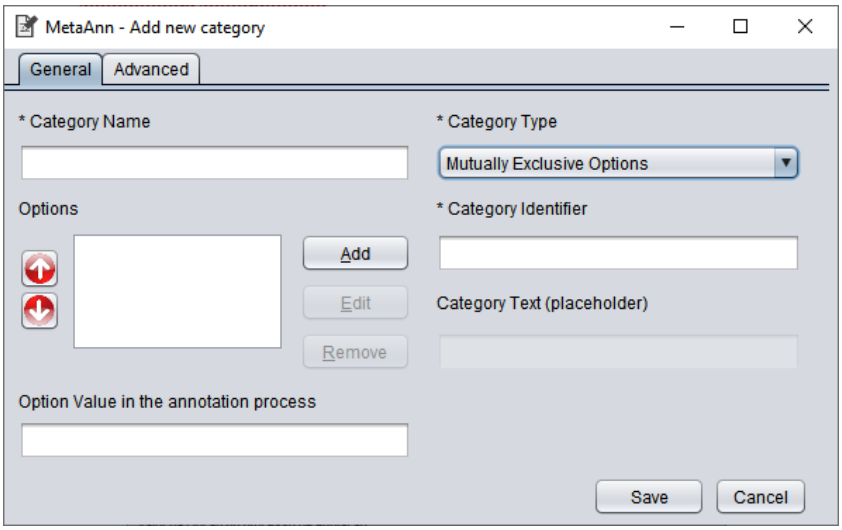
To include more than one file in the same directory, just hold the “Ctrl” key and click on all the files you want to add. To select a range of files, just click on the first file, hold the “Shift” key and click on the last file in the range. This standard feature of operating systems can be very useful for selecting all files in a given directory, for example.

Analogously to the inclusion, it is possible to remove a file from the corpus set for annotation, just by selecting the file from the list and clicking on the “Remove” button. To select more than one file in the list, it is also possible to use the “Ctrl” and “Shift” keys.

The generated tool will display annotated texts in same order as they are in the MetaAnn file list. To change the display order, simply select an item from the list and change its relative position using the ↑ (up arrow) and ↓ (down arrow) buttons.

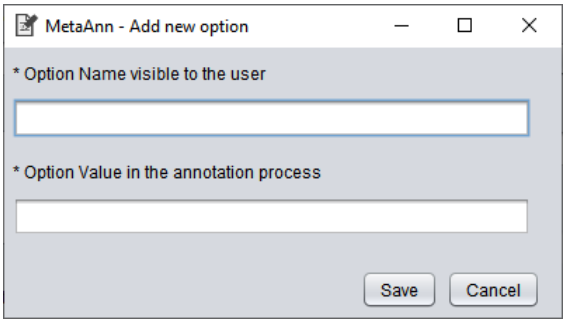
2.2 Including categories

To add a new category, just click on the “Add” button in the “Categories to be implemented” section. Then, the category addition window will appear. MetaAnn works with three types of categories: Mutually Excluding Options, Multiple selection list and free text. More information about the difference between categories` types can be found at MISSÃO & ROMAN (2013).



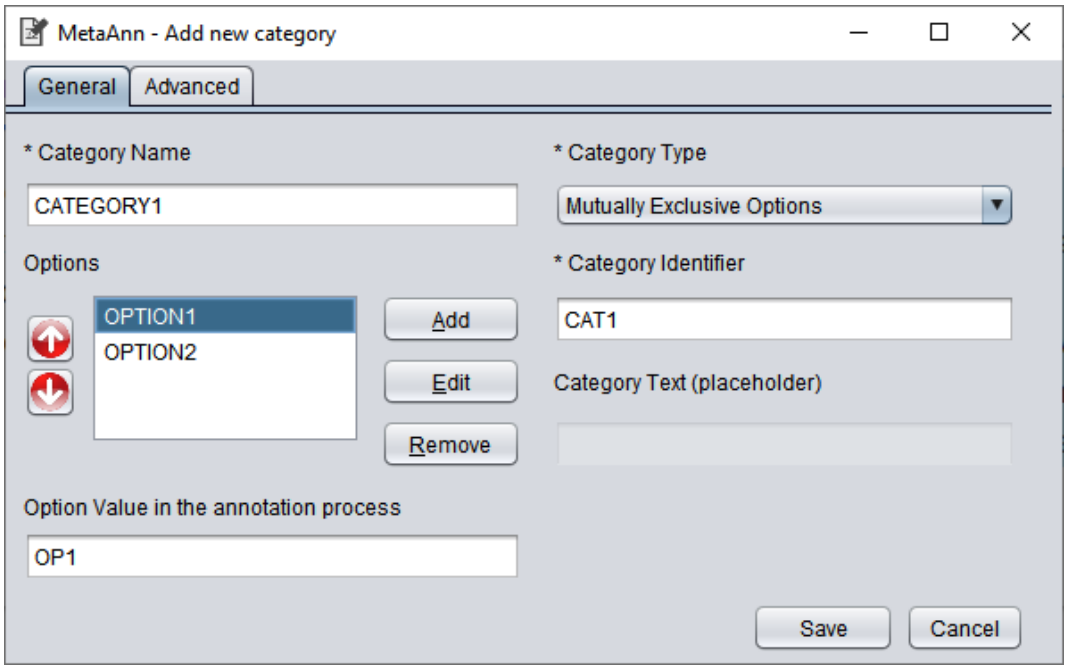
For the first two types of categories, it is required to include two or more options. The basic difference between them is that categories of the type "Mutually Excluding Options" allows the selection of only one option for classification of the category, while categories of the type "Multiple selection list" allows the selection of more than one option.

To include an option, click the “Add” button, next to the "Options" box, and the option addition window will open:



In this window, there is two items related to the option: the “Option Name visible to the user” is the text that will appear for the option in the generated tool. In

contrast, the “Option Value in the annotation process” is the value that used in annotation. By clicking on the “Save” button, the option is saved and displayed in the list of options on the previous window.



After registering the options, it is possible to manage them in the category configuration section. When selecting a category, the “Edit” button and the “Remove” button are enabled. The “Edit” button opens a screen identical to the addition screen, but with the data filled in, where the user can change the option data. When clicking on the “Remove” button, the option will be removed from the list.

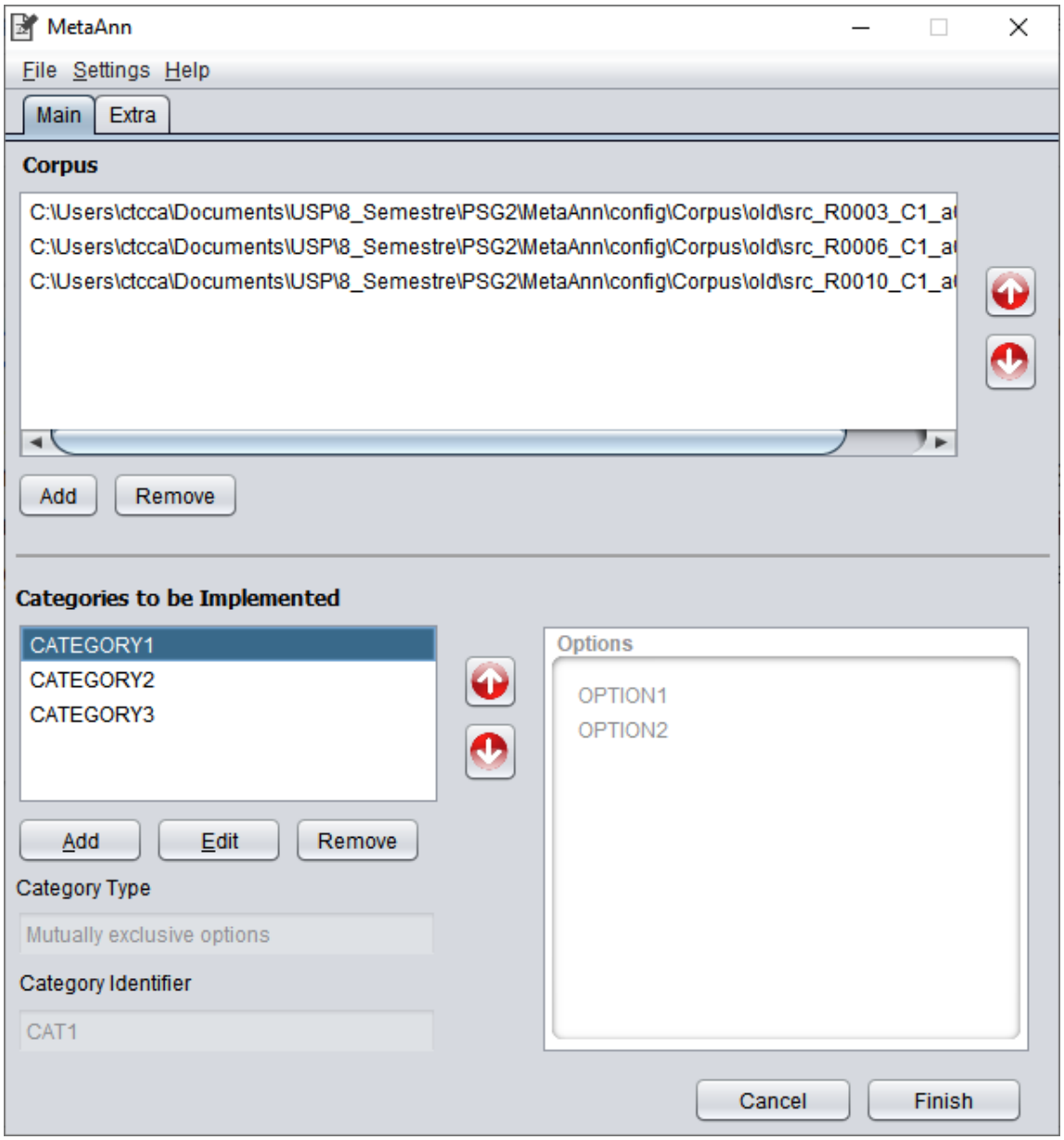
The system display the options on the tool in the same order as they are arranged in the list of options. If you need to change the display order, just use the ↑ (up arrow) and ↓ (down arrow) buttons that are next to the list.

For “Free text” categories, it is not possible to configure options; the user makes the classification directly, by typing in the field related to the category. For this type of category, the “Category Text” field is enabled for typing. In the generated tool, the content of this field appears into the field for the category, as a placeholder, to add an explanation of what needs place in the field's content.

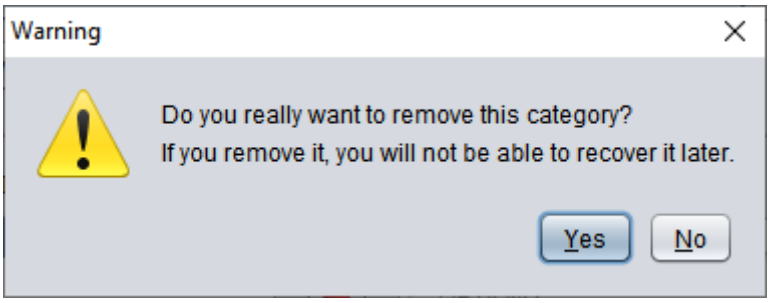
After setting up a category, just click the save button. If everything is correct, the category will be saved, the window will be closed, and the interaction will return to the main window, where the category will be included in the list of categories.

2.3 Managing categories

After including the categories, the user can access all of them from the list of categories in the “Main” tab. By clicking on a category, it is possible to view, just below the list, its type and its identifier and, beside, the configured options, if the type of category allows options.



It is possible to modify the category settings by selecting a category and clicking the "Edit" button. After clicking this button, an editing screen identical to the category addition screen opens, allowing you to edit any settings in the category. To remove a category, just select it and click on the “Remove” button. When attempting to remove, a confirmation screen will appear. To confirm the exclusion of the category, the user simply clicks the "Yes" button.

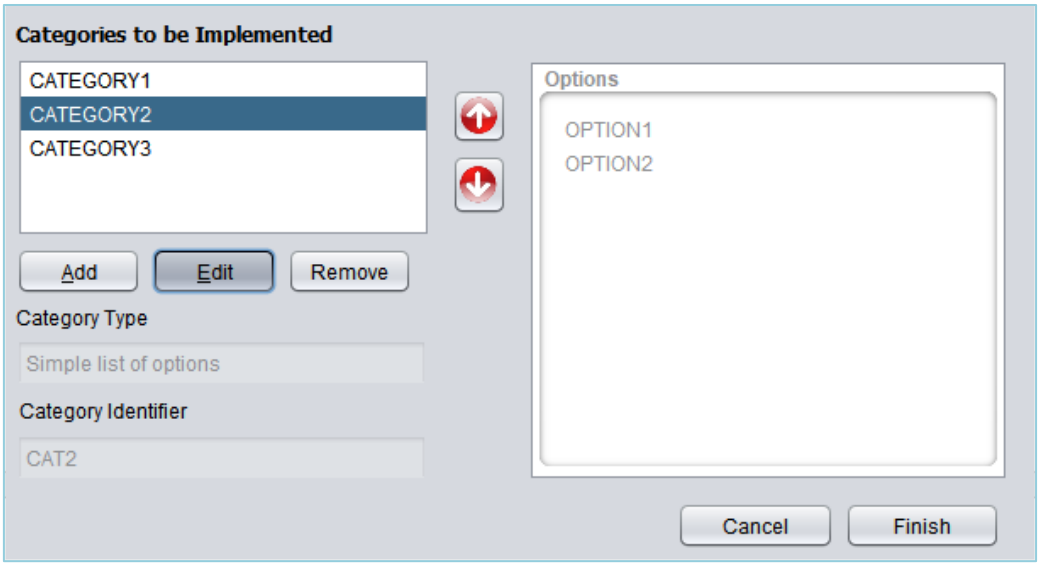


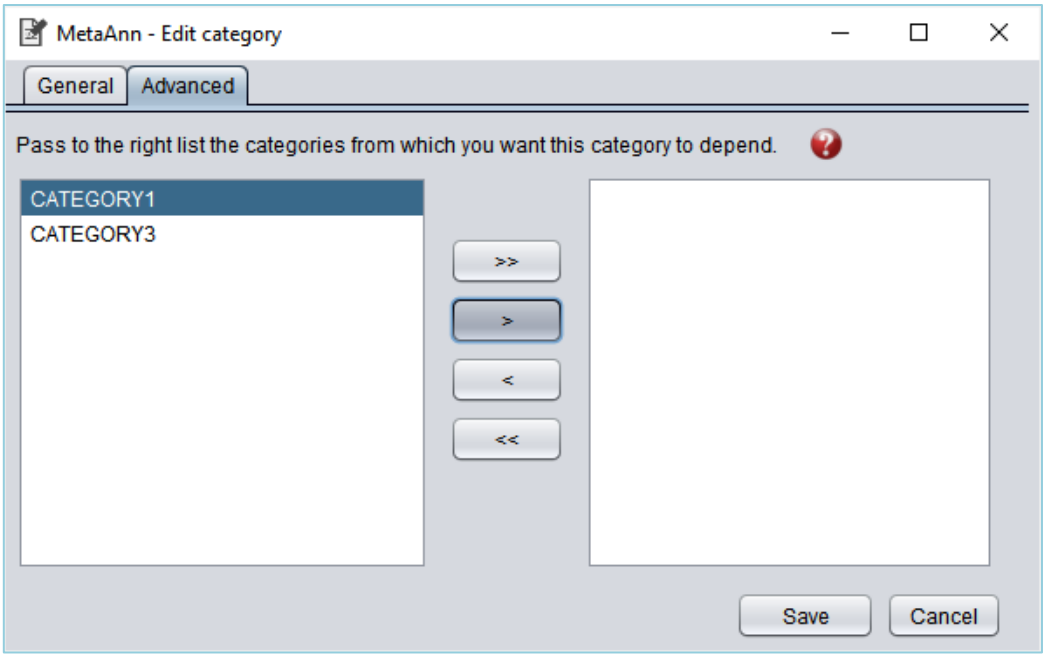
The generated tool displays the categories in the same order as they are in the category list. To change the display order, simply move the category in the list, using the ↑ (up arrow) and ↓ (down arrow) buttons next to the category list.

2.4 Defining dependency between categories

In MetaAnn, it is possible to define dependency between categories, so that, in the generated tool, filling in a category is unavailable until the user fill the category on which it depends.

To include dependency on a category in relation to another, it is necessary to use the category-editing feature. This means opening the category edit screen and configuring the dependency on the “Advanced” tab.





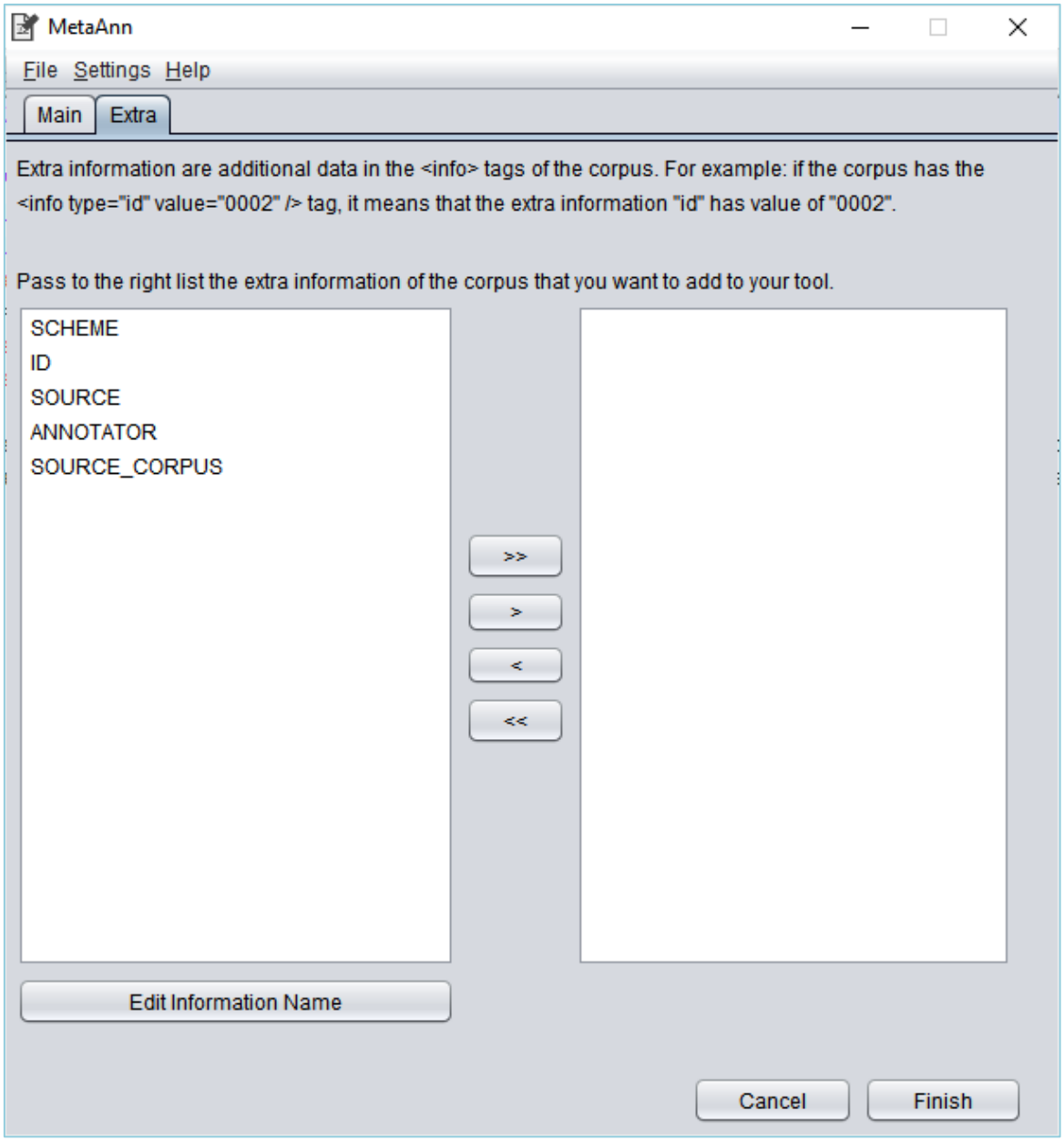
To define that a category depends on another(s), just move to the right list the categories that the user wants the current category to depend on. It is not possible to add cyclical dependency, that is, to define that a category depends on another that depends on the first.

To remove the dependency, just go back to the left list the categories added to the other list and click on the “Save” button. If a category does not appear in the left list, the following reasons can explain: it is the current category (a category cannot depend on itself), or because it generates cyclical dependency if it is added.

3. The “Extra” tab

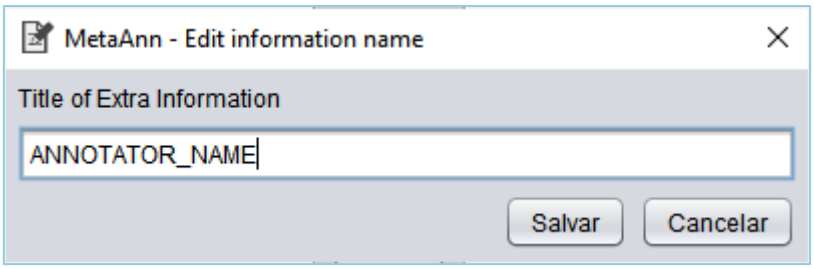
On “Extra” tab, the user can configure extra information that he wants to see in the generated tool. Extra information is the information inside the <info> tags, according to the segmented file standard defined in ROMAN (2012).

The extra information is loaded according to the first file in the list of files defined by the user in the “Main” tab. This means that, when the user defines the corpus files, the system will examine the first file, collect all additional information from it, and fill in the list of extra information on the “Extra” tab.



To view the content of extra information in the generated tool, the user simply passes the information to the list on the right. It is also possible to change the name of the information, the label displayed to the user in the generated tool. Whether the file contains a segment `<info type="annotator" value="Test">`, the standard label is “ANNOTATOR”, and the generated tool will display the information “ANNOTATOR: Test”.

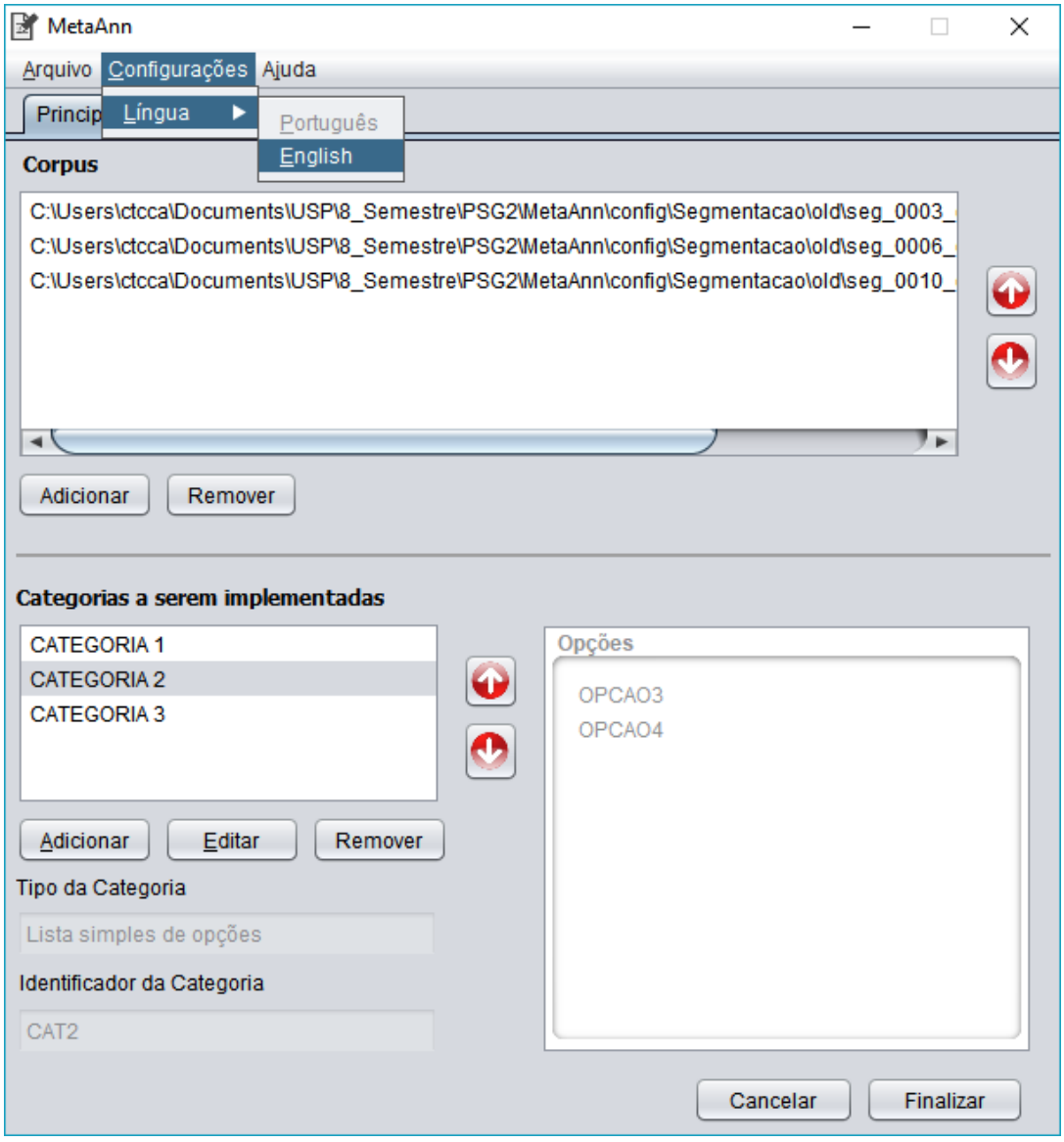
Selecting the information you want to change, just click on the “Edit Information Name” button, change the name of the information in the window that will open and save.



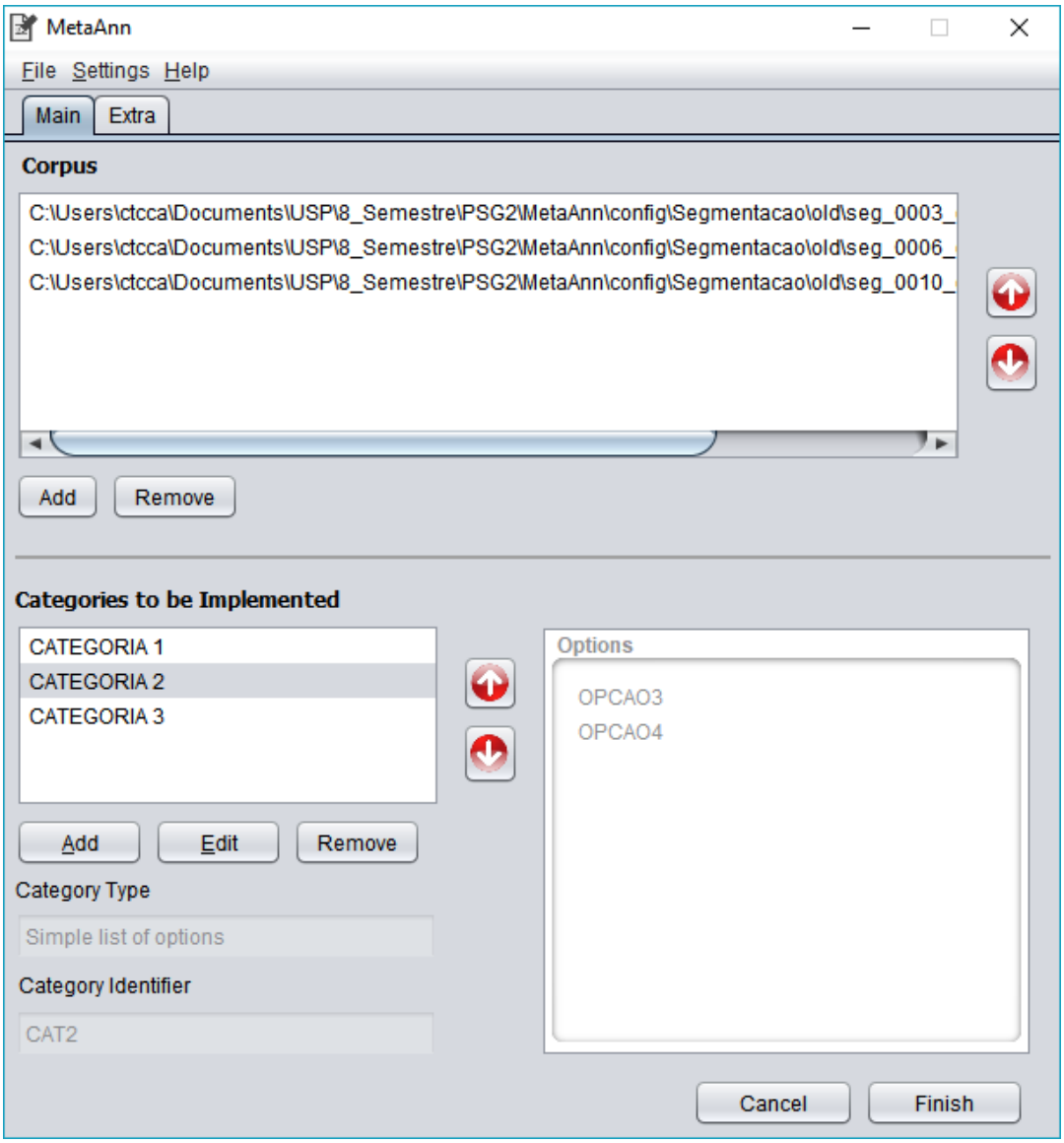
4. Settings and Help

4.1 Changing the system language

The “Settings” menu has an option to change the system language. To do this, just click on "Settings", point the cursor to "Language" and choose the desired language:



After selection, the system translates the entire context, except for the settings made by the user. The menu item for the current language is inactive, since there is no need to translate the system into the same language.

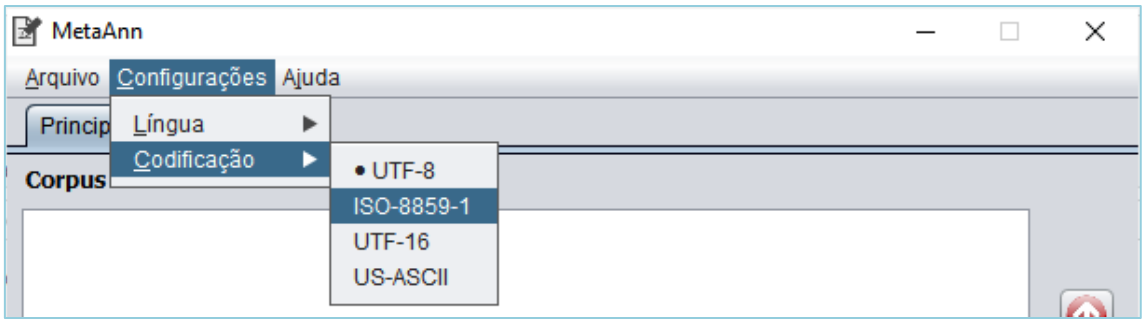


4.2 Changing the default file encoding

MetaAnn assumes that the files that make up the corpus use the standard UTF-8 character encoding. However, if the user is working with files with another encoding, it is possible to configure different encodings.

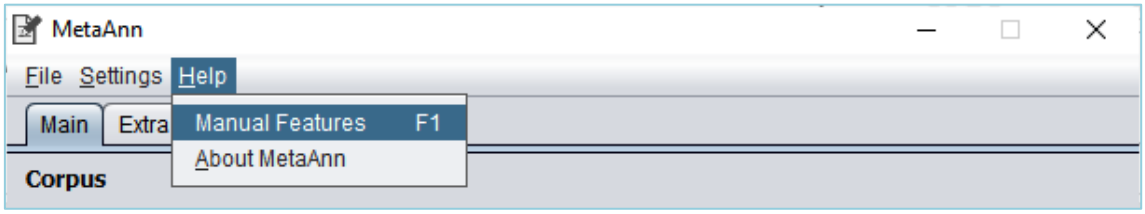
In addition to UTF-8, MetaAnn also supports ISSO-8859-1, UTF-16 and US-ASCII encodings. To change the default encoding, just click on the "Settings" menu, then "Encoding", and select one of the available encodings.

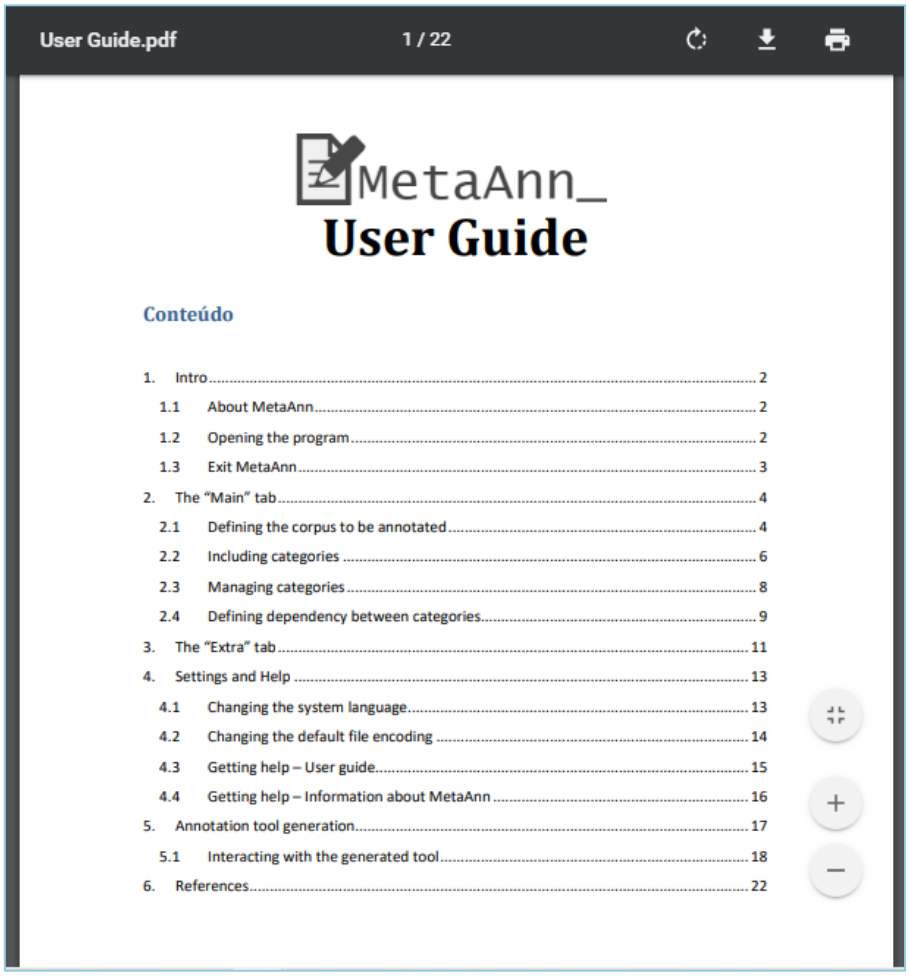
The generated tool will use the selected encoding to read the corpus files and display the information.



4.3 Getting help – User guide

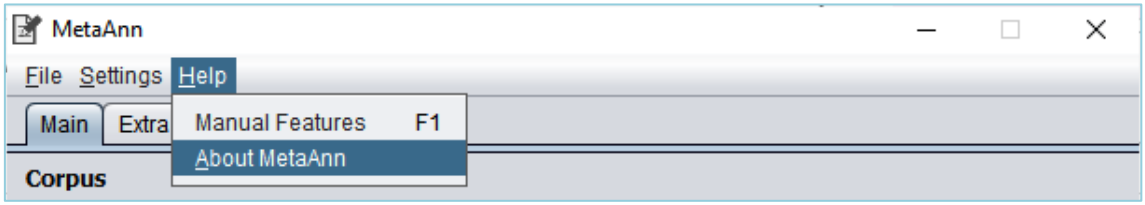
By clicking on the Help menu and then on “Manual Features”, the user guide will open with the standard PDF reader of the current computer.

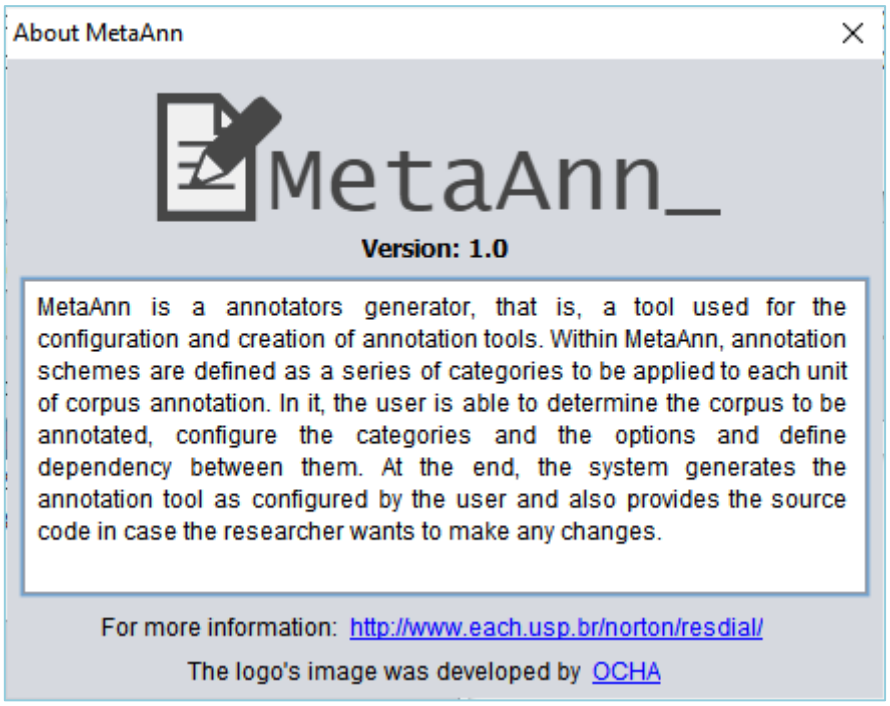




4.4 Getting help – Information about MetaAnn

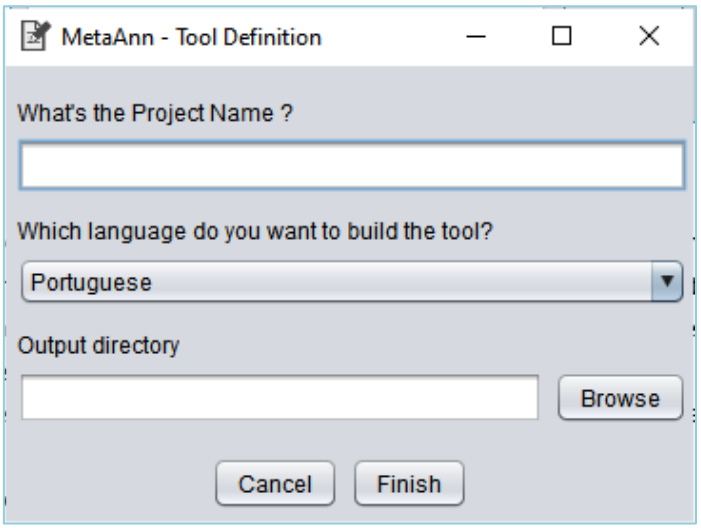
To obtain information about the system, just click on the "Help" menu and then on the "About MetaAnn" item. A window containing information about the system will appears:



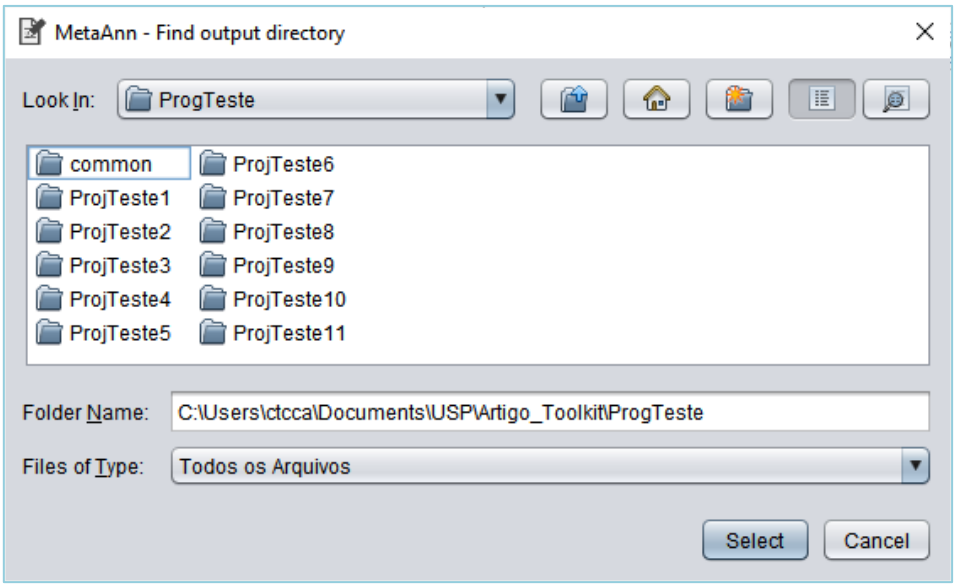


5. Annotation tool generation

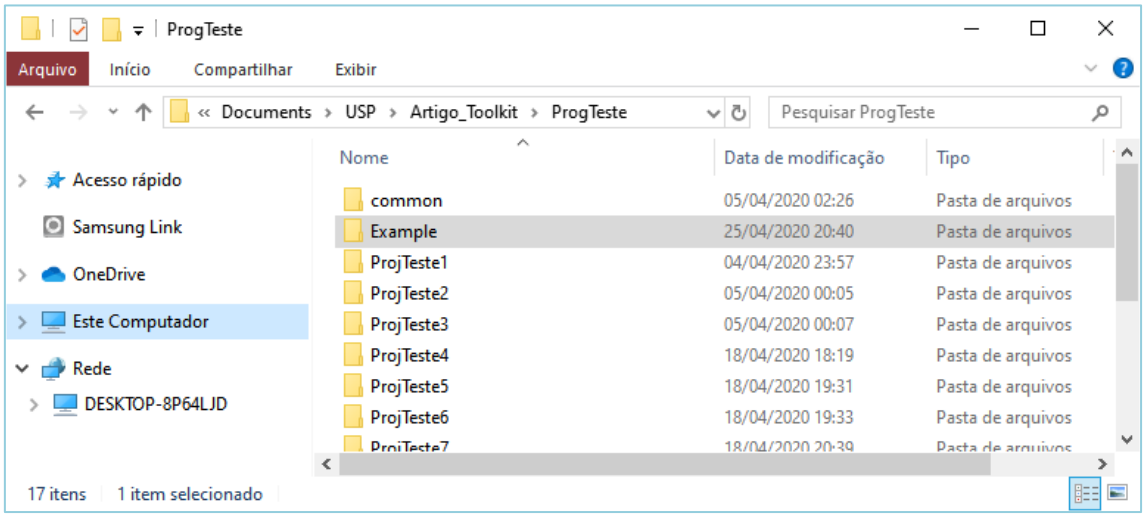
After you finish configuring the annotation tool, it's time to generate it. To do so, just click on the “Finish” button at the bottom of the main window, and fill in the information for the window that appears:



All fields in this window are required. The project name is the name that will appear in the title bar of the generated tool, and in the directories that will be generated to allocate the project. The user also has the option to choose which language the annotation tool will have: Portuguese or English. Finally, it is necessary to select the directory where the annotation tool files will be located. To do this, just click on the “Browse” button and, in the window that opens, choose the output directory.

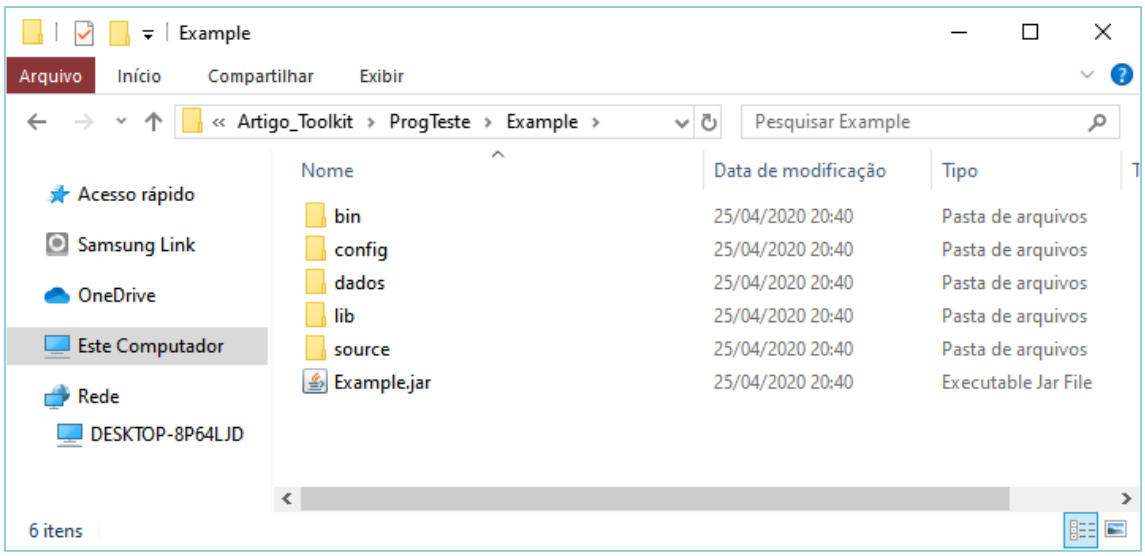


When clicking on the “Finish” button, the system will build the annotation tool and allocate it in the output directory determined by the user. After the generation is complete, the system returns to the main window. The annotation tool will be in the given folder:



5.1 Interacting with the generated tool

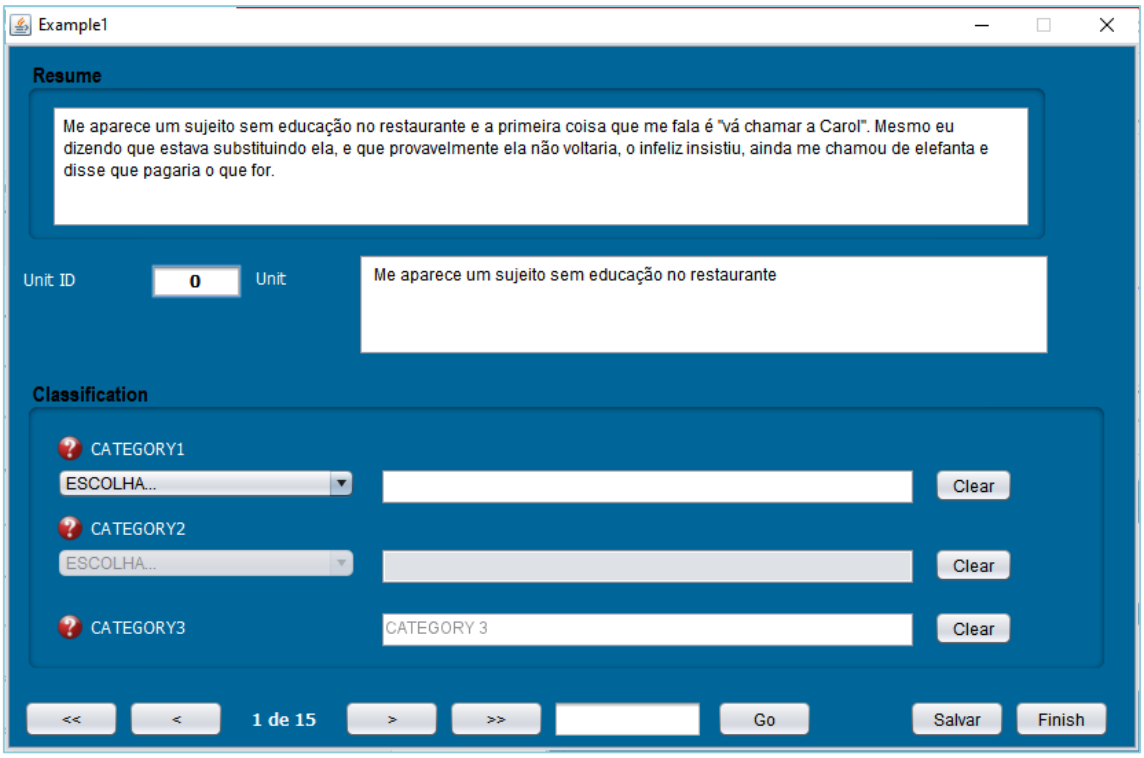
The generated tool directory has the following structure:



The “source” folder contains the source code files of the generated program; the “bin” folder contains the compiled classes of the source code. The “config” directory has configuration files, about internal configuration of system. The “lib” directory has some external libraries used by the application. Finally, the folder called

“dados” has the xml text files of the *corpus* that will be annotated, defined by the user on MetaAnn configuration.

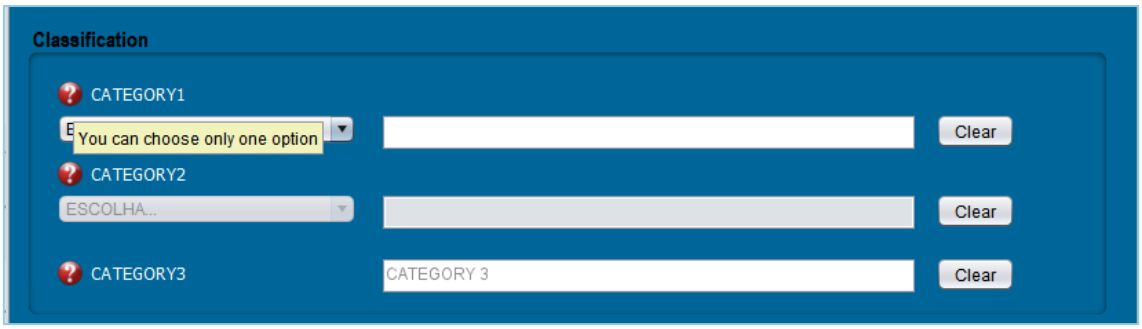
To use the tool, is not necessary to know or access any of these folders. This is only necessary if the researcher is interested in making changes to the program. To start the annotation, just double click on the *jar* file with the name of the project:



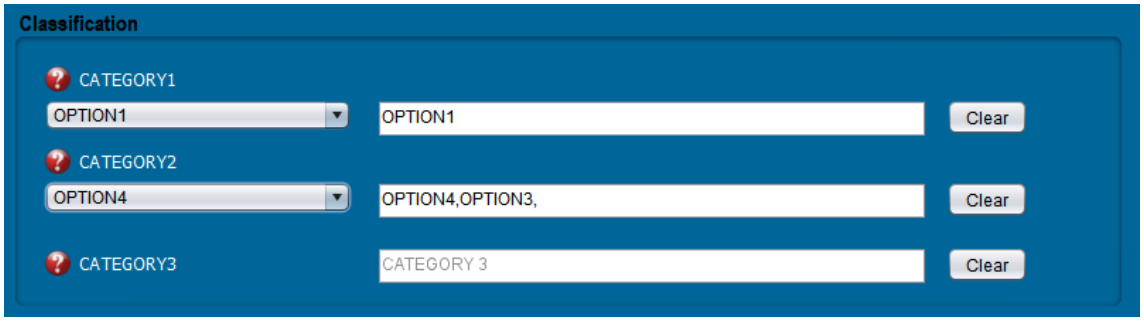
Three sections compose the generated annotation tool: the current corpus and current segment data, the classification section and the action buttons section. The first section contains the full text of the file in the corpus and, below, data related to the current segment: the Unit Id, the segment text and any additional information configured by the researcher in MetaAnn.

In order to classify each segment, the classification section groups all the categories defined by the researcher through MetaAnn in the second section. Categories of the type "Mutually Exclusive Options" and "Multiple selection list" are classified through the selection list, on the other hand, categories of the type "Free Text" have their value typed directly in the text field next to the category title.

Next to the name of each category, a help icon helps the user to classify each segment according to the determined category:



A category dependent on another is only released for filling when the category on which it depends is filled. “Multiple selection list” categories can contain multiple values. The category text defined in MetaAnn stays within the “Free Text” category field until the user starts typing the data in that field.



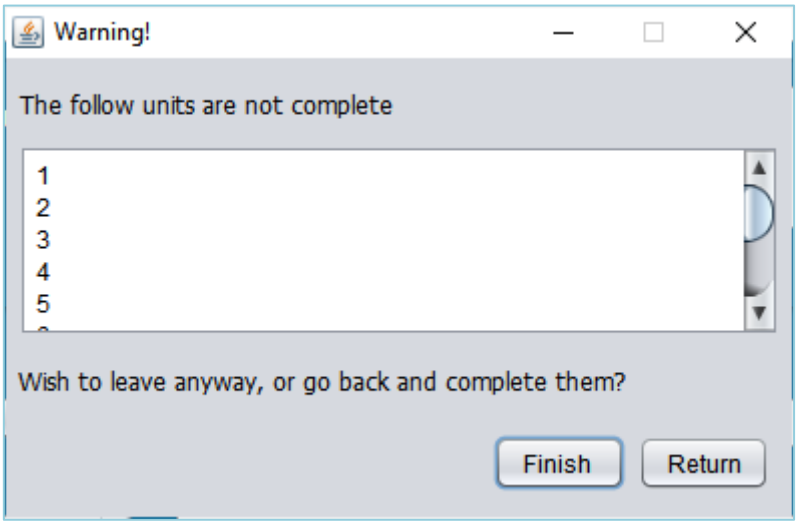
For list categories, if the user wants to clear the classification, he can select the option "CHOOSE..." within the list of options. For all categories, the button “Clear” in front of the box with the classification clears the classification.

The buttons that manage navigation compose the last section. The set is formed by the following buttons:

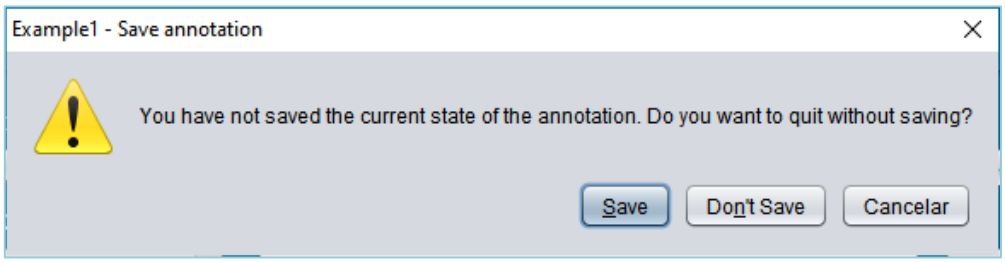
- << - Go back to the first segment
- < - Go back to the previous segment
- > - Advance to the next segment
- >> - Advance to the last segment
- Go - Navigate to the segment whose number (ordinal) is defined in the field beside
- Save - Saves the current annotation status
- Finish - Quit the tool

When exiting the tool, if all segments are not properly annotated, a window indicating which units of the corpus have not been fully classified is presented, giving

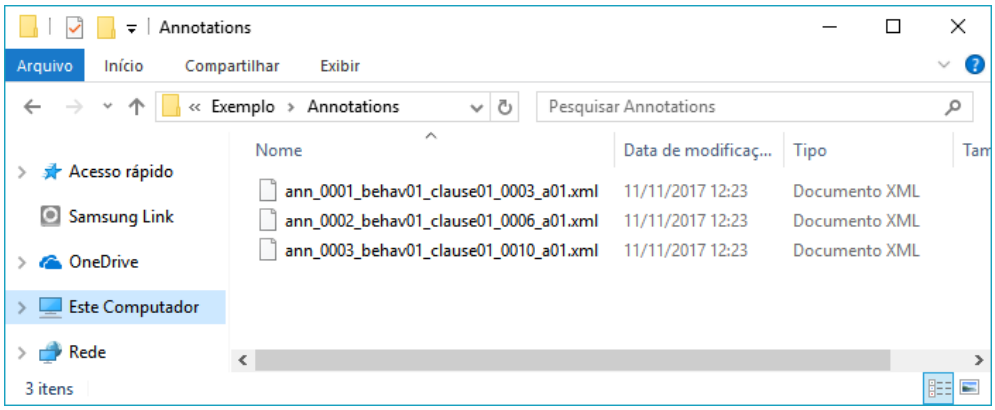
the user the option to continue with the generation of the annotation files or return to the classification.



If the user is trying to leave without saving the status of the annotation, he will view a message, giving the option to save and exit, exit without saving or cancel the close and return to the annotation task:



When the user saves an annotation, an “Annotations” folder is created, in the same directory as the annotation program. For each classified corpus file, an annotation result file will be created, defined according to ROMAN (2012).



6. References

ROMAN, N. T. **ResDial – Descrição da Codificação (v.1.0).** 2012. Retrieved from: <http://ppgsi.each.usp.br/arquivos/RelTec/PPgSI-001_2012.pdf>.

MISSÃO, T. E. I.; ROMAN, N. T. MetaAnn: Um gerador de ferramentas para anotação de textos. In: **Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology (STIL 2013).** Fortaleza, CE, Brazil: [s.n.], 2013, p. 11-20. Retrieved from: <http://www.lbd.dcc.ufmg.br/colecoes/stil/2013/002.pdf>.