

FromThePage transcription and annotation tool user guide

FromThePage is a browser based transcription and annotation tool. It works with the following browsers: *Mozilla Firefox*, *Chrome*, *Opera*.

FromThePage is useful, if one has text documents in image format and wants to transcribe and annotate them. The transcription and annotation result will be saved to a database in XML format.

Here are the instructions to follow in order to use **FromThePage**:

1

You have to go to the URL <http://10.8.83.147:3000>



2

You have to sign in (top right corner of the page).

There are 4 types of user rights: [administrator](#), [collection owner](#), [transcriber](#) and [guest](#).

The [collection owner](#) is a person who creates a collection: he uploads the scanned text to be transcribed and defines the annotation tags.

The [transcriber](#) transcribes and annotates the texts.

The [guest](#) is only allowed to make a few transcriptions.

The [collection owner](#) receives his identification details from the administrator.

The following 9 steps, from **3** to **11**, have to be executed by the [collection owner](#) in order to set up a new collection.

3

In order to start a new transcription project, you have to click on [Start A Project](#). In the top right corner of the page you will see an [Actions](#) menu. Click on [Create a Collection](#).



4

Once you have created a collection, you can upload scanned files to be transcribed.

Start A Project Your Works

Upload PDF or ZIP File

Using this form you can import your page images into FromThePage. You should select a target collection to import into and attach a ZIP or PDF file containing page images to be used for your project.

Coco

Click to browse a file... Browse

- ZIP files may contain folders containing images, PDFs, or folders containing pdfs.
- Each folder will be treated as a different document, so do not mix pages from different documents in the same folder.
- Each PDF will be treated as its own document, so do not split pages from the same document among more than one PDF.
- For example a ZIP file with 3 images, 2 PDFs, and 1 folder containing 5 more images would create 4 works: the top level images in one, each PDF in their own work, and a last work containing the 5 images from the folder.

Upload File

5

You also have to define the set of transcription tags. Go to the *Subjects* tab and create the first category.

FROM THE PAGE

Coco

Overview Statistics Subjects

In the top right corner of the page you will see an *Actions* menu that will let you add new categories and define their styles and types. You have to create **root categories**.

Dashboard Collections FAQ Signed In As collection_owner

Coco

Overview Statistics Subjects Settings Export Collaborators Add Work

Categories my_first_category

my_first_category

There are no attributes for the category selected.

This category has no description.

description.

Actions

- Rename Category
- Add Child Category
- Add Root Category**
- Delete Category
- Assign Category Scope
- Define Category Style
- Define Category Description
- Define Category Attributes
- Define Category Attribute Values
- Define Attribute Sequences
- Apply All Categories Changes
- Discard All Styles
- Delete All Categories

About Terms & Conditions Privacy

Policy

Then you have to fill in all the forms listed in the **Actions** drop down menu for each category and in the **same order as** they are listed in the **menu**:

6

The **Assign Category Scope** menu allows you to decide if the category will be usable only in the **simple mode**, only in the **advanced mode** or in both modes.

The **simple mode** allows you to add a number of attributes to each category and all of these attributes will receive a value during annotation. None of them will be skipped.

The **advanced mode** offers the possibility to choose a number of compulsory attributes for each category, all of which will receive a value, like in the simple mode. The advanced mode also offers you a possibility to fill additional attributes in case some other attribute was given a certain value.

Example:

For example, we may need to make detailed descriptions of word forms. We may need to describe all the verb forms of a text by defining their tense, voice, person etc. In this case, the value of an attribute may allow or not allow the existence of another attribute. Let's take the example of an **adjective**. In German a singular **adjective** has 3 declension paradigms, a masculine, a feminine and a neutral, while a plural **adjective** has only one declension paradigm. Therefore, if the value of the **number** attribute is **singular**, the annotator needs to define the **gender**, but if the value of the **number** attribute is **plural**, he doesn't need to define the **gender**. This use case demonstrates the necessity to define complex relations between attributes and their values.

Here is an example of a sequence of attributes:

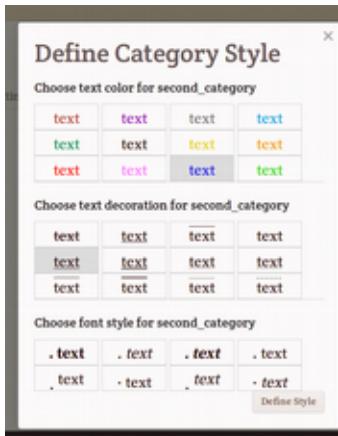
part of speech -> verb -> tense (present indefinite -> person)

adjective -> number (sg -> gender)

*noun -> gender
number*

7

The **Define Category Style** menu will let you choose the colour and decoration of the tagged text.



It is **not compulsory** to define a category's style, but it is **highly recommended**. If no particular style is defined, the annotator won't know if a certain piece of text has been tagged or not.

Tags can overlap. This should be taken into consideration while choosing categories' styles. For example, if one wants to tag parts of speech and syntactic roles, he could use different colors for different parts of speech and different types of underlining for different syntactic roles.

Buttons corresponding to styles that have already been chosen for categories of this collection appear in gray.

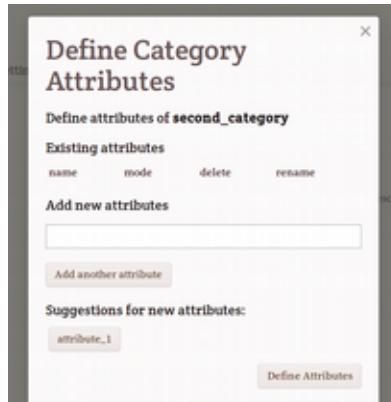
After choosing the style it is important to push the button *Define style*. Otherwise, all the menu information will be lost.

8

The *Define category description* menu allows you to provide a description of a category and its attributes, anything you want the transcriber to know. The description will appear on the *Subjects* page next to the category name.

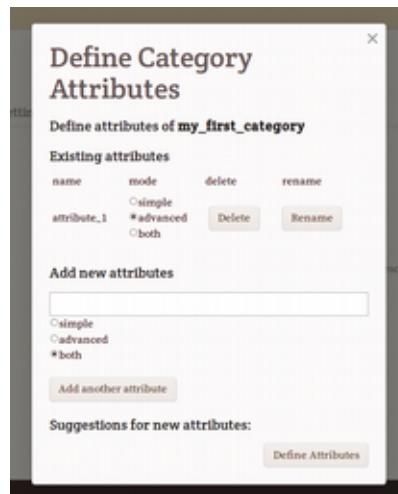
9

The *Define Category Attributes* menu will let you define different attributes of the same category. You can either add attributes to a category or not, it's **not compulsory**.



If other categories of the same collection used in the same scope already have some attributes, those attributes will appear at the bottom of the dialog window as **suggestions for attributes** of the current category. If you push the button with the suggested attribute name, it will appear in the existing attributes' list and will become one of the attributes of this category.

If a category is used in both modes, it's possible to restrict its attribute usage to only one mode.



If the category is used in only one mode, its attributes will also be used only in the same mode as the category.

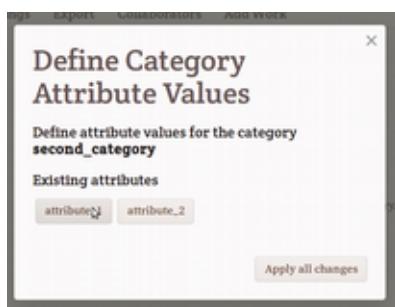
This menu also allows to **delete** and **rename** attributes.

After making all the necessary changes you have to confirm them by pushing the *Define Attributes* button. Otherwise, all the changes will be lost.

10

The *Define Category Attribute Values* menu will let you define values for the attributes you have just defined.

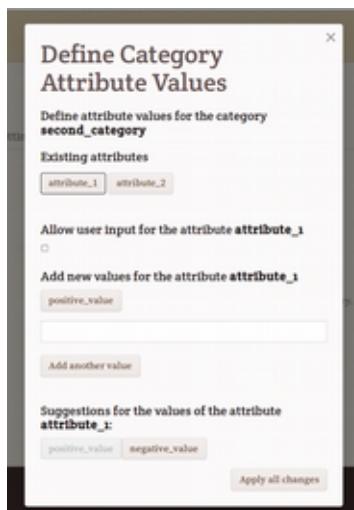
In order to define an attribute's values, you have to click on the button with the attribute's name.



A menu will appear.

You can either make a list of possible values or let the annotator type a new value while transcribing or both. In order to let the annotator type a value while transcribing, you have to check the *Allow user input* check box.

Each attribute must have a possibility to receive a value. Otherwise, the tool will show an error message and the annotator won't be able to use this category.



If **other categories** of the same collection have the **same attributes** as the current category, values defined for these attributes for other categories of the same collection will appear as **suggestions** for values of the attributes of the current category.

Example:

Lets imagine you need to tag errors and distinguish syntactic and orthographic errors. In this case you need to create a root category **error** and define an attribute **type**. Then you need to create 2 possible values for it: **synt** and **ort**.

If you also want the annotator to give the correct form, you have to introduce a special attribute for that (for example, **correct_form**), and check the check box **allow user input**.

Here is an example outcome:

```
<error-id28 tagcode="1496064202555" type="ort" correct_form="these" class="medium-error-id28">thees</error-id28>
```

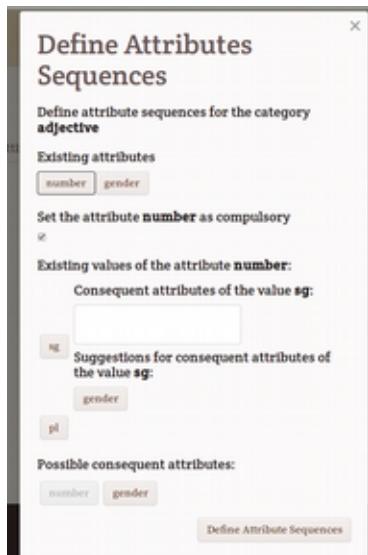
11

If you have made a category available in the **advanced mode**, you have to define attribute sequences using the **Define Attribute Sequences** menu.

After clicking on the attribute's name you will see a menu letting you define attributes that the transcriber will have to fill if he chooses certain values of the current attribute.

First, you have to decide which attributes will be compulsory. At least one attribute must be compulsory. Otherwise, this category tags will have **no attributes at all in the advanced mode**. The choice of compulsory attributes is made by checking the **Set the attribute X as compulsory** check box.

Each of the predefined values of the chosen attribute is listed below (**sg** and **pl** on the picture). By clicking on one of them, we open a dialog that allows us to define “consequent” attributes of this value. All the possible attributes are listed at the bottom of the dialog window (**Possible consequent attributes**). These are attributes of the same category. Just under the white area where the consequent attributes will be listed there may be a list of **suggestions for consequent attributes** of the chosen value, coming from the same values of attributes of other categories of the same collection. If other categories of the same collection don't have attributes with the same values or these values don't have consequent attributes, the suggestions list will be empty.



You have to be very attentive **not to create infinite loops**. For example, if a *singular* value of the number attribute induces the gender attribute, a *neutral* value of the gender attribute induces a case attribute and the *Nominative* value of a case attribute induces a number attribute, the annotator will never be able to finish the annotation. The program doesn't verify the validity of the sequences and won't warn you, if you create an infinite loop.

After defining all the necessary sequences for this category, you have to push the *Define Attribute Sequences* button. Otherwise, all the defined sequences will be lost.

12

When you have defined all the necessary categories, their styles and types, you have to click on *Apply All Categories Changes* in the *Actions* menu in order to see the style applied. If you or the annotators don't see the categories' styles, they have to empty their browser's cash.

If new styles are defined when the transcribers are already working on the collection, in order to see these new styles, they have to **delete the site's cookies field “collection_id” from their browser** and empty their browser's cash.

Before starting the transcription process, transcribers can get acquainted with the tagging categories by visiting the *Subjects* page of the collection where all the categories, their descriptions and their attributes are visible.

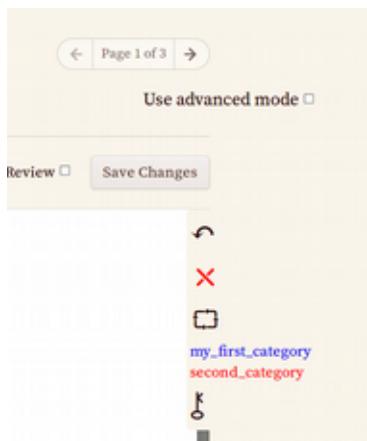
The screenshot shows a web interface for managing collection subjects. At the top, there are three tabs: 'Overview', 'Statistics', and 'Subjects', with 'Subjects' being the active tab. Below the tabs, there are two main sections. On the left, under 'Categories', there is a list of categories: 'my_first_category' and 'second_category'. On the right, for the selected category 'second_category', there is a detailed view. This view includes the category name 'second_category', a section titled 'Attributes:' with items 'attribute_1' and 'attribute_2', and a descriptive text at the bottom stating 'This is the description of the second category.'

13

Now that all the categories, their attributes and values have been defined, the transcription process can start.

In order to start transcribing, go to the *Works* tab and click on a work, then choose the page you want to transcribe. You will see the scanned page and an empty space for the transcription next to it. On the right side you will see a list of categories you have just defined.

The annotator has to **choose the annotation mode**. It is the simple mode by default, but he can also choose the advanced mode by checking a check box at the top right corner of the window.



The simple mode lets the annotator use categories having the simple mode scope, the advanced mode lets the annotator use categories having the advanced mode scope.

Both modes allow the use of categories having the “both” flag. However, the list of attributes that have to be filled may not be the same. As you remember, attributes of such categories can themselves receive a scope: **simple**, **advanced** or **both**. Attributes having received the “simple” flag are only usable in the simple mode, attributes having received the “advanced” flag are only usable in the advanced mode and attributes having received the “both” flag are usable in both modes.

In the simple mode all the attributes defined for this category have to be given a value during annotation, while in the **advanced** mode it depends on the defined **sequences**. For example, if no attribute is defined as **compulsory**, no attributes will be proposed during annotation.

There are 2 ways of inserting tags:

- 1 First insert a tag and then type the text inside.
- 2 First type the text, then highlight a portion of it with a mouse and tag it.

The tag can be added either using **hot keys** or using **buttons** on the right side of the screen.

Here is the list of **hot keys**:

Alt+C : insert a tag or tag a highlighted piece of text

When you press **Alt+C**, a pop up menu will appear offering a choice of categories available in this transcription mode. You can choose one of them with the mouse, with an arrow key or by typing the first letters of the category name. Once you have chosen the category, if the latter has attributes, each of them will appear accompanied by a list of possible values or a text input field where you will be able to type a new value. When all the attributes will have their values assigned, the tag will be inserted into the text.

Alt+X: get the cursor out of the current tag in order to continue typing outside the tag

Alt+R: make a pop up menu disappear. In this case no tagging is performed.

Alt+N : delete a tag from the portion of the text where the cursor is

A menu listing all the tags present at the cursor position will appear, the user has to choose the one he wants to delete.

This functionality is also available through the “cross” looking button on the right side of the screen.

Alt+M : modify a tag from the portion of the text where the cursor is

A menu listing all the tags present at the cursor position will appear, the user has to choose the one he wants to modify. Once the tag chosen, its attributes and their values will appear. The user will be able to change the values of these attributes. He won't be able to add or delete attributes. In order to do this, he should delete the tag and insert another one in its place. And if the category configuration defined by the collection owner will suggest other attributes as consequences of the newly chosen values (in the advanced mode), the attributes finally present in the tag will be different.

If the user modifies attribute values of a certain tag, the program doesn't check if they are compatible with each other according to the system of sequences of attributes and values defined by the collection owner (in the advanced mode). It means that in case of annotation errors the result may be slightly different from the one expected by the collection owner. The annotation quality is the responsibility of the annotator.

The tag modification functionality is also available through the “electric circuit” looking button on the right side of the screen.

All these **hot keys may be modified**. In order to change them, you have to push the “key” looking button in the right part of the window.

14

Transcriptions are automatically saved every 3 minutes. However, when you have finished transcribing, you have to push the button **Save changes** to be sure that all the changes have been saved.

If there are errors in the XML code of the page (for example, due to accidental mouse mismanipulation), an error message will appear and the page won't be saved. If you see an error message, it means that there is a mistake in the work you have done during the last 3 minutes.

Here is an example of a text tagged with ***FromThePage***:

```
<?xml version='1.0' encoding='UTF-8'?>
<page>
    <greeting_id8 class='medium-greeting_id8' mode='1' tagcode='1501593576726'>
        <adjective_id6 class='medium-adjective_id6' mode='1'
tagcode='1501595124365'>Dear</adjective_id6>
        Reader
    </greeting_id8>,
    <main_clause_id13 class='medium-main_clause_id13' mode='1' tagcode='1501596400575'>
        <subject_id11 class='medium-subject_id11' mode='1'
tagcode='1501595065526'>I</subject_id11>
        <predicate_id12 class='medium-predicate_id12' mode='1'
tagcode='1501595071144'>
            <verb_id10 class='medium-verb_id10' mode='1' person='1'
tagcode='1501595105142' tense='present ind'>am</verb_id10>
            glad
        </predicate_id12>
    </main_clause_id13>
    <subordinate_clause_id14 class='medium-subordinate_clause_id14' mode='1'
tagcode='1501596407114'>
        you
        <verb_id10 class='medium-verb_id10' mode='1' person='2'
tagcode='1501595080621' tense='past ind'>read</verb_id10>
        this tutorial.
    </subordinate_clause_id14>
    <br/>
    <closing_id9 class='medium-closing_id9' mode='1' tagcode='1501593710728'>
        <adjective_id6 class='medium-adjective_id6' mode='1'
tagcode='1501595133332'>Best</adjective_id6>
        regards
    </closing_id9>
</page>
```