

Annotation Scheme - 冰激凌

Ayden Janssen, Jonas Walliser and Shiyao Zhang

Because we want to use sequence classification, each word is a markable. We use three attributes: entity, name and alignment.

The entity attribute uses the BIO format and therefore has the values B, I, and O. It shows if the annotated word refers to a character or not.

The name attribute is used to identify which character the word refers to and acts kind of like an ID. It has no fixed values, as each fairy tale has different characters. A word that is not an entity has the value X.

The alignment attribute can have the values *good*, *neutral*, *bad* or *X*, and shows how the character is portrayed in the story. The alignment for one character can also change over the course of the fairy tale. The *X* value is only used if the word is not an entity.

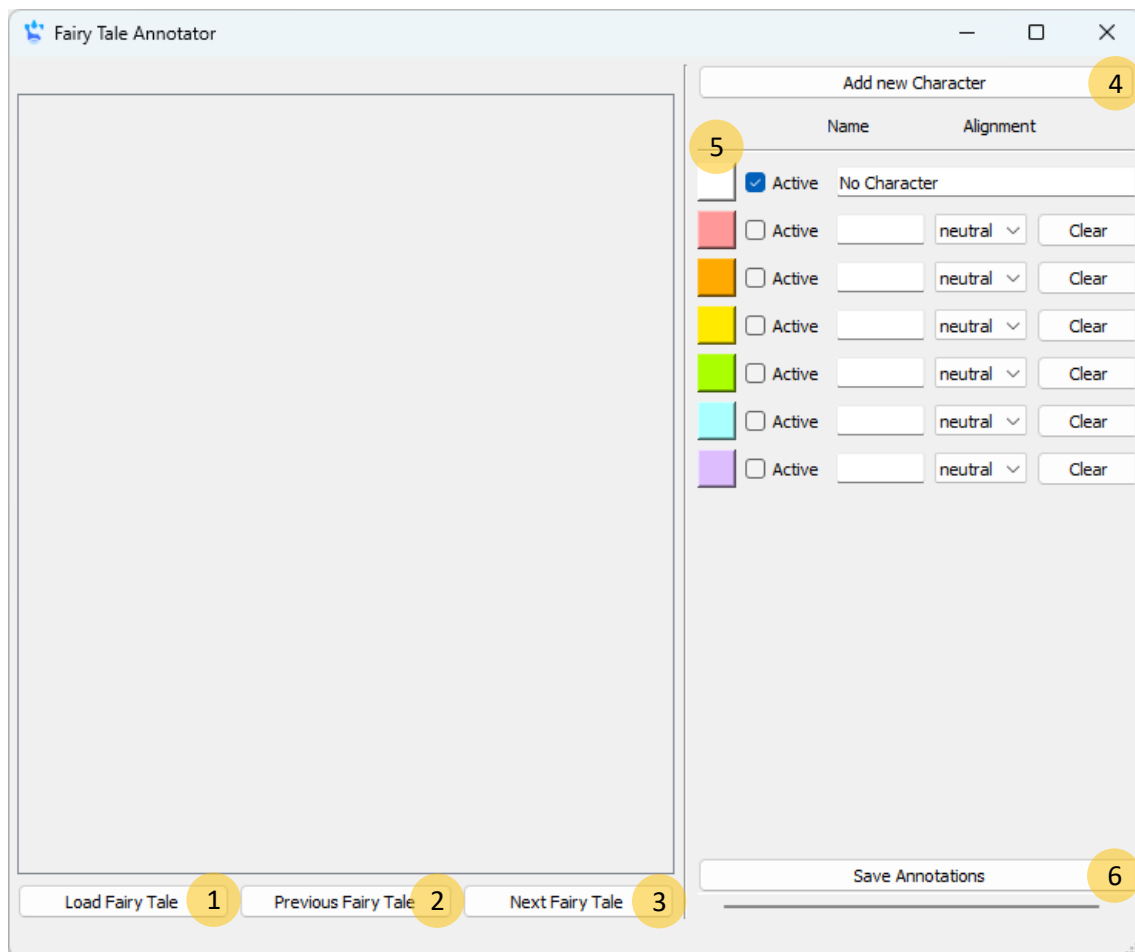
The annotations are saved as csv files for each fairy tale:

Word	Entity	Name	Alignment
There	O	X	X
was	O	X	X
once	O	X	X
a	B	Father	neutral
rich	I	Father	neutral
man	I	Father	neutral
whose	O	X	X
wife	B	Mother	good
lay	O	X	X
sick	O	X	X
and	O	X	X
when	O	X	X
she	O	X	X
felt	O	X	X
...

Annotation Manual

The annotation application

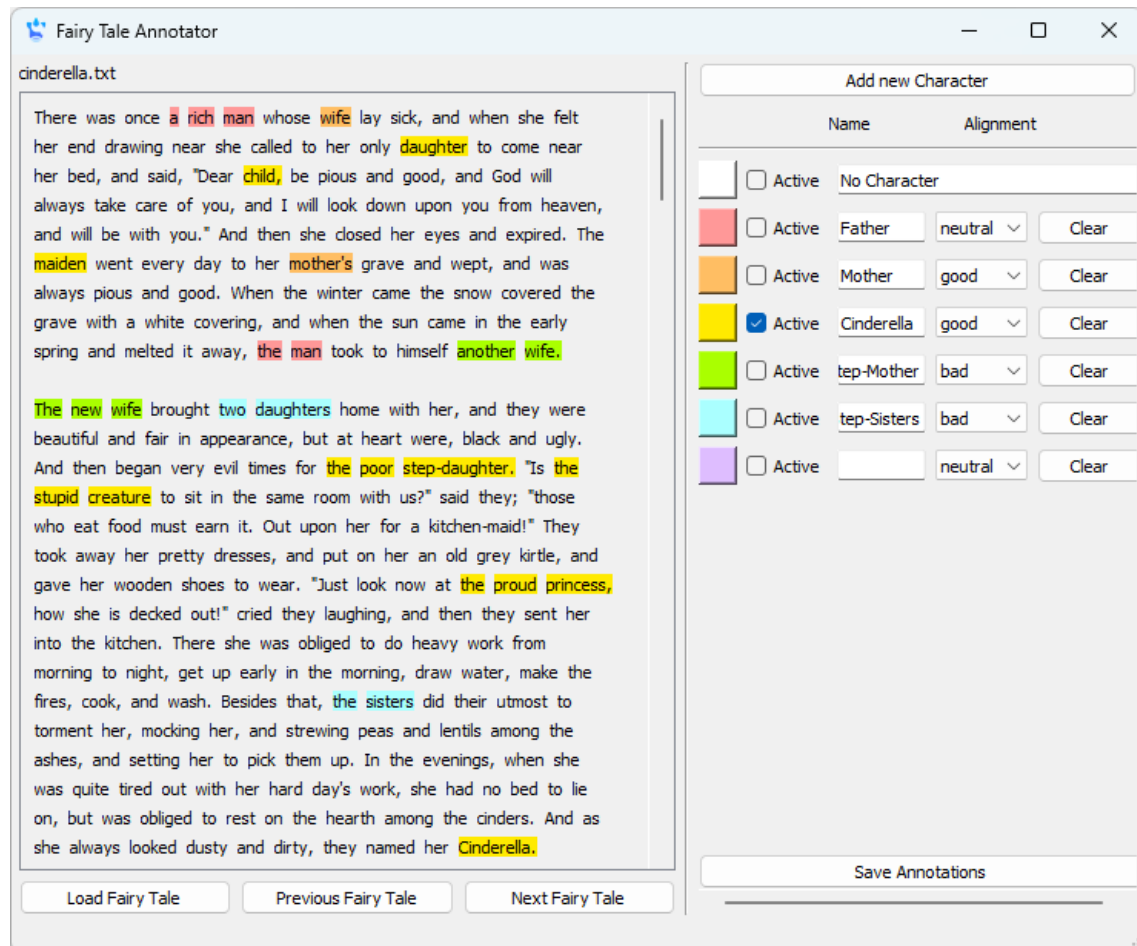
When opening the application, no fairy tale will be loaded initially. To load a fairy tale, click the **Load Fairy Tale** button (1) and choose one of the textfiles from the corpus. By using the **Previous Fairy Tale** (2) and **Next Fairy Tale** (3) buttons, you can go through the files in alphabetical order in the current directory, without having to load each file individually. You can add more characters to the list on the right by using the **Add new Character** button (4). The color for each of the characters can be changed by clicking on the colored button (5) (colors are only used for visualization). Annotations can be exported using the **Save Annotations** button (6).



Annotating a fairy tale

After loading a fairy tale, the name of the file will be displayed on the top left, and the content will be displayed in the box on the left half. You can then fill out the character names and their alignments on the right side. If the character has an actual name, that name should be used. If they are only referred to by a specific noun (e.g. mouse or father), the noun should be used as the name. When using nouns, do not add an article in front of the noun. The options for alignment are *good*, *neutral* and *bad*. You can annotate words in the text, by setting one character to active using its checkbox and then clicking on the text. The text will be highlighted in the color

that is assigned to the character. All annotations for one character can be removed with the **Clear** button. You can also remove annotations by activating *No Character* and clicking on the text, this works like an eraser. The image below shows an example annotation for the fairy tale “Cinderella”.



Exporting the annotations

The annotations have to be exported for each fairy tale individually. You can save the annotations for the current fairy tale using the **Save Annotations** button. A csv file will automatically be created in an annotations folder within the folder of the textfile.

If you have unsaved changes and try to exit the program or change the fairy tale, you will also be given a notice to save your work before closing. Unsaved annotations will be lost upon loading a different file or closing the program.

