Annotation Guide - TRANSMATT

13/05/2020 LENTSCHAT Martin

This Guide is designed to familiarize annotators with the use of the WebAnno tool. The version used here is provided by the ICO team of IATE joint research unit and can be accessed at: https://ico.iate.inra.fr/webanno/login.html.

As an annotator you have to log in as follow:

username firstname.name

password firstname.name

Getting started with the WebAnno tool

Once logged in you can access the annotation function from the main menu. Clicking on the **Annotation** button opens a new window where you can select a project you are involved in (left column) and then the document you wish to work on (right column).

The colour of the document names indicates the document work progress status:

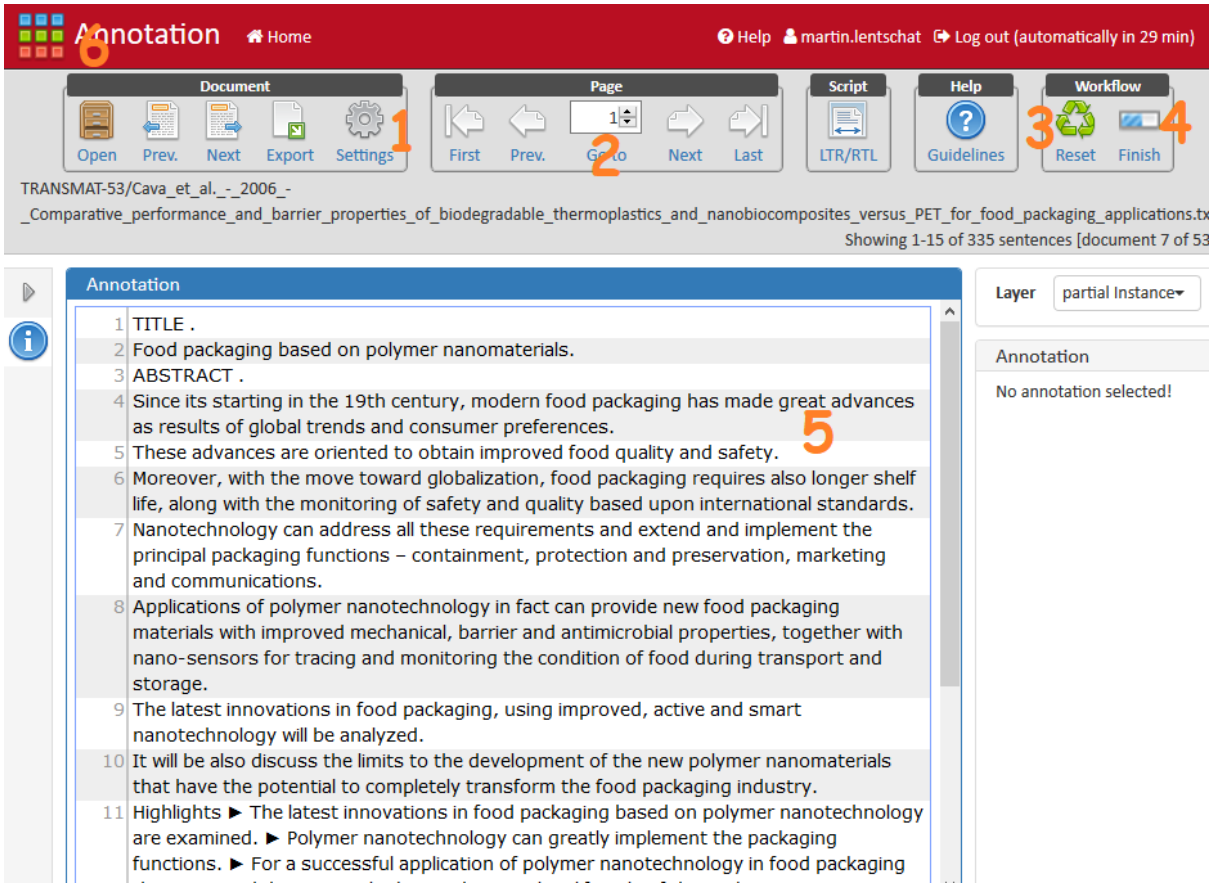
black annotation pending

blue annotation in progress

red annotation completed

The core functionalities on the annotation screen are:

1. **Settings** allows you to adjust the **text** display.
2. buttons in the **page** tab enable quick navigation through the text.
3. the **reset** button deletes all annotations on the active document.
4. the **finish** button is clicked when the annotation is completed.
5. the text you will annotate.
6. the **logo** may be clicked to return to the home page.

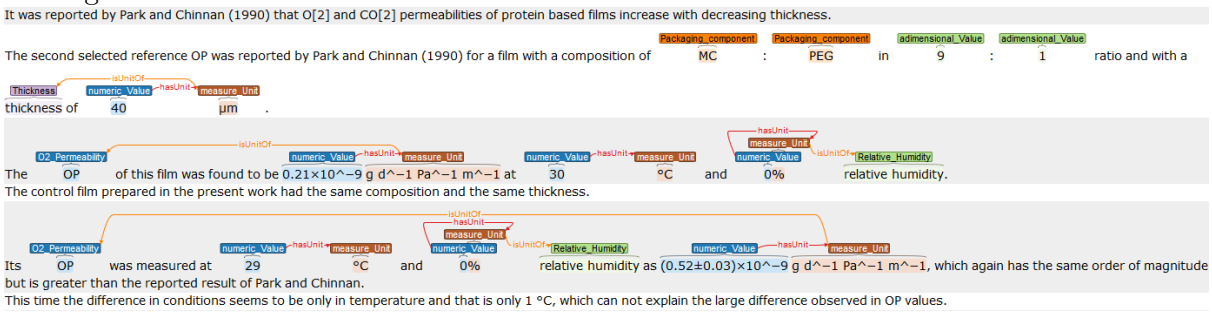


Here we seek to annotate scientific articles by locating certain *Arguments* regarding *N-Ary Relations* pertaining to permeability. This concerns several types of information present throughout the texts.

Annotation creation

To create an annotation, simply double-click a word or highlight a sequence. This selection should be as precise as possible (down to the character). The *Label* is selected in the right tab once the annotation has been created.

Some *Labels* will need to be connected by dragging the selected annotation to link it with another one. You should pay attention to the linkage direction. These linked *Labels* may not necessarily be in the same sentence, this is not problematic because a connection can be made over a long distance.

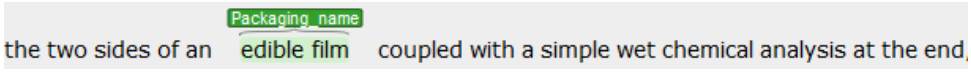


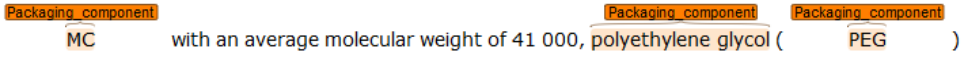
Once an annotation has been created, it can be modified or deleted by double-clicking on it again.

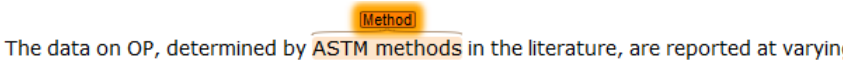
The following sections detail the selection criteria for *Labels* and connections.

Labels and symbolic argument connections

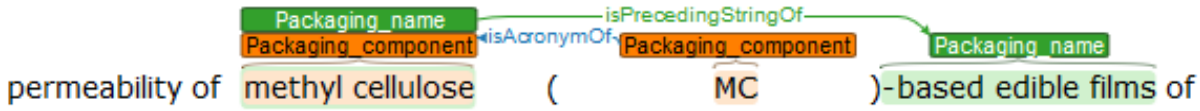
Here the symbolic arguments to be annotated consist of terms (i.e. words or word sequences) related to food packaging and the permeability measurement method. We distinguish the name used by the authors to designate and categorize a package from the components used to create it.

`Packaging_Name` 

`Packaging_Component` 

`Method` 

As these terms are 'self-supporting', they inseldom need connections (except regarding component proportions, see below). The `isAcronymOf` connection may be used (this task is optional), to indicate links between terms and their acronyms. The `isPrecedingStringOf` connection allows you to link annotations of the same *Labels*, to indicate that they concern the same term. Annotation may also be overlapped if necessary.

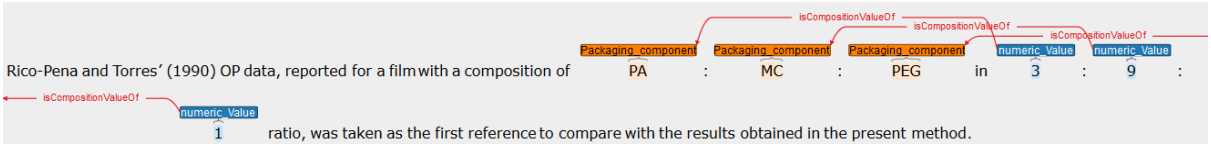


Packaging components and connections with their proportions

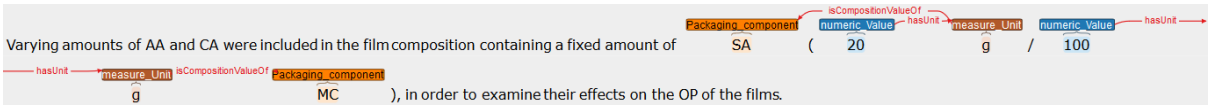
Packaging component contents are presented in documents in terms of: ratios and mass quantity. Two elements are always present (a numerical value and a packaging component), sometimes along with a measurement unit. The connection used to link the component and the numerical value or measurement unit is, when availabl, `isCompositionValueOf`.

Example :

- `Packaging_Component` and `numerical_Value`



- `Packaging_Component`, `numerical_Value` and `measure_Unit`



Labels and quantity argument connections

Instances of quantity arguments to be annotated are more complex and include multiple *Labels*. They always consist of a numerical value (`numeric_Value`) with a measurement unit (`measure_Unit`) and sometimes a term indicating its dimension :

`Thickness` the food packaging tickness, e.g. : *'thickness'*, *'thick'* ...

`Temperature` control temperature used during the permeability readings, e.g. : *'temperature', 'T', 'temp.' ...*

`Relative_Humidity` control humidity of the permeability measurements, e.g. : *'relative humidity', 'RH' ...*

`Relative_Humidity_Difference` similar to the previous, considered when measuring water vapour permeability (H2O_Permeability)

`Partial_Pressure_Difference` difference of pressure used in the permeability measurement, ex : *partial pressure difference*

`CO2-H2O-O2_Permeability` terms related to permeability measurements, e.g. : *'oxygen permeability', 'WVP', '[CO2] perm' ...*

Caution : Values not involved in the permeability relation, e.g. the packaging preparation temperature, storage humidity etc. are not concerned.

Connections indicate that the different elements constitute the same argument instancess. There are several of them:

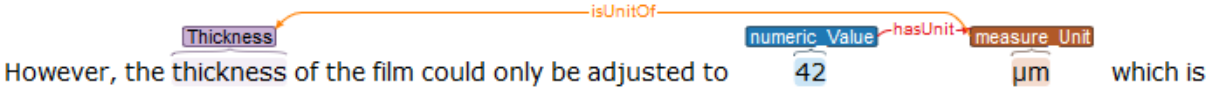
`hasUnit` goes from `numeric_Value` and link it to `measure_Unit`

`isUnitOf` goes from `measure_Unit` and link it to one of the quantity term

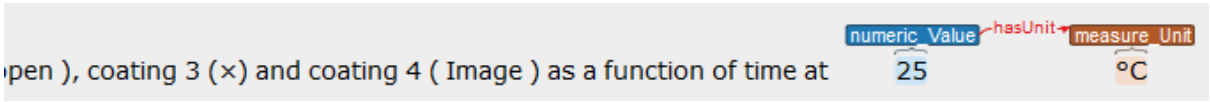
The presence of a term giving the dimension of the numerical value is important for disambiguation. It may sometimes not be present in the sentence but instead in the textual window (±1 sentence) around the measurement unit and numerical value. Yet some of the three ideally present elements are sometimes missing.

The possible cases for annotating the terms and their connections are therefore :

1 The numerical value is linked to the measurement unit, which in turn is linked to the quantity term.



2 The numerical value is linked to the measurement unit, but the quantity term is missing.



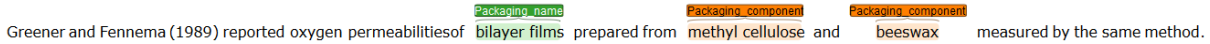
Citations

Authors of an article sometimes refer to other research articles for comparison with their work. In such cases arguments may be present, often relating to the packaging being discussed.

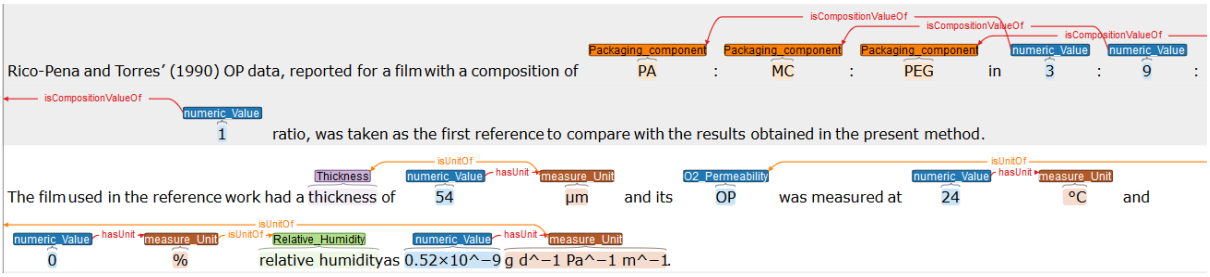
Caution : the information presented is often incomplete and not the focus of the article. Annotate only if several arguments are present, with `Packaging` and `Permeability` being mandatory.

Example :

- Invalid



- Valid



Tables

Tables are sometimes present in the documents. These consist of a legend, note or quotation and a table. The latter is easily recognizable because each line of the table is then preceded by the word *'Line'* and *pipes* (*'—'*) are included to differentiate the cells.

Given the complexity of this task for annotators and the annotation client, when annotating tables it is recommended that **only text lines be annotated** (such as the caption) while **not annotating the table cells content**.

Recommandations

- take the time to understand what you are reading and read the text several times. The text in the vincinity of an item often indicates whether an annotation is warranted.
- if you do not know whether a word should be included in an annotation or not then annotate broadly.
- annotate according to what you think is right, not what you think is consensual.
- if you are doubtful about a case and it is not covered in the present Guide, then do not annotate.
- the same information that appears many times in a text (e.g. packaging name) should only be annotated once or twice. There is no need to repeat an annotation that you are sure you have already created.