Export

After performing a query, you can click on More and then Export to export your results.There are five different exporters available: WekaExporter, CSVExporter, TokenExporter, GridExporter, SimpleTextExporter as can be seen in Figure 1. Each of them will be described in the next section.
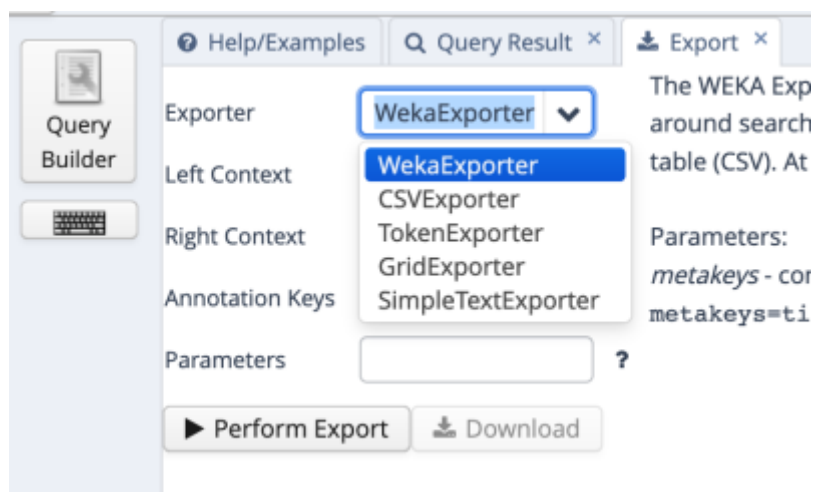

Figure 1: Different exporters and additional options for the export

WekaExporter

This exporter is very specific for the data mining application Weka.

CSVExporter

This exporter creates one line per result. In this line, you see the text you queried for as well as all the annotations available on the token level. Depending on the sub-corpus, these are the token itself as well as PoS annotations.

TokenExporter

This exporter is intended for smaller corpora than ours. Using our (sub-)corpora it often hangs even at very small queries. We recommend not to use it.

GridExporter

This exporter is the most versatile one, since you can choose the annotations that you want to export. Figure 2 shows an example in which one token to the left and one to the right are exported as well as the whole message, the message ID, the token queried for and the age_range (not visible). Additionally, the meta key for the chat ID is exported.

Figure 2: Example of a GridExport

The resulting output starts as follows:

Figure 3: Results of the export (extract)

As you can see in Figure 3, each result is preceded by a number starting with 0. You then see all the annotation keys selected in Figure 2 in the selected order: whole message, message ID, token (your query is in the center, in this case *demain* plus the left and right token that you selected with the left and right context), age_range and then the chat ID selected with `metakeys=doc`.

If you leave the field "Annotation keys" empty, your export contains all annotations available on the

token level. Very often this is too much, so it is better to make a selection as shown above.

Simple text exporter

This exporter creates a list of the token(s) you queried for with the number of preceding and following tokens you selected in the options. The results are numbered. No additional information is exported.

Additional options

Next to the type of export, you have the option "Left and right context", which is the same for all export formats. Here, you can define the number of entities to be exported to the left or right of your search query. The entity is in the same unit as your query, i.e. if you query for tokens, you can select the number of tokens to be shown, while if you query for messages, this is the number of messages.

The other options, "Annotation keys" and "Parameters" depend on the export format and are explained to the right when you select an export option.

Under "Parameters", you can add annotations that pertain to the chat. More precisely, you can add all annotations that are listed under "Meta Annotations" in the information display per sub-corpus. To list that kind of information, you use the form: `metakeys=doc` to display the chat ID. More values can be added with commas.

Once you click `Perform Export`, the system will create the export in the memory and you can click `download` to have it downloaded to your own computer.

Exports are very hungry in resources, thus, it might take a while to create an export or the server might even hang. The simpler your query, the less problems you have. **Hint**: instead of formulating a complex RegEx query, it might be more useful to create several simpler queries and then merge the resulting files.

From:
https://sms.linguistik.uzh.ch/ -

Permanent link:
https://sms.linguistik.uzh.ch/02_browsing/05_support/01_export

Last update: **2022/06/27 07:21**