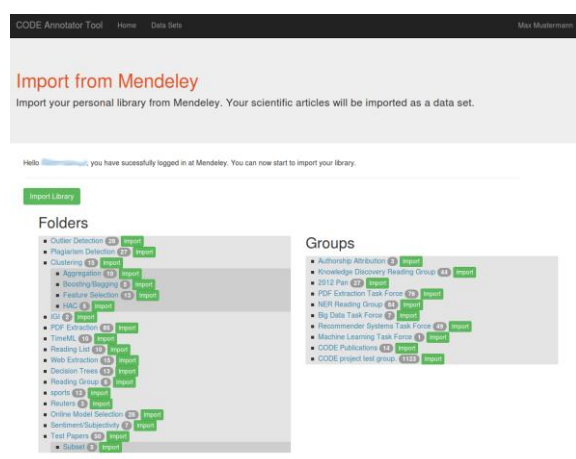# CODE Annotator Tool

Import data sets and concept hierarchies; intuitively annotate documents; apply machine learning techniques and finally share models with others


❶ Import the Data

Every user can create and store an arbitrary number of data sets. The most elegant way to create a new data set is to **import a collection of scientific articles from a Mendeley account**. One can import the whole library, a single folder or sub-folder, or all documents shared within a group.

❷ Import a Concept Hierarchy

Once a data set is selected, one or more models can be applied to this data set. For example, a model in the form of a **concept hierarchy can be imported from Mindmeister**. Such a model consists of a set of entity classes and relations between them. If the model already contains example names for the concepts, the data set is automatically annotated by the tool.
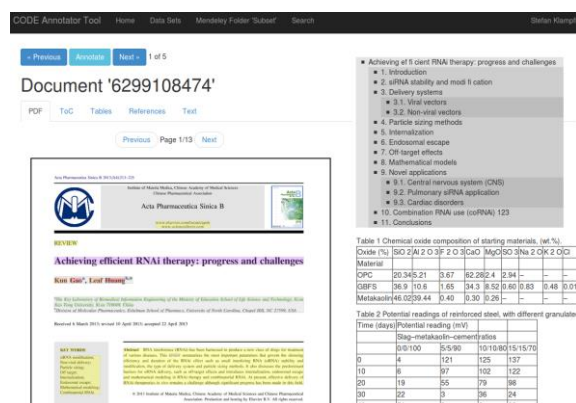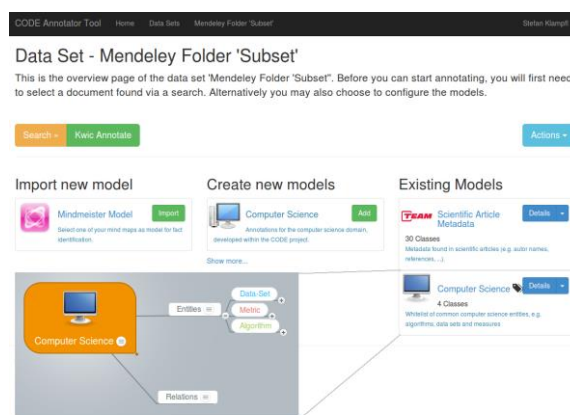
❸ View the Data & Structure

The document structure of scientific articles in PDF format is analysed by a number of tailored information extraction algorithms. The extracted **document structure, table of contents, tabular data, and references** are displayed. One can navigate through the pages of the document, highlighting the labelled components of the document.
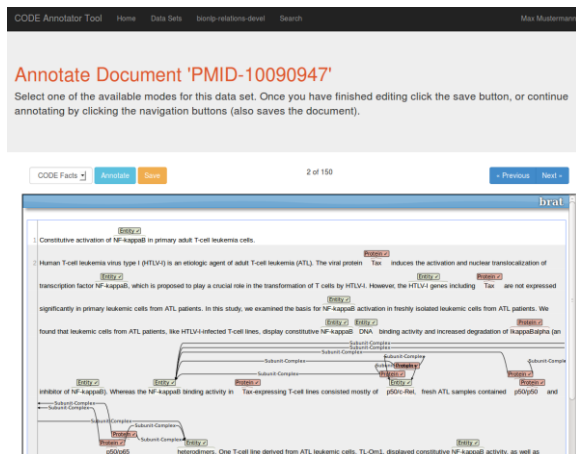
❹ Visually Annotate the Documents

For each model applied to the current data set one can view the **annotated entities** and, if available, the **relations between these entities**. These annotations can be edited or deleted, or new annotations can be added. The user is guided by a visual tool, optimised for mouse input.

❺ Quick Annotations on Search Results

Alternatively to the visual annotation, user's may choose a search driven annotation mode. The **annotation candidates are visualised using a keyword in context (KWIC) interface**. Here the user interface is tailored towards a keyboard based interaction.
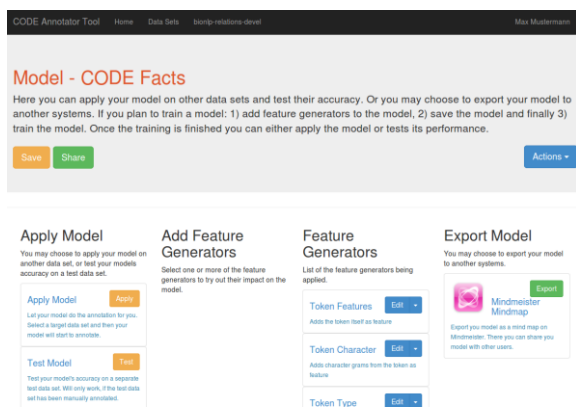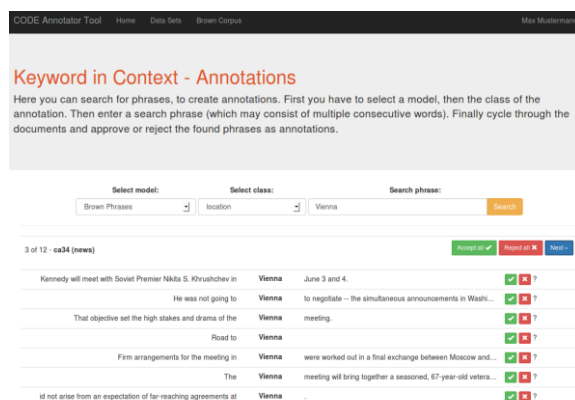
❻ Train & Share

Once annotations have been made one can train a model using **machine learning algorithms**. The user annotations thereby are the input to the process. The trained model can be applied or evaluated on any compatible data set.

Finally the trained model can be **shared with other users** and systems.

Useful Links:

CODE Project Website: http://code-research.eu/
CODE Annotation Tool: http://code-annotator.know-center.tugraz.at/

Contact: Roman Kern, Know-Center, rkern@know-center.at

Source code for the annotator tool is dual licensed and also available under an open-source license (AGPL3).