

CODE Annotator Tool

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

The screenshot shows the 'Import from Mendeley' section of the tool. At the top, it says 'Hello [User Name], you have successfully logged in at Mendeley. You can now start to import your library.' Below this, there are two main sections: 'Folders' and 'Groups'. The 'Folders' section lists several imported items like 'Outer Detection', 'Phagosome Detection', 'Clustering', etc., each with an 'Import' button. The 'Groups' section lists 'Knowledge Discovery Working Group', '2012 Pan', 'PDF Extraction Task Force', etc., also with 'Import' buttons.

2 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.

1 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.

This screenshot shows the 'Data Set - Mendeley Folder "Subset"' page. It includes sections for 'Import new model' (with a 'Mindmeister Model' option), 'Create new models' (with a 'Computer Science' model listed), and 'Existing Models' (listing 'Scientific Article Metadata' and 'Computer Science'). A diagram illustrates the 'Computer Science' model structure, showing entities like 'Data-Set', 'Metric', and 'Algorithm' connected to a central 'Computer Science' node.

The screenshot shows a document page with the title 'Achieving efficient RNAi therapy: progress and challenges'. It includes a table of contents, a sidebar with research highlights, and two tables of data extracted from the document.

Oxide (%)	SiO ₂	Al ₂ O ₃	CaO	MgO	Na ₂ O	K ₂ O	ClO ₄
Material	20.34(5.21)	3.67	62.26(2.4)	2.94	—	—	—
OPC	20.34(5.21)	3.67	62.26(2.4)	2.94	—	—	—
OPF	20.34(5.21)	3.67	62.26(2.4)	2.94	—	—	—
Metacelite	46.02(9.44)	0.40	3.0	0.25	—	—	—

Time (days)	Potential reading (mV)
0/0/100	5/5/90
0	121
10	97
20	19
22	3
40	51

3 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.

The screenshot shows a document titled "Annotate Document 'PMID-10090947'". It includes a toolbar with "CODE Facts", "Annotate", and "Save" buttons. The main content area displays a complex biological diagram with various proteins (e.g., Tax, p50, p65) and DNA binding sites. Annotations are shown as boxes and arrows indicating interactions between these components.

5 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.

4 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.

The screenshot shows a search interface with fields for "Select model: Brown Phrases", "Select class: location", and "Search phrase: Vienna". Below is a table of search results:

	Vienna	Action
Kennedy will meet with Soviet Premier Nikita S. Khrushchev in	June 3 and 4.	<input checked="" type="checkbox"/> <input type="checkbox"/>
He was not going to	Vienna to negotiate – the simultaneous announcements in Washi...	<input checked="" type="checkbox"/> <input type="checkbox"/>
That objective set the high stakes and drama of the	Vienna meeting.	<input checked="" type="checkbox"/> <input type="checkbox"/>
Road to	Vienna were worked out in a final exchange between Moscow and...	<input checked="" type="checkbox"/> <input type="checkbox"/>
Firm arrangements for the meeting in	Vienna meeting will bring together a seasoned, 67-year-old vetera...	<input checked="" type="checkbox"/> <input type="checkbox"/>
The	Vienna id not arise from an expectation of far-reaching agreements at	<input checked="" type="checkbox"/> <input type="checkbox"/>
	Vienna .	<input checked="" type="checkbox"/> <input type="checkbox"/>

The screenshot shows a dashboard for "Model - CODE Facts". It includes sections for "Apply Model", "Add Feature Generators", "Feature Generators", and "Export Model". Buttons for "Save" and "Share" are also present.

6 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).