

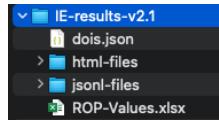
ROP ENTITY ANNOTATION GUIDELINE

Yinghao Li; September 13, 2021

Note: check [this link](#) for the latest version of this document.

CHECKING RESULTS

The result folder should contain several folders and files.



- Html_files: the folder that contains HTML result files with polymer names and sentences containing ROP property values highlighted.
- Jsonl_files: the folder that contains jsonl result files (will be introduced later)
- Dois.json: the dois of the articles of positive results.
- ROP-Values.xlsx: A table that contains the positive sentences, articles and extracted values from them. It also has hyperlinks to the original articles as well as the local result files in "HTML_files" folder.

A	B	C	D	E	F	G	H	I
article_idx	sentence_id	DOI	Tc	ΔH	ΔS	sentence	local_file	reliability
2	1	10.1021/ma501541g	130 °C	-41 kJ mol ⁻¹	-101 J mol ⁻¹	The values of Tc, as well as of the polymerization enthalpy ΔH p and entropy (ΔS p) were determined from the slope and intercept of the linear regression line, values of ΔH p = -41.0.1021&sl;ma501541g.json	10.1021&sl;ma501541g.json	7.4
3	1	10.1021/ma501541g		-41 kJ mol ⁻¹	-101 J mol ⁻¹	From the slope and intercept of the linear regression line, values of ΔH p = -41.0.1021&sl;ma501541g.json	10.1021&sl;ma501541g.json	7.4
4	1	10.1021/ma501541g	130 °C			Tc for unimolecular monomer concentration was then determined to be 130 °C. 10.1021&sl;ma501541g.json	10.1021&sl;ma501541g.json	7.4
5	1	10.1021/ma501541g	104 °C			This is not far from the value reported for TrMA polymerized via RAFT in toluene. 10.1021&sl;ma501541g.json	10.1021&sl;ma501541g.json	7.4
6	1	10.1021/ma501541g		-55 kJ mol ⁻¹		The determined value of ΔH p is considerably higher than that for MMA (about 10.1021&sl;ma501541g.json)	10.1021&sl;ma501541g.json	7.4
7	1	10.1021/ma501541g			-100 J mol ⁻¹	The calculated ΔS p value lies on the higher end of the interval -120 to -100 J mol ⁻¹ . 10.1021&sl;ma501541g.json	10.1021&sl;ma501541g.json	7.4
8	2	10.1021/ma0114887		0.5 kJ mol ⁻¹	2 J mol ⁻¹ K ⁻¹	The thermodynamic parameters for the polymerization reaction were determined from the fit of ln([MDO]e/[MDO]s) vs 1/T (Figure 1) to a line and yielded ΔH = 10.1021&sl;ma0114887.json	10.1021&sl;ma0114887.json	7.3
9	2	10.1021/ma0114887		5 kJ mol ⁻¹	2 J mol ⁻¹ K ⁻¹	A plot of ln([MDO]e/[MDO]s) vs 1/T (Figure 1) was fit to a line and yielded ΔH = 10.1021&sl;ma0114887.json	10.1021&sl;ma0114887.json	7.3
10	2	10.1021/ma0114887	90 °C			Using initial MDO concentrations of 3 M in toluene, we measured [M]e for the 10.1021&sl;ma0114887.json	10.1021&sl;ma0114887.json	7.3
11	2	10.1021/ma0114887		-12 kJ mol ⁻¹	-42 J mol ⁻¹	As expected, the polymerization of MDO is characterized by a typical entropy 10.1021&sl;ma0114887.json	10.1021&sl;ma0114887.json	7.3
12	2	10.1021/ma0114887	101 °C			For [MDO]0 = 3 M, we calculate a Tc of 101 °C. 10.1021&sl;ma0114887.json	10.1021&sl;ma0114887.json	7.3
13	1	10.1021/jacs.6b07974		-5.9 kJ mol ⁻¹	-40.1 J mol ⁻¹	Next, a Van't Hoff plot of ln[M]leq versus 1/T gave a straight line (Figure S6) 10.1021&sl;jacs.6b07974.json	10.1021&sl;jacs.6b07974.json	6.5
14	3	10.1021/jacs.6b07974	221 K			Third, the ceiling temperature (Tc) was calculated to be 221 K (-52 °C) at [MB] 10.1021&sl;jacs.6b07974.json	10.1021&sl;jacs.6b07974.json	6.5
15	3	10.1021/jacs.6b07974	-52 °C			Third, the ceiling temperature (Tc) was calculated to be 221 K (-52 °C) at [MB] 10.1021&sl;jacs.6b07974.json	10.1021&sl;jacs.6b07974.json	6.5
16	5	10.1021/jacs.6b07974	147 K			Third, the ceiling temperature (Tc) was calculated to be 221 K (-52 °C) at [MB] 10.1021&sl;jacs.6b07974.json	10.1021&sl;jacs.6b07974.json	6.5
17	3	10.1021/jacs.6b07974	-126 °C			Third, the ceiling temperature (Tc) was calculated to be 221 K (-52 °C) at [MB] 10.1021&sl;jacs.6b07974.json	10.1021&sl;jacs.6b07974.json	6.5
18	3	10.1021/jacs.6b07974	-126 °C			As expected, the same analysis performed at lower temperatures reduces the 10.1021&sl;jacs.6b07974.json	10.1021&sl;jacs.6b07974.json	6.5
19	3	10.1021/jacs.6b07974	-150 °C			As expected, the same analysis performed at lower temperatures reduces the 10.1021&sl;jacs.6b07974.json	10.1021&sl;jacs.6b07974.json	6.5
20	3	10.1021/jacs.6b07974		1.8 kcal/mol		As expected, the same analysis performed at lower temperatures reduces the 10.1021&sl;jacs.6b07974.json	10.1021&sl;jacs.6b07974.json	6.5
21	3	10.1021/jacs.6b07974		0.7 kcal/mol		As expected, the same analysis performed at lower temperatures reduces the 10.1021&sl;jacs.6b07974.json	10.1021&sl;jacs.6b07974.json	6.5
22	4	10.1021/ma021122+	135 °C			The ceiling temperature for the formation of the poly(ethylene)sulfone is the hi 10.1021&sl;ma021122+.json	10.1021&sl;ma021122+.json	6.4
23	4	10.1021/ma021122+	48 °C			We can estimate, however, that, according to the Clausius-Clapeyron equation 10.1021&sl;ma021122+.json	10.1021&sl;ma021122+.json	6.4
24	4	10.1021/ma021122+		-70.2 kJ mol ⁻¹		A previous experimental study 16 reported -70.2 kJ mol ⁻¹ in the gas phase and 10.1021&sl;ma021122+.json	10.1021&sl;ma021122+.json	6.4
25	4	10.1021/ma021122+		-82.8 kJ mol ⁻¹		A previous experimental study 16 reported -70.2 kJ mol ⁻¹ in the gas phase and 10.1021&sl;ma021122+.json	10.1021&sl;ma021122+.json	6.4
26	4	10.1021/ma021122+	25 °C			However, the initial spectrum was easily restored under evacuation even at 25 10.1021&sl;ma021122+.json	10.1021&sl;ma021122+.json	6.4
27	4	10.1021/ma021122+	135 °C			However, the initial spectrum was easily restored under evacuation even at 25 10.1021&sl;ma021122+.json	10.1021&sl;ma021122+.json	6.4
28	4	10.1021/ma021122+		-5 kJ mol ⁻¹		However, the analysis above of our activation energy data and thermochemical 10.1021&sl;ma021122+.json	10.1021&sl;ma021122+.json	6.4
29	4	10.1021/ma021122+		-5 kJ mol ⁻¹	219.3 J mol ⁻¹	Given that the entropy change for this reaction is primarily due to the removal 10.1021&sl;ma021122+.json	10.1021&sl;ma021122+.json	6.4
30	5	10.1021/je0206781	360 °C			An illustrative example of N-alkylation is the reaction of cyanuric acid (1 mol) 10.1021&sl;je0206781.json	10.1021&sl;je0206781.json	5.8
31	5	10.1021/je0206781	170 °C	112.5 kJ/mol		This was not found to be the case, as ΔH p values of 124.6, 125.6, and 112.5 kJ/mol 10.1021&sl;je0206781.json	10.1021&sl;je0206781.json	5.8
32	6	10.1021/acs.macromol.5b01592		1 kJ mol ⁻¹	2 J mol ⁻¹ K ⁻¹	The obtained ΔH p values were plotted as a function of reaction temperature. 10.1021&sl;acs.macromol.5b01592.json	10.1021&sl;acs.macromol.5b01592.json	5.7

TO VALIDATION A SENTENCE

- Go to ROP-Values.xlsx and go through the sentences/values top-down. The results are ordered according to their reliability scores. So, results at the top of the file are more “reliable” than results below.
- If a result is obviously correct/incorrect, you can directly add the values to the spreadsheet or discard it.

3. If it is hard to tell, you can go the local HTML file by clicking the corresponding link. The sentences are listed at the top of the file as well as highlighted in the body. You can also go to the original website by clicking the doi link for more information.

TO GIVE FEEDBACK

The **preferred** feedback method is manually annotated sentences/articles with NER labels and positive/negative sentence-level labels. You can also provide an unformatted feedback file in natural language.

The following section describes how to provide NER labels and positive/negative sentence-level labels.

DOCCANO

Doccano is a powerful open-source web-based NLP annotation tool. It has modern, stable and user-friendly annotation interface.

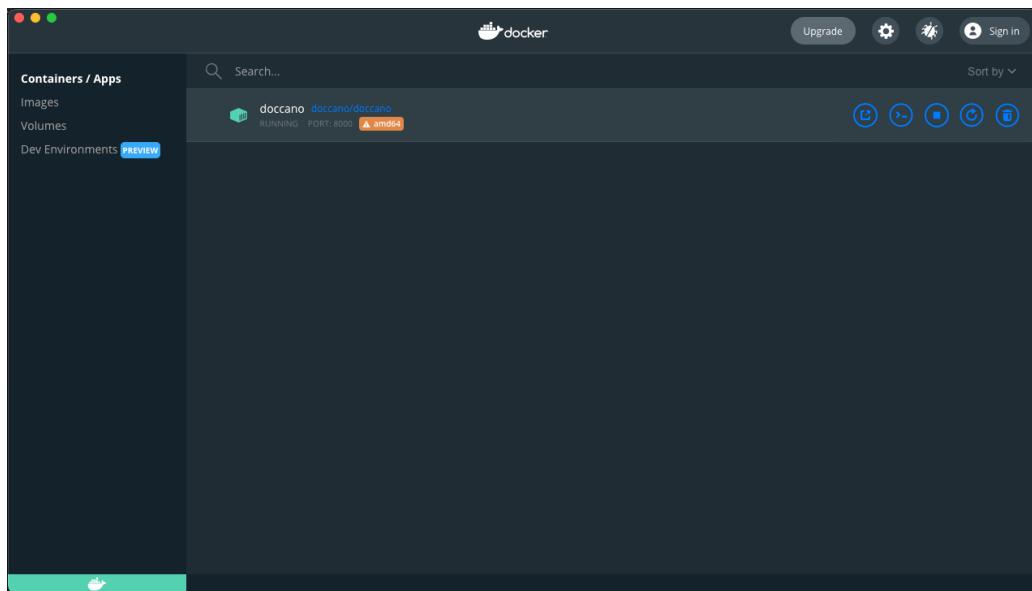
INSTALLATION

Recommended method: using Docker (<https://www.docker.com/>)

1. Download and install Docker: <https://www.docker.com/get-started>
2. Start Docker process.
3. Follow the instruction on <https://github.com/doccano/doccano> in “Docker” section to install Doccano in your Docker. Notice that you need to specify your username and password.
4. Start the docker and go to <http://127.0.0.1:8000/> to annotate data.

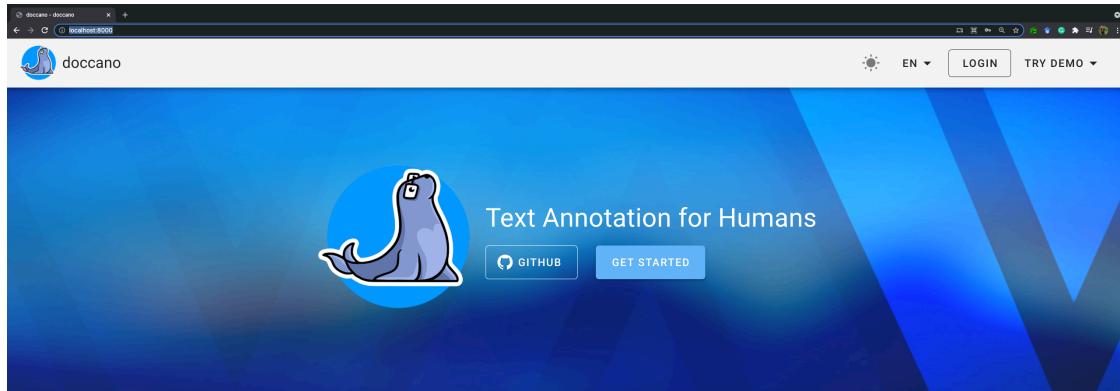
Other methods: Follow the instruction on <https://github.com/doccano/doccano>. Notice that other installation methods are not tested.

Your Docker UI (if the version of Docker you installed is with UI) should look like this if the doccano container is running:



CREATING PROJECT

Once you have successfully installed and ran image, you should be able to see the web page below with URL <http://localhost:8000/> or <http://127.0.0.1:8000/>



The best features



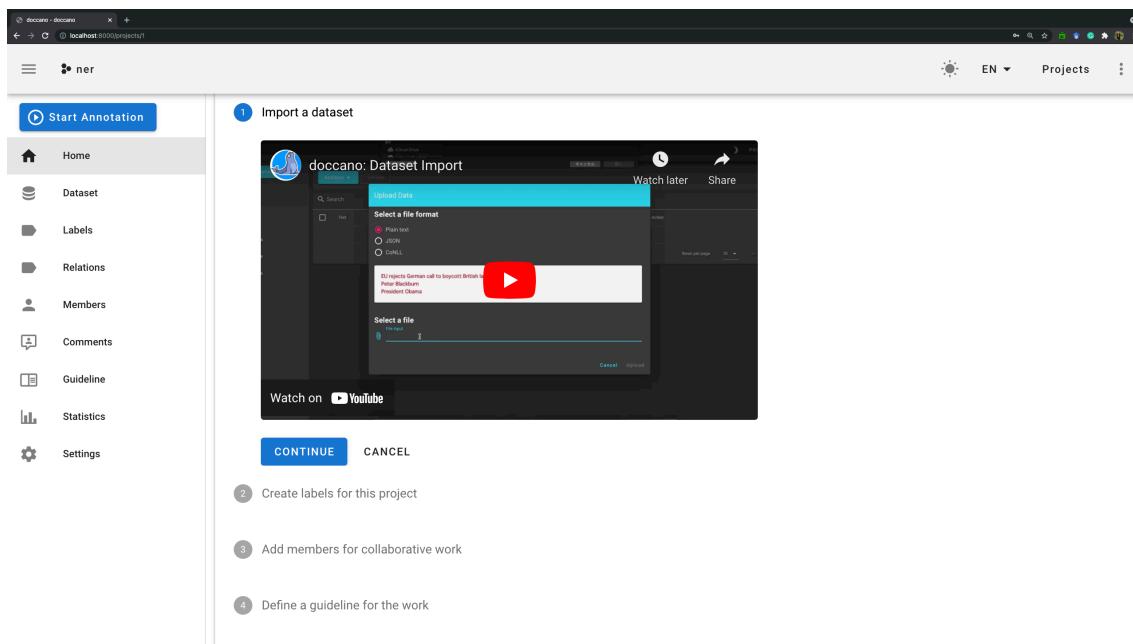
The first thing to do is logging into the system with the username and password you just specified.

A screenshot of a login form titled 'Login'. It has two input fields: 'Username' containing 'yli' and 'Password' containing a masked value. Below the fields is a 'Login' button.

Then, you need to create a project to hold the dataset. You can choose whatever project name and description you like. For project type, you need to choose sequence labeling. Save the project

A screenshot of an 'Add Project' form. It includes fields for 'Project name' (containing 'ner'), 'Description' (containing 'ner'), and 'Project type' (set to 'Sequence Labeling'). There are also two checkboxes: 'Randomize document order' and 'Share annotations across all users'. At the bottom are 'Cancel' and 'Save' buttons.

Once you have created the project, you should be able to see the page below:



The screenshot shows the Doccano web interface. On the left, there is a sidebar with the following navigation options: Home, Dataset (which is selected), Labels, Relations, Members, Comments, Guideline, Statistics, and Settings. A blue button labeled "Start Annotation" is at the top of the sidebar. The main content area has a title "1 Import a dataset". Below it is a modal window titled "doccano: Dataset Import". The modal has a "Upload Data" section with a "Select a file format" dropdown containing "Plain text", "JSON", and "GML". There is also a "Select a file" input field with a red play button icon. Below the modal are two buttons: "CONTINUE" and "CANCEL". To the right of the modal, there is a numbered list of steps: 2 Create labels for this project, 3 Add members for collaborative work, and 4 Define a guideline for the work.

Doccano has (not that) detailed video instruction on how to annotate data on this platform. It should cover most of the cases. You can give it a look if you are interested.

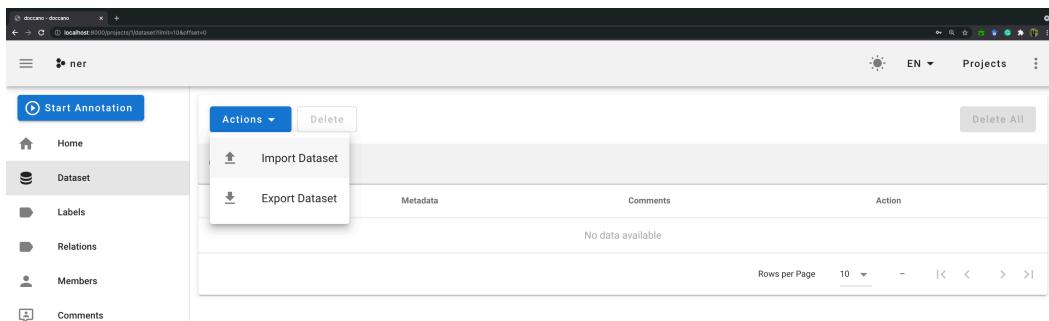
ANNOTATING DATA

The automatically generated ROP property sentences and polymer names are currently not perfect. To improve the performance of the models, we need larger training dataset. Doccano allows us to directly modify the predicted labels, which can be much faster than annotating a document from scratch.

The input of Doccano is in JSONL format, which are provided in the “jsonl_files” folder. The files have the same names apart from the suffix as the result files in the “html_files” folder, so you can search for them by the file name with your file manager. I do not know how to implement a “reveal in folder/finder” function in the Excel tables thus we have to stick to this less convenient way of finding files.

For the documents you think that are worth labeling, to correct the labels generated by the automated NLP system or introducing new annotations, you can follow the procedures below:

1) Import datafile: In Doccano webpage, click “Dataset”, “Actions”, “Import dataset”



The screenshot shows the Doccano web interface. The sidebar on the left is identical to the previous one. The main content area shows a table with columns: Actions, Delete, Import Dataset, Export Dataset, Metadata, Comments, and Action. A modal window is open over the table, showing the "Actions" dropdown menu with "Import Dataset" selected. At the bottom of the table, it says "No data available". There are buttons for "Rows per Page" (set to 10), and navigation arrows.

Choose “JSONL” data format and drop the candidate file into the box, click “inject”

The screenshot shows a web-based annotation tool interface. On the left, a sidebar menu includes 'Start Annotation', 'Home', 'Dataset' (which is selected), 'Labels', 'Relations', 'Members', 'Comments', 'Guideline', 'Statistics', and 'Settings'. The main area is titled 'Upload Data' with a dropdown menu set to 'JSONL'. It contains fields for 'Column Data' (text) and 'Column Label' (label), both currently empty. An 'Encoding' dropdown is set to 'utf_8'. Below these fields is a code block containing JSONL data:

```
{"text": "EU rejects German call to boycott British lamb.", "label": [0, 2, "ORG", ...]}  
{"text": "Peter Blackburn", "label": [10, 15, "PERSON"]}  
{"text": "President Obama", "label": [10, 15, "PERSON"]}
```

Below the code block is a grey area with the placeholder 'Drop files here...'. A green progress bar at the bottom indicates the file '10.1021/ma501541g.jsonl' has been uploaded. To the right of the progress bar is a button labeled 'Upload complete' with a link 'View in dataset'. At the bottom right of the interface is a blue 'Ingest' button.

Goto “Dataset” or use “Backward” button, you will be able to see the file you just imported

The screenshot shows the 'Dataset' interface. At the top, there are buttons for 'Actions' (with a dropdown arrow), 'Delete', and 'Delete All'. Below this is a search bar with the placeholder 'Search'. The main area displays a table with columns: 'Text' (checkbox), 'Metadata' (checkbox), 'Comments' (checkbox), and 'Action' (button labeled 'Annotate'). There is one row of data:

Text	Metadata	Comments	Action
<p>doi: 10.1021/ma501541g title: ATRP of POSS Monomers Revisited: Toward High-Molecular Weight Methacrylate-POSS (Co)Polymers Abstract: For the first time, ATRP was successfully employed for homopolymer...</p>	<p>{ "doi": "10.1021/ma501541g" }</p>	<p>0</p>	<p>Annotate</p>

At the bottom of the table are buttons for 'Rows per Page' (set to 10), '1-1 of 1', and navigation arrows.

2) Annotating data: click the “Annotate” button, you will see the annotation interface with some words/sentences already labeled by the BERT models

The screenshot shows the annotation interface. On the left, a sidebar menu includes 'Start Annotation', 'Home', 'Dataset' (selected), 'Labels', 'Relations', 'Members', 'Comments', 'Guideline', 'Statistics', and 'Settings'. The main area displays a text document with several words highlighted in blue boxes with the label 'POLYMER':

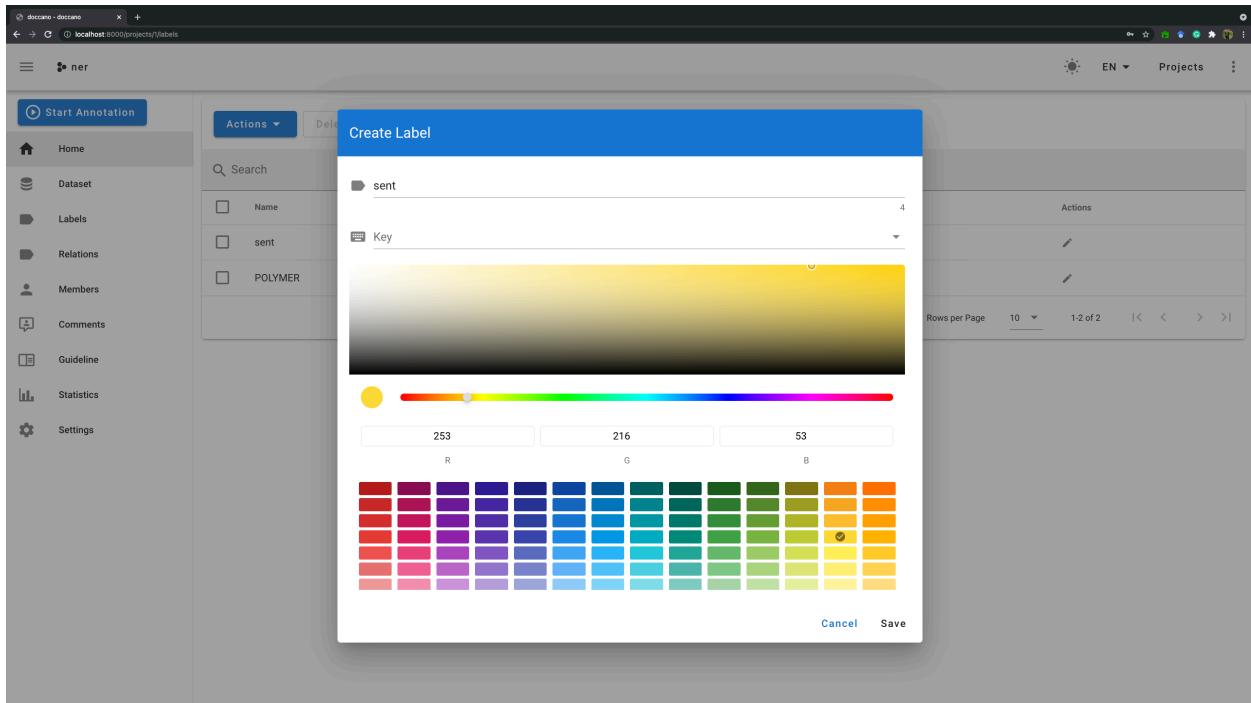
For the first time, ATRP was successfully employed for homopolymerization of a commercial methacrylate-functionalized polyhedral oligomeric silsesquioxane (POSS) monomer, iBuPOSSMA, to high molecular weights.

It was found that iBuPOSSMA has a low ceiling temperature (T_c); therefore, low temperatures and/or high initial monomer concentrations need to be used in order to avoid low degrees of polymerization that had been observed previously.

The values of T_c , as well as of the polymerization enthalpy ΔH_p and entropy ΔS_p were determined to be 130 °C (at $[M]_0 = 1 \text{ M}$), -41 kJ mol $^{-1}$, and -101 J mol $^{-1}$ K $^{-1}$, respectively.

Under optimized conditions, poly(iBuPOSSMA) homopolymers having low dispersity and high M_n , ranging from 23 000 to 460 000, were obtained in a well-controlled ATRP process. Moreover, various block copolymers having high- M_n blocks were prepared by copolymerization of iBuPOSSMA with methyl methacrylate and styrene.

The above figure shows 2 entity types: “sent” and “POLYMER”. “sent” entity represents the sentences that are predicted to contain ROP property values. “POLYMER” entity represents the polymer names. You can change their colors in the “Labels” panel to make it easier to separate them apart (“Labels” -> “Action”)



Result:

It was found that

iBuPOSSMA
POLYMER

has a low ceiling temperature (T_c); therefore, low temperatures and/or high initial monomer concentrations

need to be used in order to avoid low degrees of polymerization that had been observed previously.

The values of T_c , as well as of the polymerization enthalpy ΔH_p and entropy ΔS_p were determined to be 130 °C (at $[M]_0 = 1 \text{ M}$), -41 kJ mol⁻¹, and -101 J mol⁻¹ K⁻¹, respectively.
sent

Under optimized conditions, poly(iBuPOSSMA)
POLYMER

homopolymers having low dispersity and high M_n , ranging from 23 000 to 460 000, were obtained in a well-

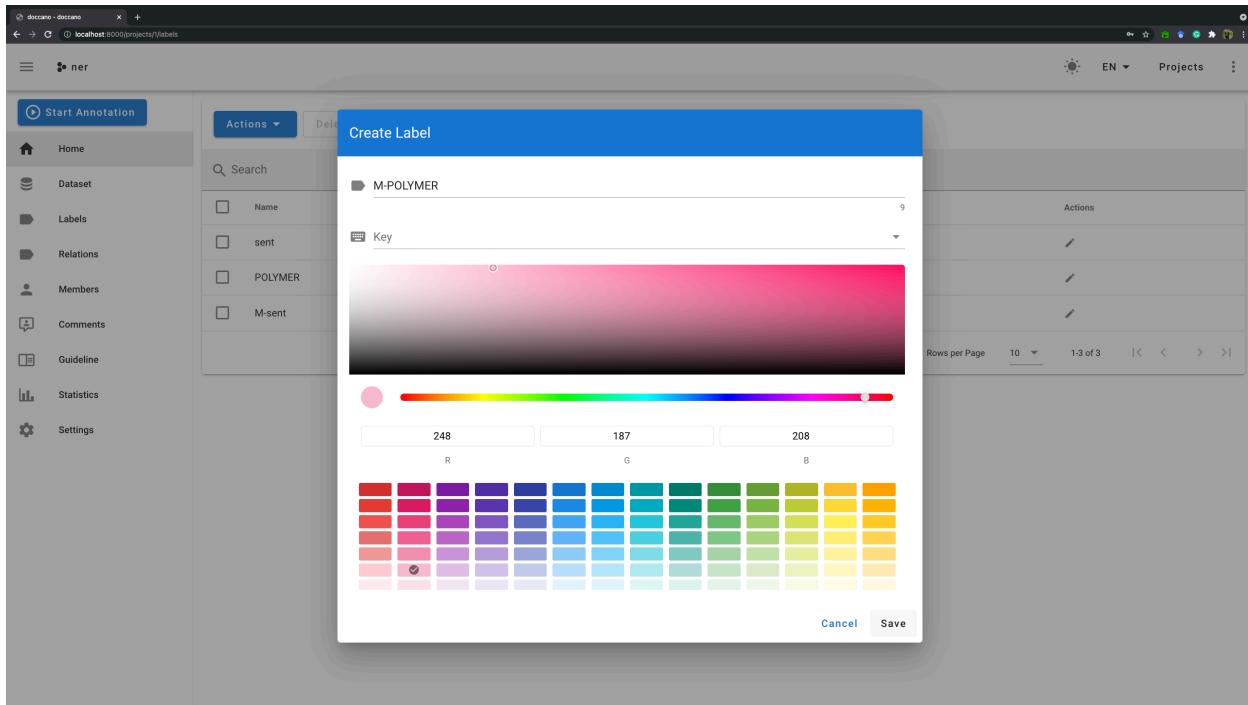
controlled ATRP process. Moreover, various block copolymers having high- M_n

poly(iBuPOSSMA)
POLYMER

blocks were prepared by copolymerization of iBuPOSSMA with methyl methacrylate and styrene.

NOTICE: You should check the results in the Excel table and HTML result files **BEFORE** going through this labeling step. This system is not suitable for direct result presentation.

To insert your labels or confirm model output, you should create entity labels starting with “M-”, in this case, “M-sent” and “M-POLYMER” (the labels are case-sensitive). Notice that entity labels not starting with “M-” will be discarded during my future post processing. You can also assign a hotkey to each entity label:



To confirm an annotation, click that entity and, in the drop-down list, select the corresponding entity label started with “M-”:

The values of T_c, as well as of the polymerization enthalpy ΔH_p and entropy ΔS_p were determined to be 130 °C (at [M]₀ = 1 M), -41 kJ mol⁻¹, and -101 J mol⁻¹ K⁻¹, respectively.

sent

sent

POLYMER

M-sent

M-POLYMER

controlled ATRP process. Moreover, various block copolymers having high-M_n

poly(iBuPOSSMA)
POLYMER

The entity label is successfully changed:

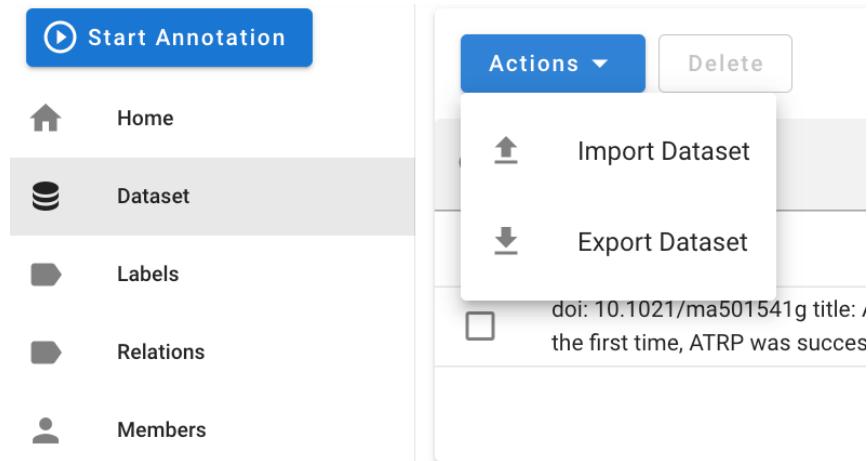
The values of T_c, as well as of the polymerization enthalpy ΔH_p and entropy ΔS_p were determined to be 130 °C (at [M]₀ = 1 M), -41 kJ mol⁻¹, and -101 J mol⁻¹ K⁻¹, respectively.

M-sent

Under optimized conditions, **poly(iBuPOSSMA)**
POLYMER

To insert an annotation, simply select the text that belongs to the entity and choose the corresponding entity label started with “M-” in the drop-down list. To delete an annotation, hover your cursor above the entity and click the “X” button at the top-left corner.

3) Saving results: You can export your dataset with the “Export dataset” button in the “Action” list in “Dataset” section.



Notice that the platform only allows you to export the entire dataset in one JSONL file. It is suggested to label some document -> export results -> delete all document in the dataset -> import and label other documents. This procedure keeps your workspace clean.

Important: according to the GitHub page, you need to use ` docker container stop doccano -t 5` to stop your doccano container while keeping your work. Your progress may be lost otherwise.