



# @Note2 tutorial

Hugo Costa ([hcosta@silicolife.com](mailto:hcosta@silicolife.com))

Ruben Rodrigues ([pg25227@alunos.uminho.pt](mailto:pg25227@alunos.uminho.pt))

Miguel Rocha ([mrocha@di.uminho.pt](mailto:mrocha@di.uminho.pt))

The document presents a typical workflow using @Note2 platform capabilities for a selected case study “Stringent response” in bacteria “*Escherichia coli*”.

**Website:** <http://www.anote-project.org>

**General email:** [anote@silicolife.com](mailto:anote@silicolife.com)

**Repository - libraries:** <https://github.com/orgs/biotextmining/>

**Repository - @Note application and plugins:** <https://sourceforge.net/projects/anote/>

## Contents

Getting started .....	3
Workflow for a specific case study.....	4
Finding texts using the Publication Manager.....	4
Creating Clusters using Carrot2 <i>Plug-in</i> .....	8
Creating a Corpus .....	9
Using the Corpora <i>Plug-in</i> .....	9
Using Corpus Loader plug-in .....	11
Lexical Resources plug-in .....	11
Entity Recognition Plug-in. ....	19
Other Entity Recognition Plug-ins .....	21
Manual Curation and Editing annotations .....	22
Relation Extraction .....	27

# Getting started

To download and install the application please follow the steps in:

<http://anote-project.org/download>

and simply follow the recommended steps depending on your system. Take note of the details mentioned in:

[http://anote-project.org/wiki/index.php?title=Getting\\_Started#Installation\\_Notes](http://anote-project.org/wiki/index.php?title=Getting_Started#Installation_Notes)

To first configure the application, @Note runs a wizard the first time you start it. Check the details in the following wiki item:

[http://anote-project.org/wiki/index.php?title=Starting\\_to\\_use\\_@Note](http://anote-project.org/wiki/index.php?title=Starting_to_use_@Note)

Before starting to see what @Note can offer, it is important to notice that you are able to save your session or to load a previously saved session from file. Check the following How To's for details:

[http://anote-project.org/wiki/index.php?title=Save\\_Session](http://anote-project.org/wiki/index.php?title=Save_Session)

[http://anote-project.org/wiki/index.php?title=Load\\_Session](http://anote-project.org/wiki/index.php?title=Load_Session)

# Workflow for a specific case study

Currently, @Note2 has many plug-ins. It is possible to work with only a few or with all of them. In this section, we present a case study for “Stringent Response in Escherichia Coli” that shows how the user can take advantage of some of @Note2 capabilities. Naturally, this document does not cover the full set of potential operations supported by @Note2, which has more far options. To know all existing functionalities, please check the @Note2 Wiki <http://anote-project.org/wiki/>.

## Finding texts using the Publication Manager

The Publication Manager (PM) plug-in provides a system for recovering PubMed information with a combination of keywords and other specific fields. The result is a set of publications containing information such as: title, abstract, journal. The publications are organized in Queries.

To perform PubMed searches, right click on an object of the Publication Manager data-type in the clipboard or press the “New Query” button on the Publication Manager View (bottom left). A new search will be performed with the keywords selected in the pop-up window.

The screenshot shows a 'Pubmed Search' dialog box with the following fields and options:

- Query Name:** [KEYWORDS]:[ORGANISM]:[Date]
- Query Main Information:**
  - Keywords : Stringent Response
  - Organism : Escherichia Coli
- Query Properties:**
  - Authors: [Empty field]
  - Journal : [Empty field] Type : All
  - Publication Date : 2010 To 2012
- Repository Information:**
  - All Documents
  - Pubmed Central
  - MedLine
- Article Details:**
  - All Documents
  - Available Abstract
  - Open Access Articles
  - Available Full Text

Buttons at the bottom: Cancel, Help, Ok.

For this case study, we looked for “stringent response” related publications (keywords) for “Escherichia coli” (organism), published between the years 2010 and 2012 and with available free full texts, as shown in the figure above. Pressing the “ok” button launches a GUI showing the number of publications that match the query (at the time of writing of this document there were 36). Here, you can cancel the query or continue to recover the full information. Check details about this operation in the PubMed search wiki item:

[http://anote-project.org/wiki/index.php/Pubmed\\_Search](http://anote-project.org/wiki/index.php/Pubmed_Search)

The result of this operation is a new query that appears on the Publication Manager View. The details about this view can be checked in the following link:

[http://anote-project.org/wiki/index.php/Publication\\_Manager\\_View](http://anote-project.org/wiki/index.php/Publication_Manager_View)

In this view, you can load queries to the clipboard pressing the rightmost button (plus symbol) in the Queries table for a specific query. A new Query object is added to the clipboard, hierarchically under the Publication Manager (named “Query: 1”). Clicking on Query 1, you can visualize the respective data. There are three tabs in this view: one for viewing the publication details, another for assigning document relevance and the last to view clustering results (this option will be presented later on this document).

Within the “Query View” (figure below), you can check the list of publications and their details and can also select publications to perform the Journal Retrieval option. The full set of functionalities within this view are given in:

[http://anote-project.org/wiki/index.php/Query\\_View](http://anote-project.org/wiki/index.php/Query_View)

The screenshot shows the @Note2.2.5.0 application window. The title bar reads "@Note2.2.5.0". The menu bar includes "File", "Corpora", "Corpus", "Process", "Publication Manager", "Resources", "Workflow", "Settings", and "Help". The interface is divided into several sections:

- Clipboard:** A sidebar on the left containing a tree view with "note2" as the root, followed by "Publication Manager", "Stringent Response:Escherichia Coli", "Corpora", and "Resources".
- Publication Manager:** The main window title, with a sub-tab "Stringent Response: E. coli".
- Query Information:** A form with fields for:
  - Keywords: Stringent Response
  - Organism: Escherichia Coli
  - Query Origin: PUBMED
  - Date: 2017-06-29 12:56:07
  - From Date: 2010
  - To Date: 2012
  - Complete Query: (free full text[fb]) AND ("2010[Publication date]"2012[Publication Date]) AND (Stringent Response) AND ((E. Coli) OR "Escherichia Coli")
- Property Table:** A small table with columns "Property" and "Option":
 

articleDetails	freshfulltext
toDate	2012
fromDate	2010
- Publications:** A table with 36 rows. The total count is "Total 36". The table has columns: Details, Title, Authors, Date, PMID/OtherID, Link, PDF, and a checkbox column. The first few rows are:
 

Details	Title	Authors	Date	PMID/OtherID	Link	PDF	
<input type="checkbox"/>	Genome-wide analysis of aminoacylation (charging) levels of tRNA using microar...	Zaborske J, Pan T	2010	ID: 578308305492...			<input checked="" type="checkbox"/>
<input type="checkbox"/>	Antagonistic regulation of dgkA and plsB genes of phospholipid synthesis by mul...	Wahl A, My L, Dumoulin R, St...	2011	ID: 106057227141...			<input checked="" type="checkbox"/>
<input type="checkbox"/>	Transcription regulation of the Escherichia coli pcnB gene coding for poly(A) poly...	Nadratowska-Wesolowska B...	2010	ID: 115952622247...			<input checked="" type="checkbox"/>
- Operations:** A row of buttons: Refresh, Add Publication, Update Query, Workflow Information Extraction, and Document Retrieval.
- Bottom Bar:** Three tabs: "Query View" (selected), "Query Relevance View", and "Query Clusters View".

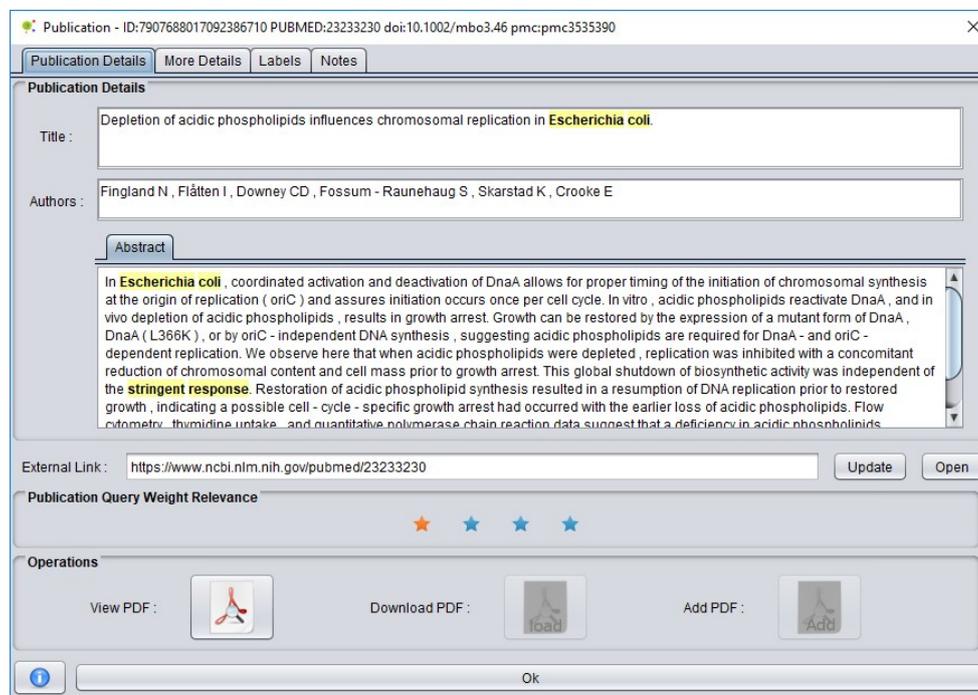
The document retrieval operation is performed by marking the set of desired publications and, then, pressing the “Journal Retrieval” button. @Note tries to download the available PDF files

for you. When this operation is over, the successfully downloaded publications change the icon status in the publications table. Publications that contain PDF files appear with the following icon . Details on this operation are given in:

[http://anote-project.org/wiki/index.php/Publication\\_Retrieval](http://anote-project.org/wiki/index.php/Publication_Retrieval)

In this case study, we selected three publications to obtain the full texts in PDF (PMID: 23233230, 22880033, 22393021). To do this, select the three publications and press the “Journal Retrieval” Button. All PDFs should be downloaded. Note that all these papers are freely available.

In the publications table within the query view, it is also possible to view details about each publication. When you select the leftmost button in a publication, details as the abstract, authors, journal and others are shown. In this GUI, it is also possible to try to download the publication file if it does not exist yet or to view the PDF file if it is already available. The details for the publication with PMID: 23233230 are shown in the figure below.



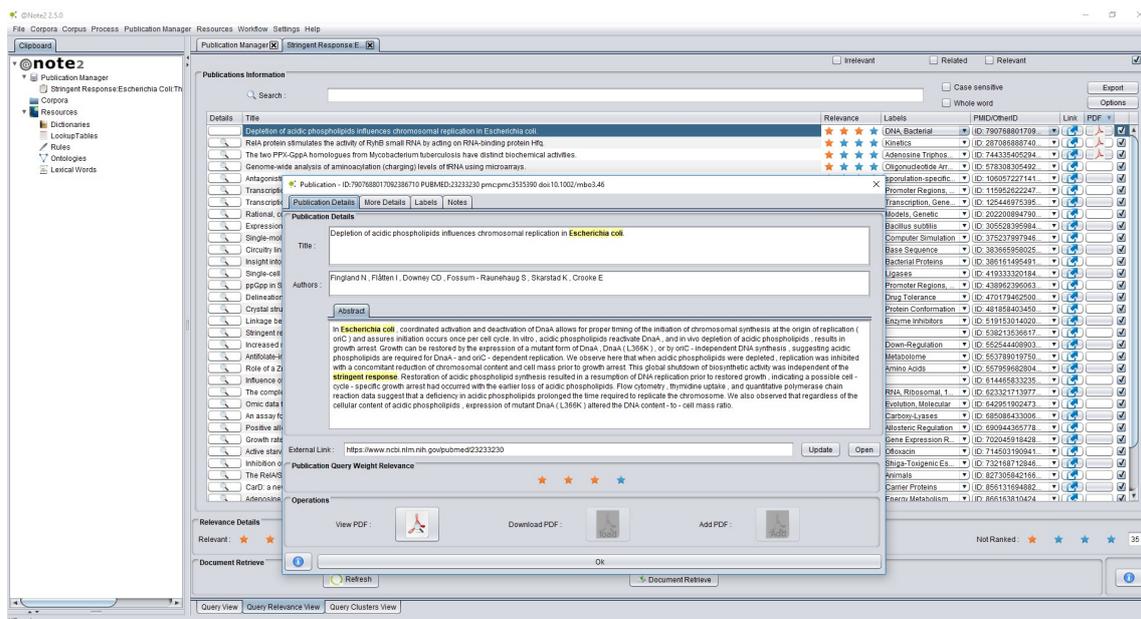
Pressing the “ViewPDF” button it is possible to view the PDF content of the paper, as shown below.



In the Query Relevance View, it is possible to check the relevance that the user gives to each of the publications in the Query. You can change the publication relevance within the Publication details view described above, clicking on the “weight relevance” button selecting the appropriate level. When the publication relevance is changed, the number of stars will be updated (relevance degrees are coded by number of colored stars). Full details for this view can be checked in:

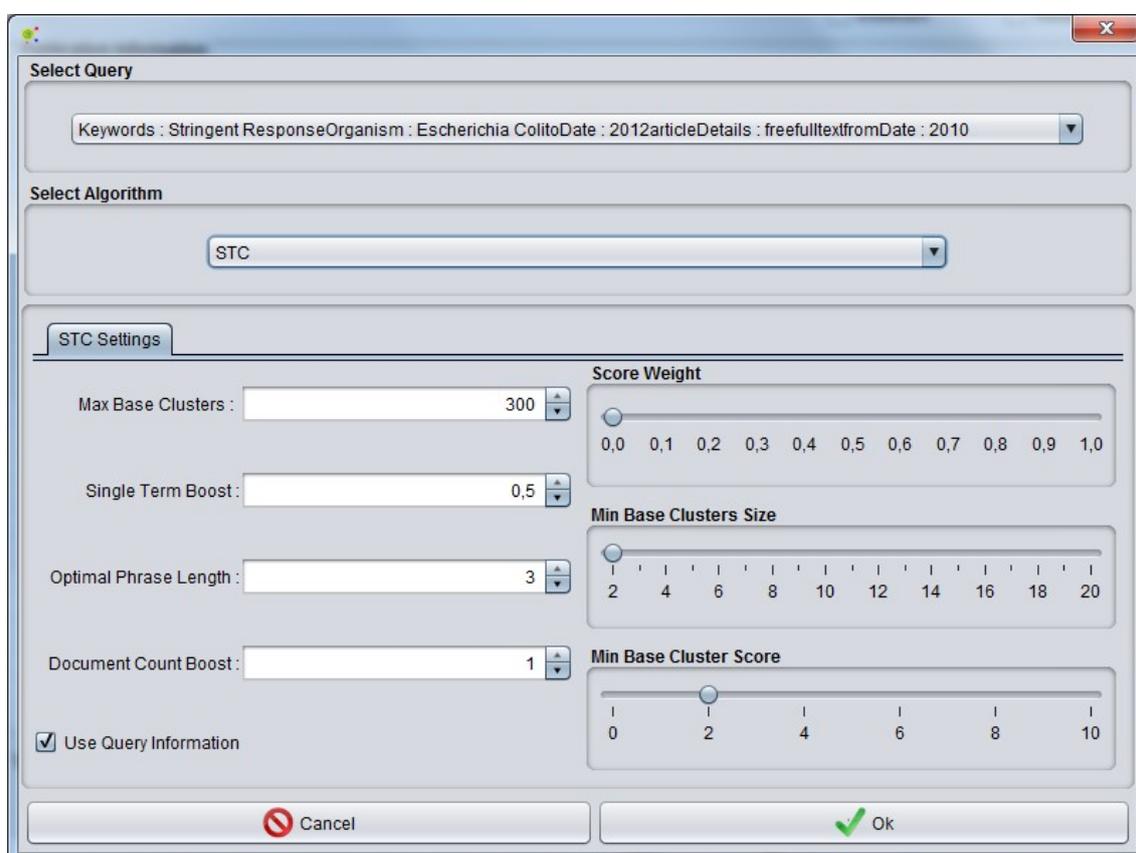
[http://anote-project.org/wiki/index.php/Query\\_Relevance\\_View](http://anote-project.org/wiki/index.php/Query_Relevance_View)

As an example, try to select publication with PMID 23233230 and assign two stars (paper considered to be “Related”) in the Weight Relevance option. After closing the publication details, the number of stars for the row of this publication changes, as shown in the figure below:



## Creating Clusters using Carrot2 Plug-in

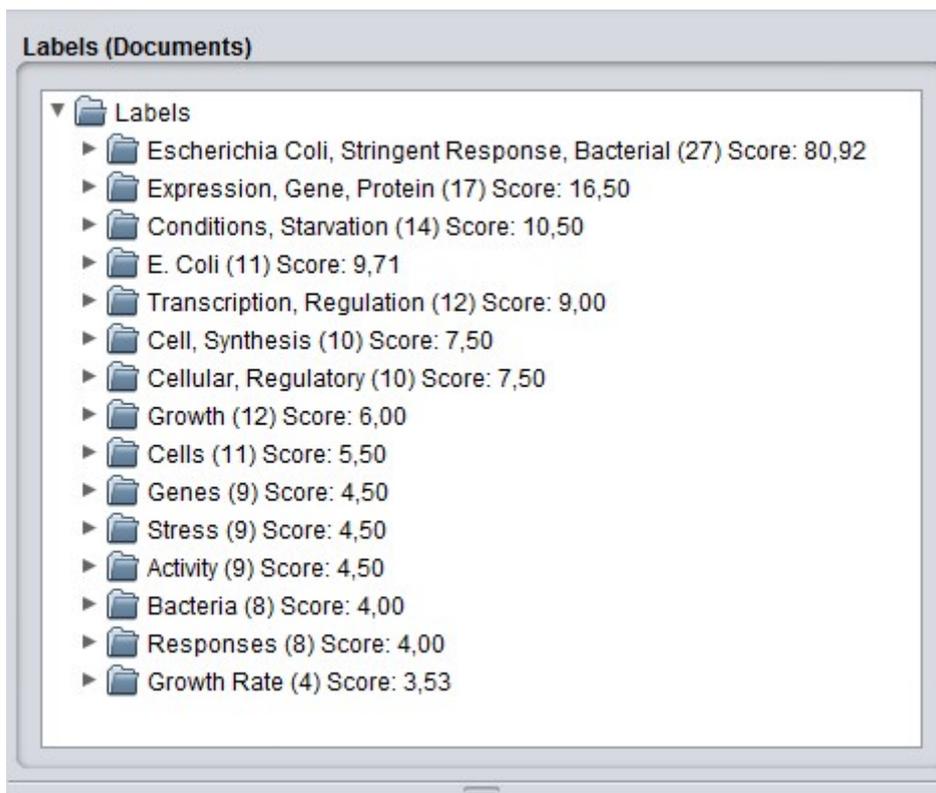
This plugin allows running @Note's clustering algorithms that make use of the [Carrot2 software](#). This allows clustering documents within Publication Manager Queries and with the results generate new Queries. To create clusters based on the results of a Query, select the "Cluster" option in the Query data-type. A Clustering Configuration GUI is launched where you can select the Algorithm and define its parameters. As a result, a new row is shown within the Cluster View. An example is shown below using the STC algorithm with default parameters.



Details can be found in:

[http://anote-project.org/wiki/index.php/Query\\_Create\\_Cluster](http://anote-project.org/wiki/index.php/Query_Create_Cluster)

An example of the results is shown in the following figure:



You can view the full results by clicking in the appropriate row on the table (Cluster View) and viewing information on the different clusters and enclosed documents. More details in:

[http://anote-project.org/wiki/index.php/Query\\_Cluster\\_View](http://anote-project.org/wiki/index.php/Query_Cluster_View)

Also, you can create a new Query using the results of clustering (using one or more clusters).

[http://anote-project.org/wiki/index.php/Create\\_Query\\_from\\_Cluster\\_results](http://anote-project.org/wiki/index.php/Create_Query_from_Cluster_results)

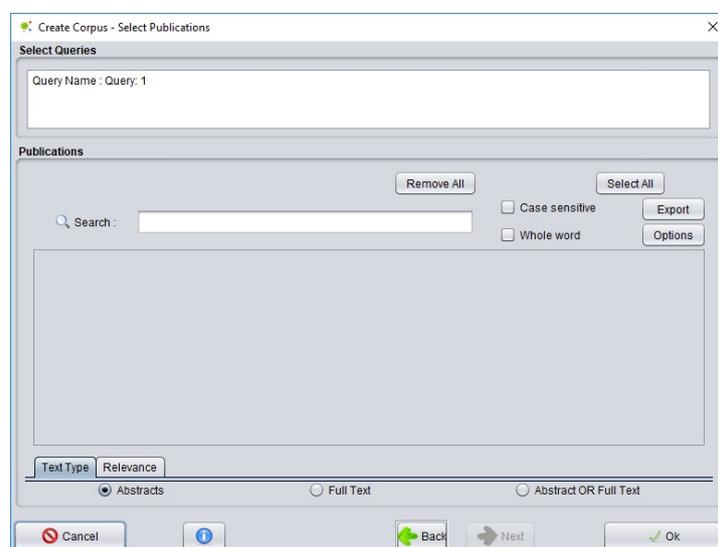
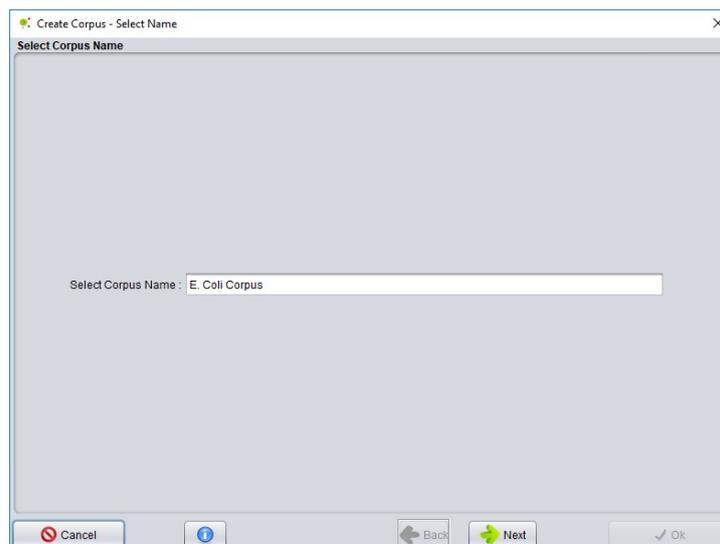
## Creating a Corpus

### Using the Corpora Plug-in

The core @Note2 plug-in is the Corpora Plug-in that define Corpora (sets of Corpus) and makes possible the integration of extraction processes (like entity extraction or relation extraction). A Corpus, i.e. a set of documents, can have processes indexed, i.e. Corpus Process objects, that have annotations for a document (entity/event annotations).

Basically, there are two ways to create a corpus: using queries in the Publication Manager or loading an already existing corpus from literature using specific loaders. The second case is presented in the next sub section. Here, we show how the user can create a corpus based in queries built in the Publication Manager plug-in (see previous section).

To create a Corpus based on Publication Manager queries, you have to click the Corpora data-type icon in the clipboard and select the option “Corpus -> Create -> From Publication Manager”. Selecting this option causes the Corpus Creation Wizard to be launched allowing the selection of the desired queries, the filtering of publications by relevance level or content options (abstract or full texts). For example, in this case study, we will select all publications that have an abstract available, using the query created in the last section. In the corpus name you can put, for instance, “E. coli Corpus”, in the selection of the queries, select “Query 1” and keep all default options for the selection of publications.



After this operation, a new Corpus is created and will be available in the Corpora View with the name “E. coli Corpus”. Full details on this operation can be checked in the following wiki link:

[http://anote-project.org/wiki/index.php/Create\\_Corpus\\_By\\_Publication\\_Manager](http://anote-project.org/wiki/index.php/Create_Corpus_By_Publication_Manager)

## Using Corpus Loader plug-in

This sub section presents another form to create a corpus by loading existing ones. In this case, five corpus loaders are available: GENIA Event Corpus, GENETAG-05 from MedTag, ABGene from MedTag, CEMP from CHEMDNER task, BioNLP corpus from A1 A2 format, AIMED Protein and @Note v1.

To create a corpus using this option, you should right click on the Corpora data-type in the clipboard and select the specific loader for the required format, also selecting the directory in which the respective files are kept.

For example, for the case of the AIMED Protein, the files can be obtained in: <ftp://ftp.cs.utexas.edu/pub/mooney/bio-data/proteins.tar.gz>. @Note2 creates a new corpus based on the AIMED Protein information. Indexed to this corpus, information extraction (NER and RE) processes are also imported, containing information about Entities and Events present in the original corpora. Later in this document we describe NER and RE processes.

More details can be found in:

[http://anote-project.org/wiki/index.php/Create\\_Corpus\\_By\\_AIMED\\_Corpus](http://anote-project.org/wiki/index.php/Create_Corpus_By_AIMED_Corpus)

Other loaders for available formats are made available for @Note version 1 files and for Genia Event Corpus. Check the details in:

[http://anote-project.org/wiki/index.php/Create\\_Corpus\\_By\\_AnoteV1\\_Corpus](http://anote-project.org/wiki/index.php/Create_Corpus_By_AnoteV1_Corpus)

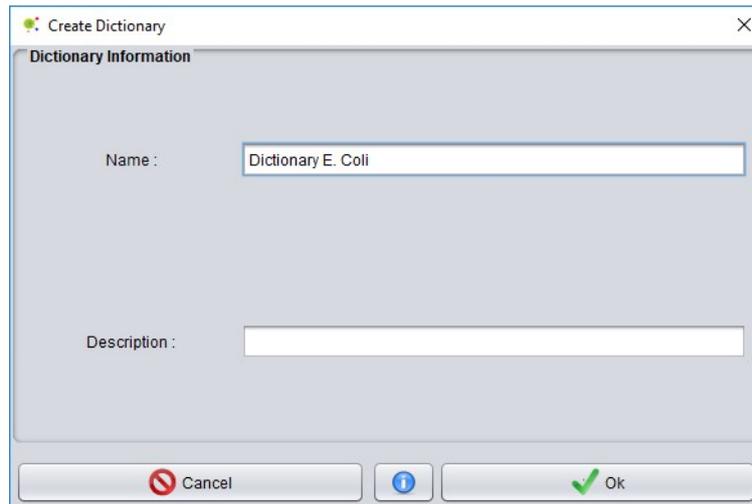
[http://anote-project.org/wiki/index.php/Create\\_Corpus\\_By\\_Genia\\_Event\\_Corpus](http://anote-project.org/wiki/index.php/Create_Corpus_By_Genia_Event_Corpus)

## Lexical Resources plug-in

At this stage, you have just created a corpus (using one of the options). You can now apply Information Extraction processes to this corpus. One of them is defined in the next section, the @Note Entity Recognition algorithm that implements Named Entity Recognition (NER) based on lexical resources.

To start, we can provide an overview of the different types of resources available. In this case study, to illustrate the options, we will create a dictionary based in flat files, a basic lookup table, build some syntactic rules, an ontology and create a set of words (Lexical Words object).

To create a Dictionary, you should right click in the dictionaries data-type in the clipboard and select "Add Dictionary" in "Resources -> Dictionaries" and in the GUI select name and notes for dictionary. In this example, "E. coli Dictionary" will be created and the user can load it to clipboard by pressing rightmost button on Dictionaries table View.



Check further details on dictionary creation in:

[http://anote-project.org/wiki/index.php/Dictionary\\_Create](http://anote-project.org/wiki/index.php/Dictionary_Create)

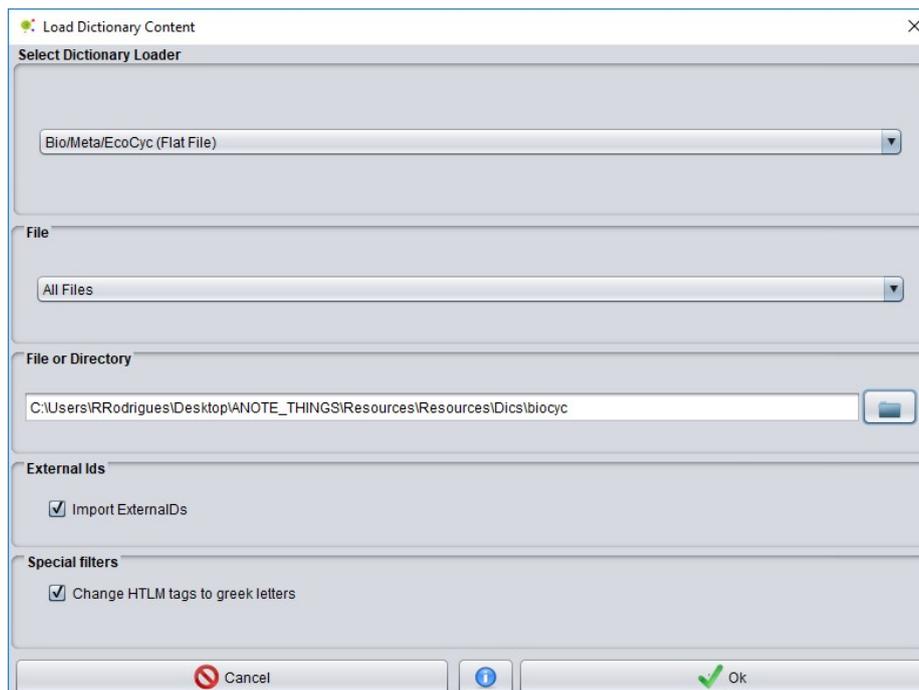
To update the dictionary content, press the right button on the Dictionary data-type and select the "Update -> Native Loaders" option. The Dictionary Update GUI appears and you can select: the source (database), the entity type, the file location and the Organism (some databases allow this selection to be made). More details in:

[http://anote-project.org/wiki/index.php/Dictionary\\_Update](http://anote-project.org/wiki/index.php/Dictionary_Update)

For this case study we downloaded *E. coli* flat files from Biocyc at the following link:

[http://darwin.di.uminho.pt/anote2/material/resources/dictionaries/BioCyc/biocyc\\_ecoli.tar](http://darwin.di.uminho.pt/anote2/material/resources/dictionaries/BioCyc/biocyc_ecoli.tar)

Afterwards, we unzipped the file. In the GUI, select the folder and options "BioCyc" in the source and "all" in the entity.



When the import process is over, the dictionary is updated with the new data. The result will look like this:

Dictionary Information and Statistics				
Terms	8650	Class	Terms	Synonyms
Synonyms	9683	protein	1656	2832
Average Term/Syn	1,12	compound	917	2725
Number of classes :	5	enzyme	1739	3938
		gene	4058	1
		pathways	280	187

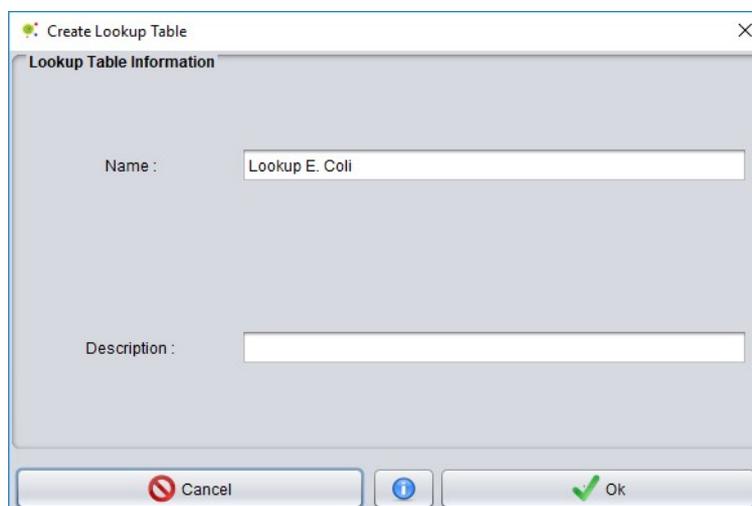
You can then add or edit dictionary information using the Dictionary View. More details on how to edit dictionary contents can be found in:

[http://anote-project.org/wiki/index.php/Dictionary\\_Management\\_Content](http://anote-project.org/wiki/index.php/Dictionary_Management_Content)

Also, information can be imported from CSV files, as detailed in:

[http://anote-project.org/wiki/index.php/Dictionary\\_Import\\_Element\\_From\\_CSV\\_File](http://anote-project.org/wiki/index.php/Dictionary_Import_Element_From_CSV_File)

Proceeding with the creation of lexical resources, we will now show how to create a Lookup table. This process starts by right clicking in the Lookup table's data-type in the clipboard and selecting the name on the GUI. In this example, we will call the table: "Lookup E. coli".



Details can be found in:

[http://anote-project.org/wiki/index.php/LookupTable\\_Create](http://anote-project.org/wiki/index.php/LookupTable_Create)

Now, it is possible to insert a term manually by right clicking on the lookup table data-type and selecting the "Add Element" option. In the GUI, the user puts a name and a class for the

element to add presses “ok”. For this case study, we will introduce 3 terms: PCR – Laboratory technique; stringent response – cell state; Water – Compound, as shown below:

Lookup Table : Lookup Table : Lookup E. coli (ID :2)

Term : PCR

Class : New Class

NewClass : Laboratory Technique

Cancel Ok

Lookup Table : Lookup Table : Lookup E. coli (ID :2)

Term : stringent response

Class : New Class

NewClass : cell state

Cancel Ok

Lookup Table : Lookup Table : Lookup E. coli (ID :2)

Term : Water

Class : New Class

NewClass :

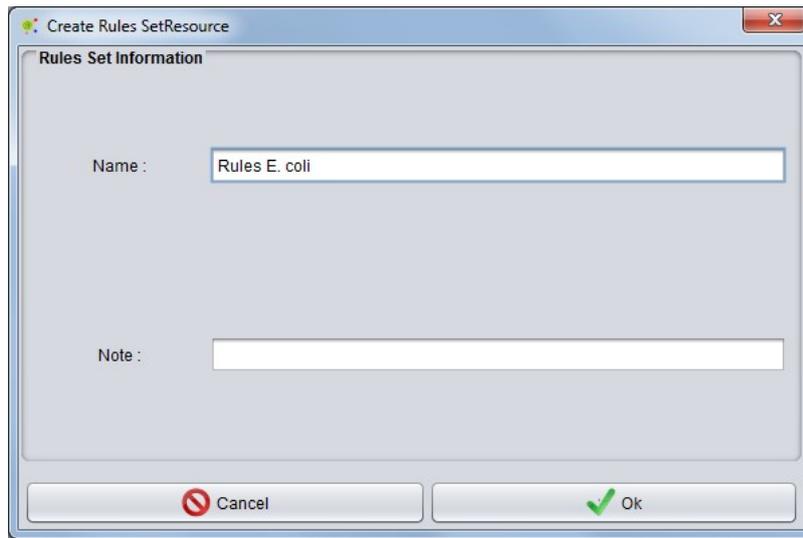
- gene
- cell state
- compound
- laboratory technique
- reactions
- protein
- pathways
- New Class

Cancel Ok

More details in:

[http://anote-project.org/wiki/index.php/LookupTable\\_Add\\_Element](http://anote-project.org/wiki/index.php/LookupTable_Add_Element)

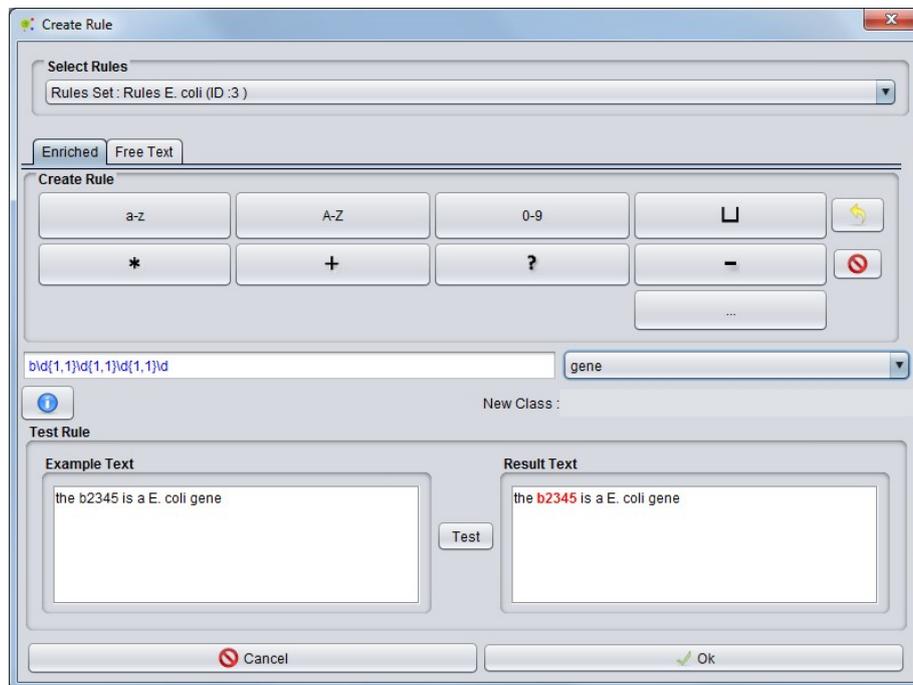
Proceeding to the creation of a Rules Set, right click in the Rules Set data-type in the clipboard and chooses a name. Here, it will be simply “Rules E. coli”.



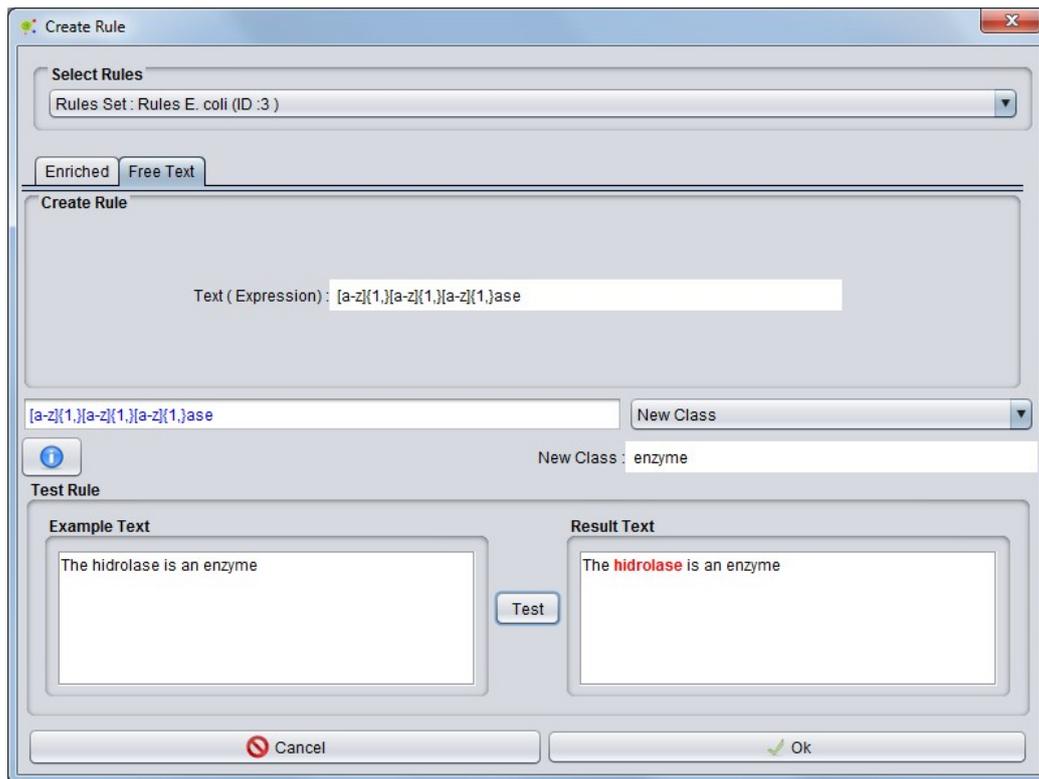
Further details can be found in:

[http://anote-project.org/wiki/index.php/RulesSet\\_Create](http://anote-project.org/wiki/index.php/RulesSet_Create)

Creating a rule set, you are able to create syntactic rules based on regular expressions. To create a rule, right click on the Rule Set data-type and select the "Add rule" option, launching a specific GUI. Here you have two options for creating rules: using an enriched visual environment or free text. As an example for the enriched environment, and based in the rule that an *E. coli* gene name starts with letter "b" plus 4 digits we create a syntactic rule for this pattern. The GUI allows to test the regular expression with examples defined by the user.



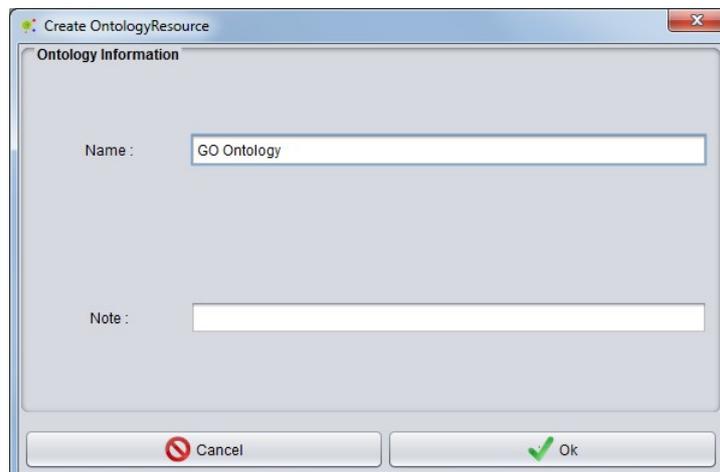
Another example, now built using free text, is to put a Rule for detecting enzymes. One word could be an enzyme if it has at least 6 letters and ends with "ase".



Find more details in:

[http://anote-project.org/wiki/index.php/RulesSet\\_Add\\_Rule](http://anote-project.org/wiki/index.php/RulesSet_Add_Rule)

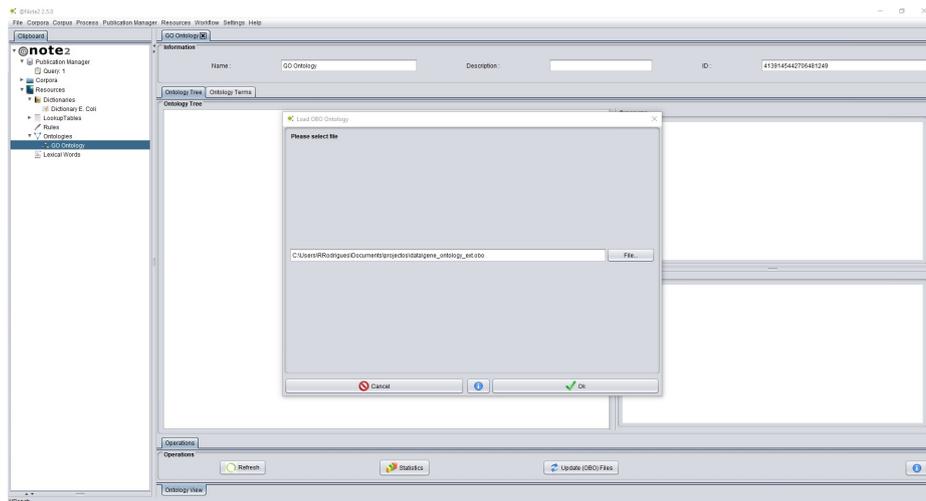
Other lexical resources available are biomedical ontologies. Ontologies are important for the disambiguation of terms or for mapping entities into ontology terms (unified terminology) and can be used as a terminology for entity recognition. So, for Ontology creation, you need to right click in the Ontologies icon or click in “add ontology” in the ontologies view. A GUI appears to add information on the name and some notes for the ontology. After pressing “ok” a new Ontology instance is created in the Ontologies View. In this example, we create an Ontology based in the Gene Ontology database, named “GO Ontology”.



Further details are available in:

[http://anote-project.org/wiki/index.php/Ontology\\_Create](http://anote-project.org/wiki/index.php/Ontology_Create)

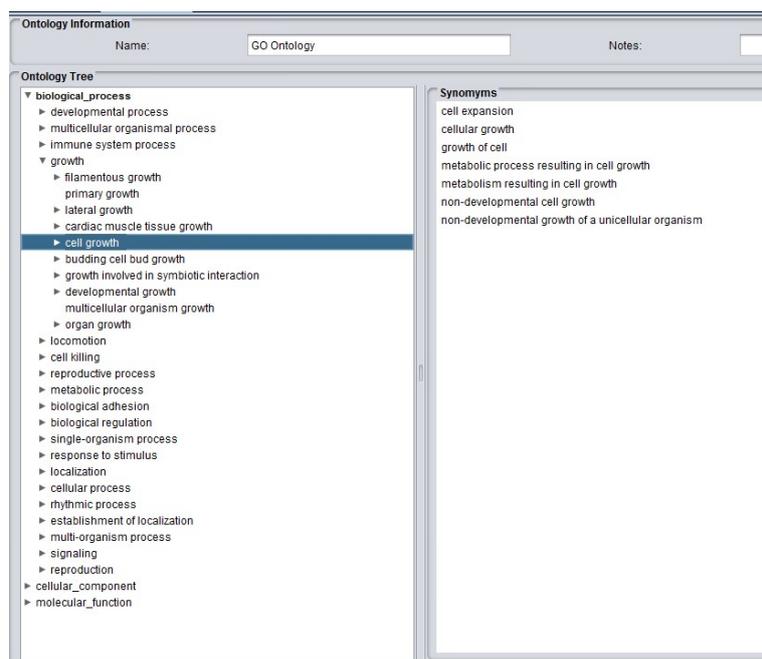
Now, you can select it and load it to the clipboard. You can fill ontologies selecting the update option (right clicking the Ontology Data type in the clipboard). Currently, it is possible to load Ontologies using the OBO format. In the GUI for “Ontology Update” you must select the OBO file and press “ok”. When the operation finishes, you can select it and view details in the Ontology view. For this case study, please download the following file: [http://www.geneontology.org/ontology/obo format 1 2/gene\\_ontology\\_ext.obo](http://www.geneontology.org/ontology/obo%20format%201%20/gene_ontology_ext.obo)



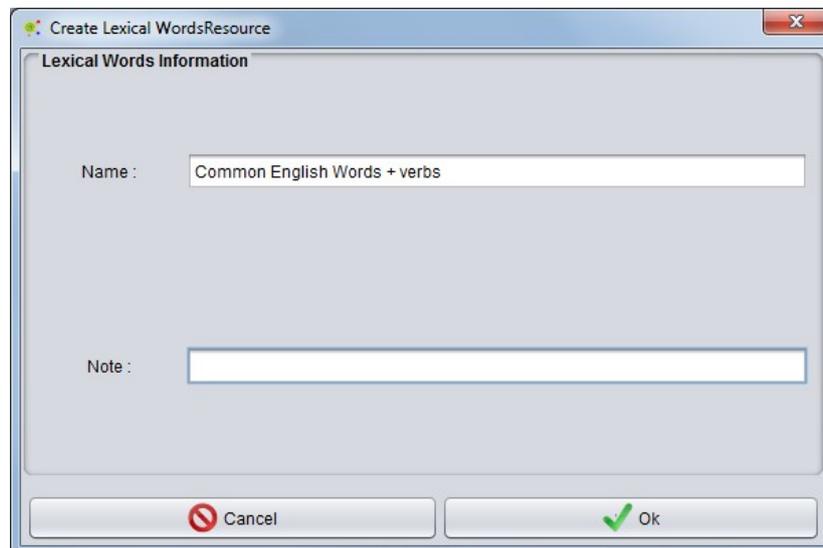
Check further details in:

[http://anote-project.org/wiki/index.php/Ontology\\_Update](http://anote-project.org/wiki/index.php/Ontology_Update)

You can view Ontologies in the Ontology View, where the terms are represented hierarchically, as shown below.



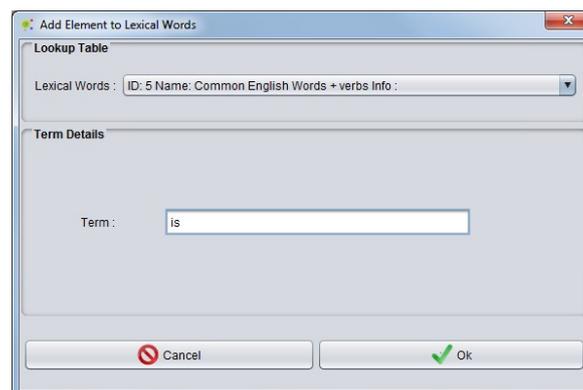
The last lexical resource presented will be the Lexical Words data-type. Each object of this type contains a list of words that can be used for several purposes like grouping important biomedical words, or representing common English words. To create an instance of Lexical Words, start by right clicking in the Lexical Words' data-type in the clipboard and enter the name on the GUI, in this case, for instance, "Common English Words + verbs".



You can find further details on:

[http://anote-project.org/wiki/index.php/LexicalWords\\_Create](http://anote-project.org/wiki/index.php/LexicalWords_Create)

Now, it is possible to insert a word manually by right clicking on the lexical works data-type and selecting the "Add Element" option. In the GUI, put a name add press "ok". For this case study, we can introduce 3 words: *is*, *was*, *to*. Check the GUI for the insertion of the first:



Further details in:

[http://anote-project.org/wiki/index.php/LexicalWords\\_Add\\_Element](http://anote-project.org/wiki/index.php/LexicalWords_Add_Element)

## Entity Recognition Plug-in.

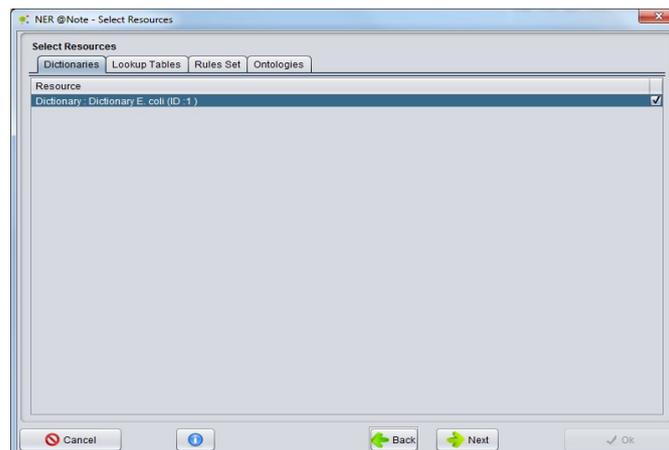
The Entity recognition plug-in performs Named Entity Recognition (NER) based on the lexical resources presented above. The result is a NER Process object that contains entity annotations for the selected documents (abstracts or full texts) present in the selected Corpus.

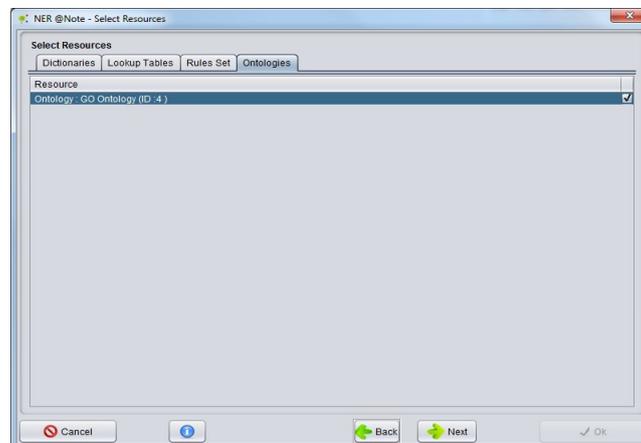
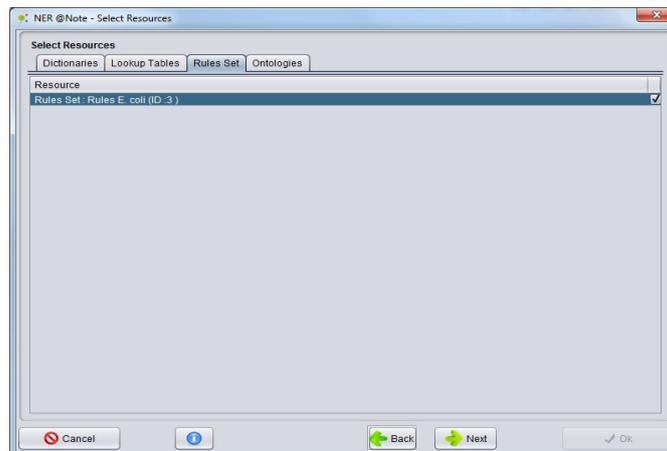
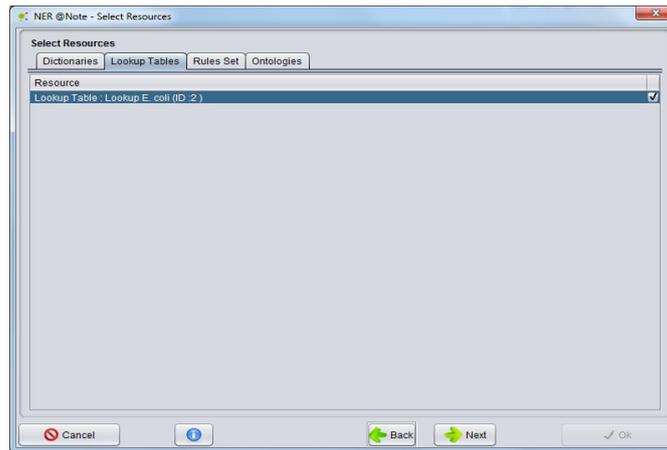
Thus, to apply a NER Process, we need to first create a Corpus. Using the one created above, you can select the Corpus data-type in the clipboard and right click to select “NER->Lexical Resources”.

A wizard for the NER process configuration is launched. The first step is to select if a new NER Process is created or if a previously created process that was applied before to another Corpus will be used. The latter option implies using a configuration taken from a previous NER process and will not be covered in this tutorial; the details of this option are given in the following How To:

[http://anote-project.org/wiki/index.php/Corpus\\_Create\\_Annotation\\_Schema\\_By\\_NER\\_Already\\_Configured](http://anote-project.org/wiki/index.php/Corpus_Create_Annotation_Schema_By_NER_Already_Configured)

In this example we will use the “E. Coli Corpus” created in a previous section, and thus will follow the first option (create a new NER). So, you should press “New NER” and proceed. The next step provides a GUI for Lexical resources selection. Here, you can choose one Dictionary and/or a Lookup table and/or a Rules Set and/or Ontology. You should select the appropriate resources and press “next”. For this example, we choose “Dictionary E. coli”, “Lookup E. Coli”, “Rules E. Coli” and “GO Ontology” created in previous section. See the details in the screenshots below:





For each lexical resource, you can choose the class or classes' terms to perform a NER. The full details of the configuration step are given in the following How To:

[http://anote-project.org/wiki/index.php/Corpus\\_Create\\_Annotation\\_Schema\\_By\\_NER\\_Lexical\\_Resources](http://anote-project.org/wiki/index.php/Corpus_Create_Annotation_Schema_By_NER_Lexical_Resources)

For this example we use all classes available for all the resources selected. Other options are available for NER. In this case study, we use the default option for POS-Processing and

Normalization, but other options can be selected being their behavior explained in detail in the Wiki documentation.

Once the configuration is complete, the NER operation will be executed. It will take a few minutes or hours depending on the number of documents and on the used lexical resources. When all operations are successfully performed, the new NER Process is added to the Corpora View.

Selecting a process in this view it is possible to load it to the clipboard. Check details in:

[http://anote-project.org/wiki/index.php/Corpus\\_Load\\_Process](http://anote-project.org/wiki/index.php/Corpus_Load_Process)

This new object offers two views: one for the entity details and another to check the annotated documents. Check details in:

[http://anote-project.org/wiki/index.php/Process\\_Entity\\_Details\\_View](http://anote-project.org/wiki/index.php/Process_Entity_Details_View)

[http://anote-project.org/wiki/index.php/Process\\_Annotated\\_Documents\\_View](http://anote-project.org/wiki/index.php/Process_Annotated_Documents_View)

In this view, an annotated document can be loaded for further analysis:

[http://anote-project.org/wiki/index.php/Process\\_Load\\_Annotated\\_Document](http://anote-project.org/wiki/index.php/Process_Load_Annotated_Document)

and its details can be visualized:

[http://anote-project.org/wiki/index.php/Annotated\\_Document\\_Default\\_View](http://anote-project.org/wiki/index.php/Annotated_Document_Default_View)

The view for this document includes a list of entities annotated with the respective offsets and a list of sentences with enclosed annotations. An alternative view will be explained further in the next section.

## **Other Entity Recognition Plug-ins**

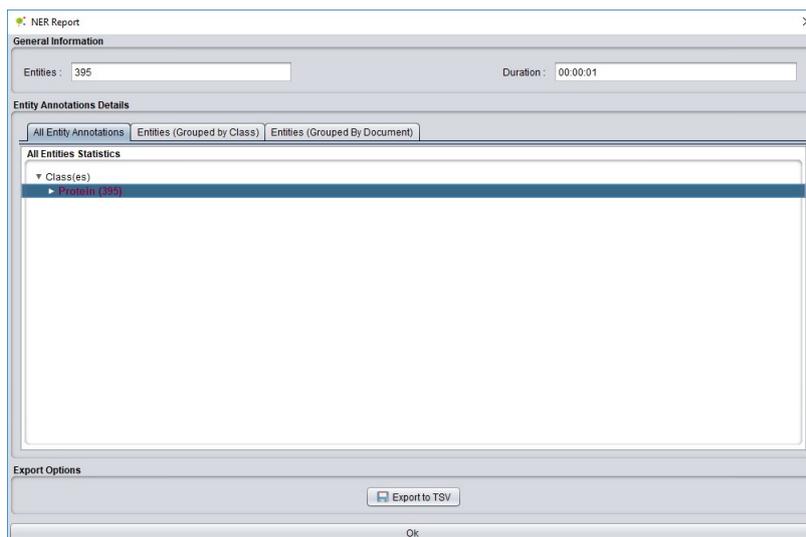
### ***ABNER***

You can also perform an NER process using the Abner algorithm (for more details on how the algorithm works please visit <http://gate.ac.uk/sale/tao/splitch16.html#sec:parsers:abner>).

To check the details for how to use ABNER within @Note:

[http://anote-project.org/wiki/index.php/Corpus\\_Create\\_Annotation\\_Schema\\_By\\_Abner](http://anote-project.org/wiki/index.php/Corpus_Create_Annotation_Schema_By_Abner)

Using default Options for the “E. coli Corpus” we get the following result:



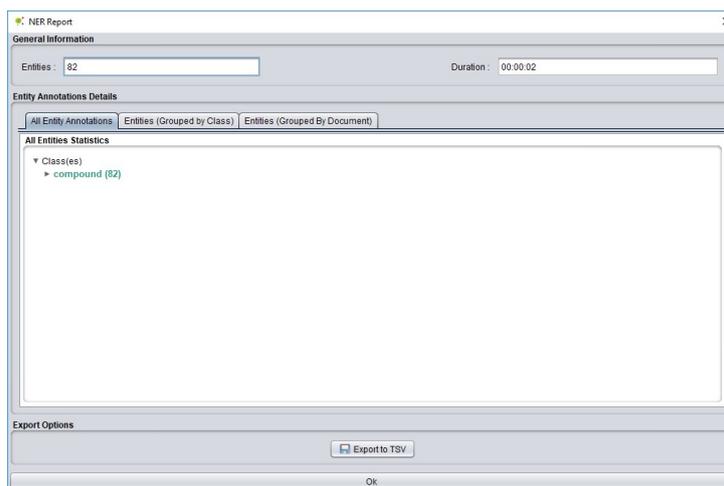
## Chemical Tagger

Another option for NER is the use of the Chemical Tagger NER algorithm (for more details please visit <http://gate.ac.uk/sale/tao/splitch21.html#sec:parsers:chemistrytagger> ).

To see the details of using this algorithm in @Note, follow:

[http://anote-project.org/wiki/index.php/Corpus\\_Create\\_Annotation\\_Schema\\_By\\_Chemical\\_Tagger](http://anote-project.org/wiki/index.php/Corpus_Create_Annotation_Schema_By_Chemical_Tagger)

Using default Options for "E. coli Corpus", the following results would be obtained:



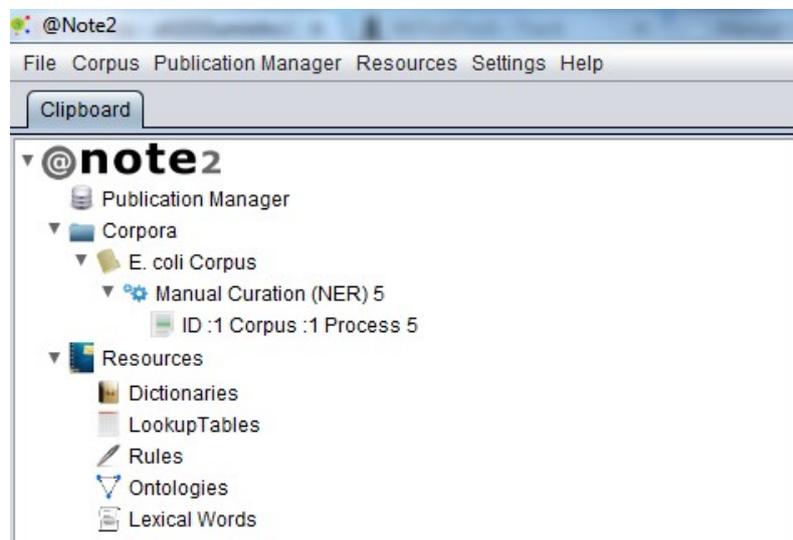
## Manual Curation and Editing annotations

Alternatively to running an NER Process, for entity identification, it is possible to annotate documents directly in text. This process is called “Manual Curation” and it is important to add/remove/edit annotations given by NER Processes, to correct errors or just for annotating documents manually.

To use Manual Curation, the user has to select a Corpus and right click over it, selecting the “Manual Curation” option. Check details in:

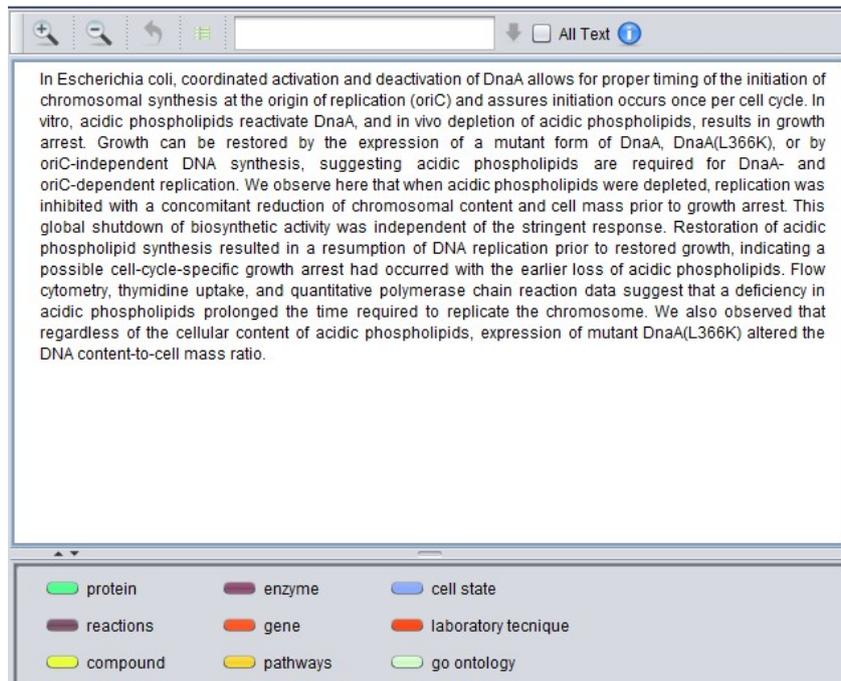
[http://anote-project.org/wiki/index.php/Manual\\_Curation](http://anote-project.org/wiki/index.php/Manual_Curation)

@Note creates a new object, in the Corpus View, called Manual Curation. Now let us load the new process to the clipboard. All Processes (NER, RE or Manual Curation) have an Annotated Documents View that organize documents in a table and allow loading individual Annotated documents to the Clipboard. As an example, let us import Annotated Document with PMID 23233230 to the clipboard.



Once loaded, the Annotated Document has two views (Entity Annotations and Curator View), if derived from NER or Manual Curation Processes, or four if derived from an RE Process (includes a Relations View).

Below, we show an example of the text for an Annotated Document (PMID 23233230):



In the Curator View, it is possible to manipulate the annotations. The user can add, remove or edit entity annotations. Users can explore the top bar menu where they can increase or decrease text font.

Another option is to select/create a Lookup Table (Lexical Resources) to save the changes in the annotations . For example, if lookup tables are selected when an annotation is added to the text, this annotation is saved in a lookup table element that can be used in future NER Processes. See more details about this option in:

[http://anote-project.org/wiki/index.php/Lookup\\_table\\_curator\\_enable](http://anote-project.org/wiki/index.php/Lookup_table_curator_enable)

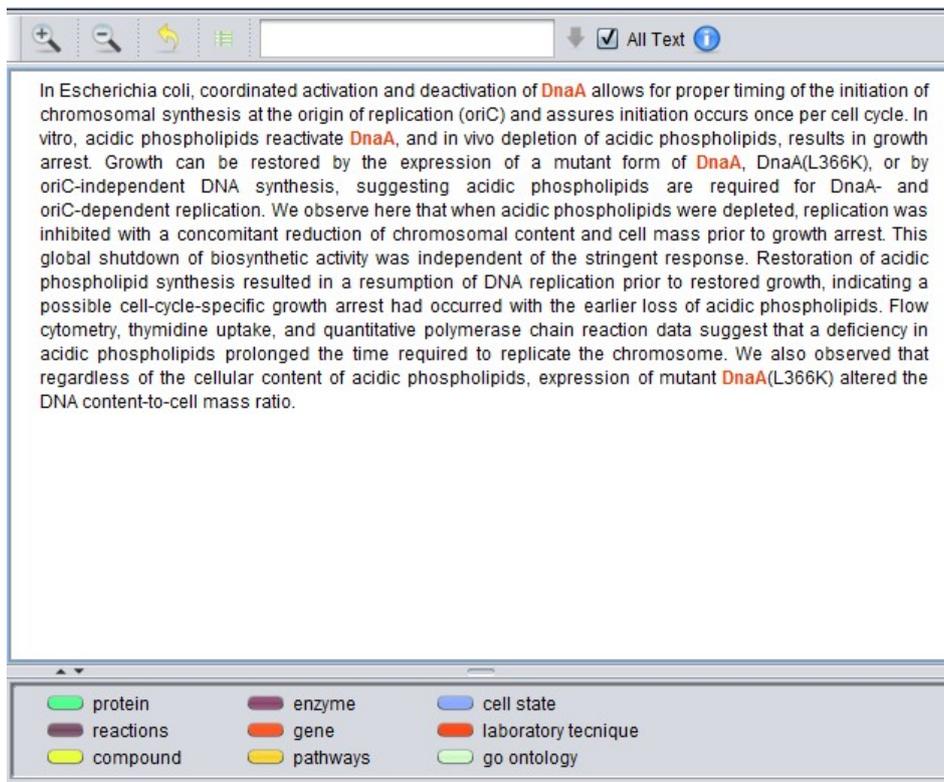
Another option in this menu is the possibility to search words in the text. Also, the button “All text” permits, when selected, to add/remove or edit all the text with the same expression. For example, when the user removes one annotation, all instances for this annotation are removed. If the “all text” option is disabled, just the selected annotation will be removed.

In the menu below, it is possible to view class entities and the color by which they are represented. Here, it is possible to change colors assigned to classes. In the right menu, it is possible to view the list of entities annotated, ordered by classes. Check further information in:

[http://anote-project.org/wiki/index.php/Curator\\_View\\_Annotated\\_Document](http://anote-project.org/wiki/index.php/Curator_View_Annotated_Document)

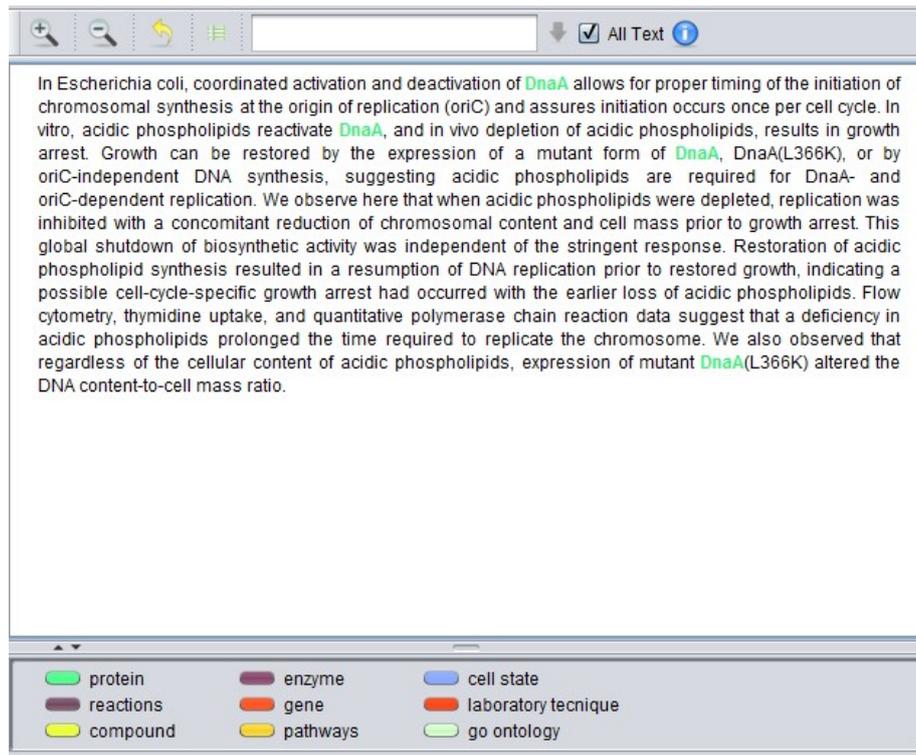
We now make a short introduction to how users can change annotations. To add a new annotation, the user selects the text and right clicks, selecting “add annotation” and a class for the annotation. If the class is not available, the user selects “new class” and a new GUI allows creating a new class.

If the “All Text” option is select, all instances for the term are annotated otherwise just the selected ones are annotated. In the following example, we select DnaA and annotate it as a gene (enabling the all text option).



[http://anote-project.org/wiki/index.php/Curator\\_Annotating\\_a\\_Term](http://anote-project.org/wiki/index.php/Curator_Annotating_a_Term)

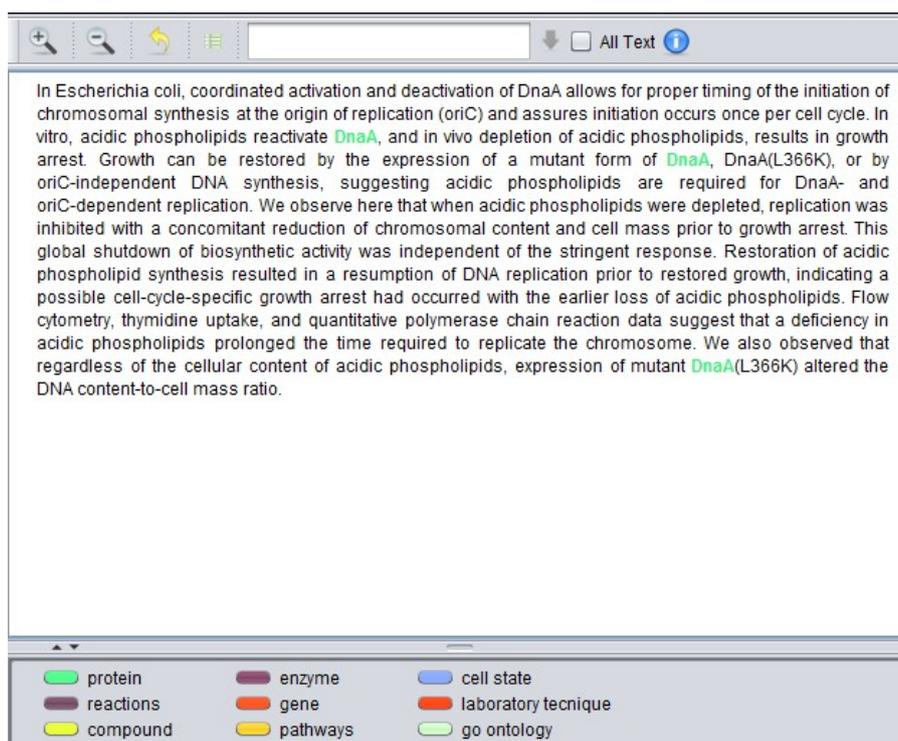
For editing an annotation, the user must select a term annotated in the text and right click it, choosing “correct tag”. A new GUI allows changing the class entity. You select the new class entity and press Apply. It is important to note that if the “all text” button is on, all instances of the term are changed, and otherwise just the selected term is changed. In the following example, all DnaA instances are changed for the class Protein.



Check more details on this option in:

[http://anote-project.org/wiki/index.php/Curator\\_Correct\\_An\\_Annotation](http://anote-project.org/wiki/index.php/Curator_Correct_An_Annotation)

It is also possible to remove annotations. The user must select the annotation and right click on it, selecting “remove tag”. If the “All text” button is pressed, all instances in the text of the annotation are removed; otherwise just the annotation selected is removed. In this example, we removed the first occurrence of the DnaA Protein entity (without the all text option).



Another important feature is a menu option that contains the possibility to link text/annotations to external databases. To get more information on how this works please check:

[http://anote-project.org/wiki/index.php/Curator\\_External\\_Links](http://anote-project.org/wiki/index.php/Curator_External_Links)

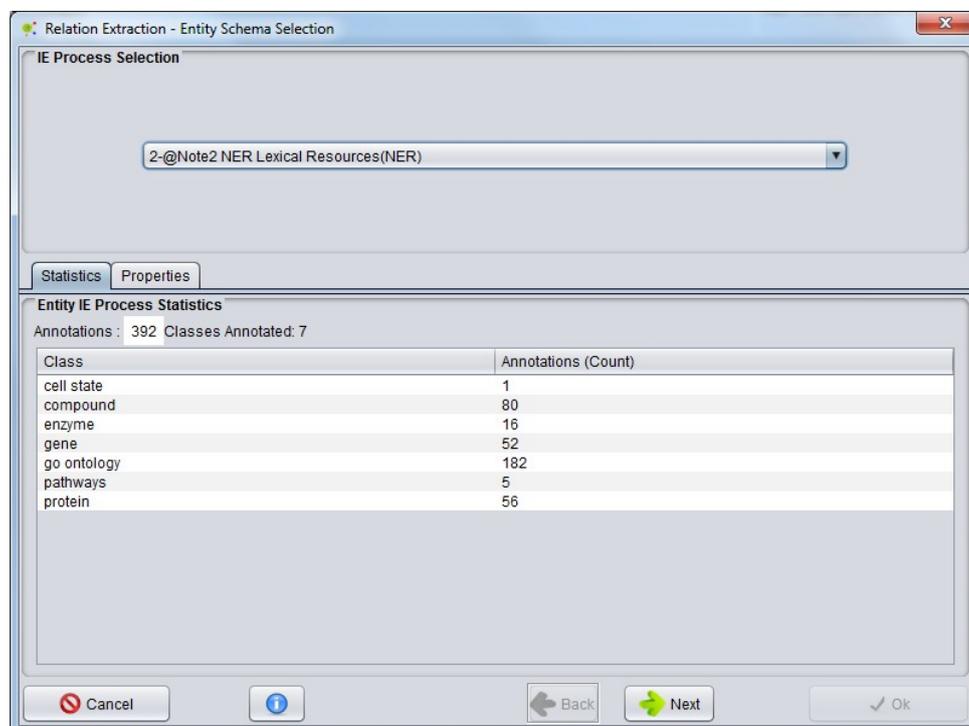
[http://anote-project.org/wiki/index.php/Linkout\\_Management](http://anote-project.org/wiki/index.php/Linkout_Management)

## Relation Extraction

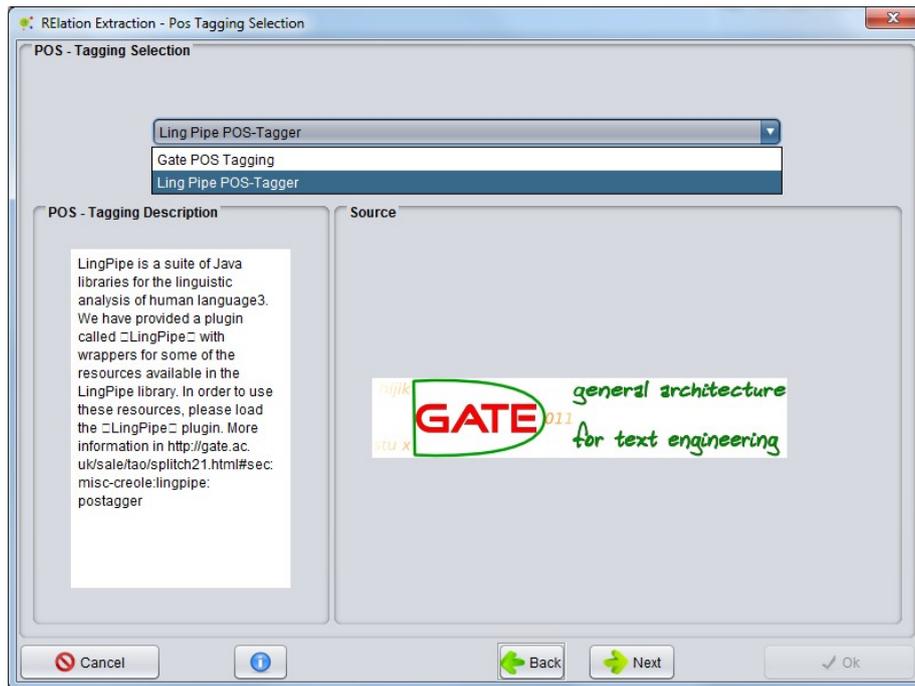
The last steps in this case study is to use the relation extraction plug-in. It permits to gather relations between entities based on an existing corpus NER Process.

### *Base in Pos-Tagging*

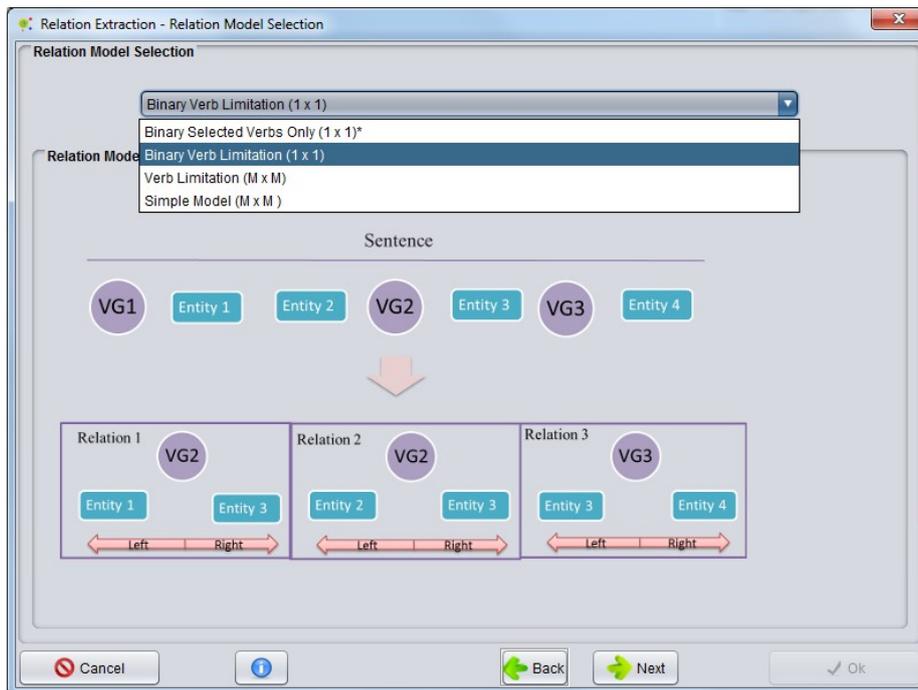
After right clicking on the Corpus data-type and selecting the RE option – Relation Extraction, a GUI to configure the process appears. After this operation was successfully finished a new RE Process appears on Corpus Processes View. In this example, we used the *E. coli* Corpus with NER annotations created above with @Note2 Lexical Resources as a basis for the RE. In the following figures, we show the configuration process used in this example in each step.



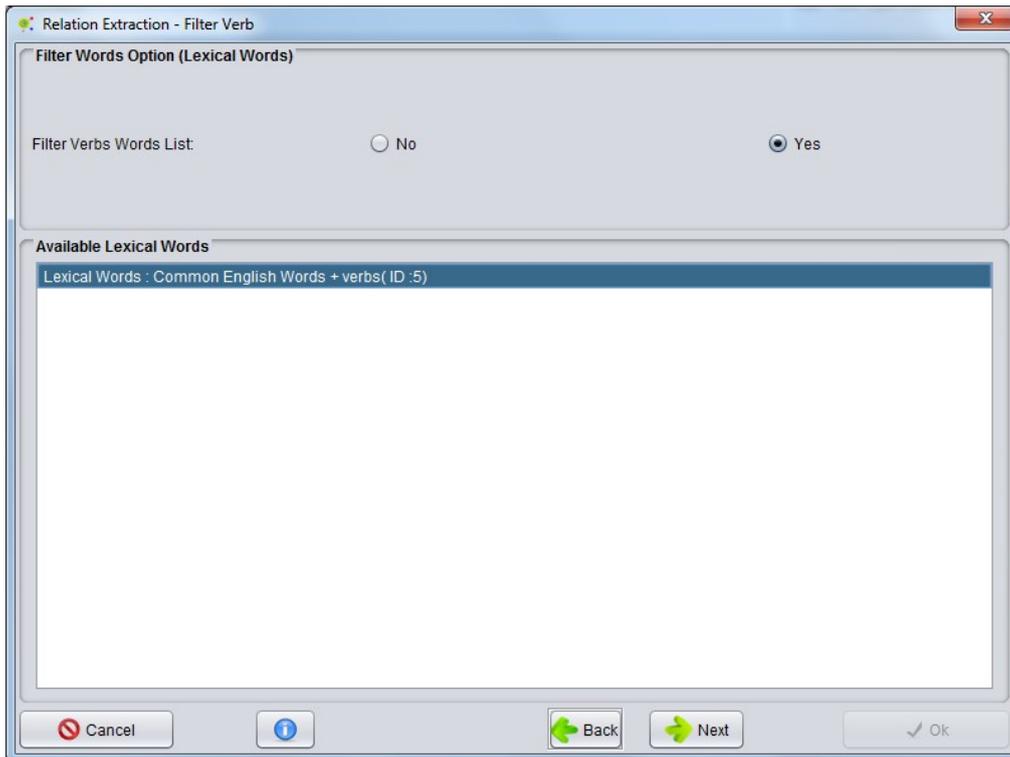
## Ling Pipe POS-Tagging



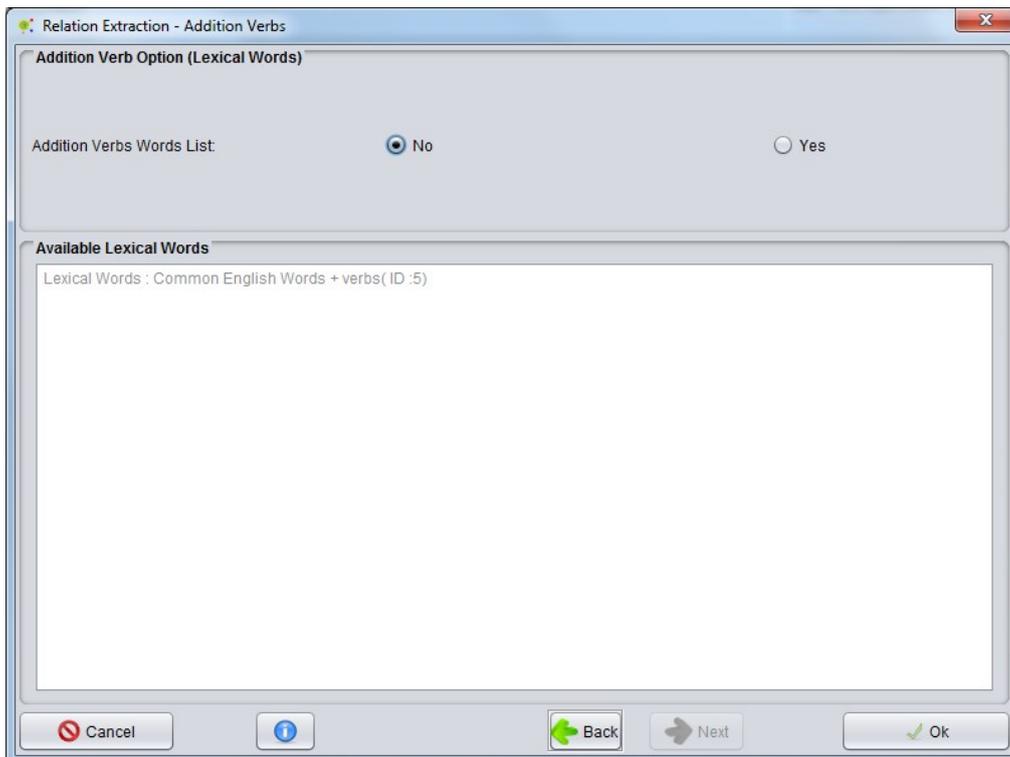
## Verb Limitation Model



Set filter verbs from a word list as true and select the Lexical Words object “Common Words and Verbs” created above (in the Lexical Resources section).



Setting the verb list to add flag to false



Performing the RE Process will create a new RE Process in Process View.

Check further details in:

[http://anote-project.org/wiki/index.php/Corpus\\_Relation\\_Extraction](http://anote-project.org/wiki/index.php/Corpus_Relation_Extraction)

Now, you can view the result of an RE Process by loading it to the clipboard. When selecting an RE Process in the clipboard, four view tabs are available. Two are identical to NER Processes (Annotated Documents View and Entity Stats View). Another two are designed to view relation statistics and to view the relations extracted.

For further information:

[http://anote-project.org/wiki/index.php/Process\\_Relations\\_Statistics](http://anote-project.org/wiki/index.php/Process_Relations_Statistics)

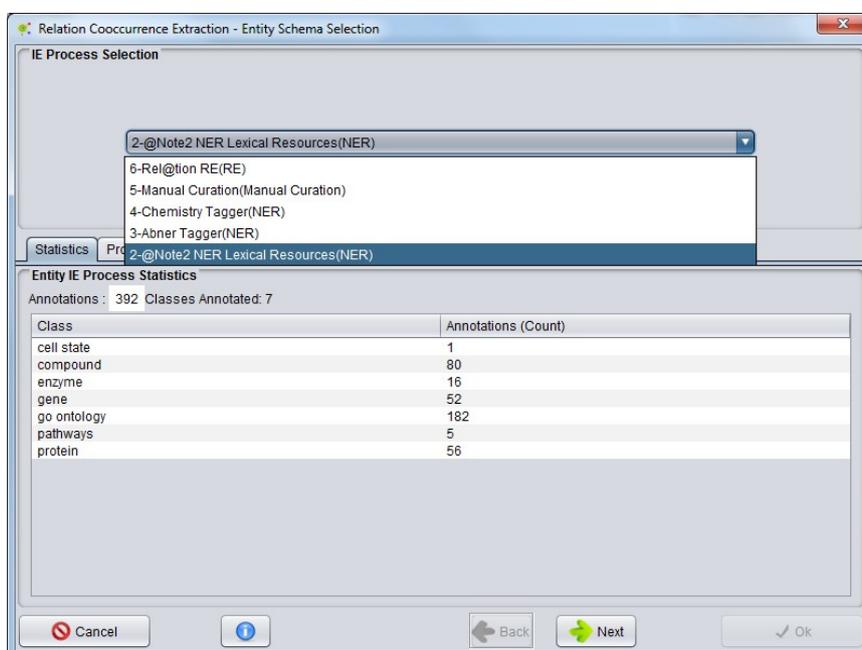
[http://anote-project.org/wiki/index.php/Curator\\_Relations\\_View\\_Annotated\\_Document](http://anote-project.org/wiki/index.php/Curator_Relations_View_Annotated_Document)

### *Based in Co-occurrence*

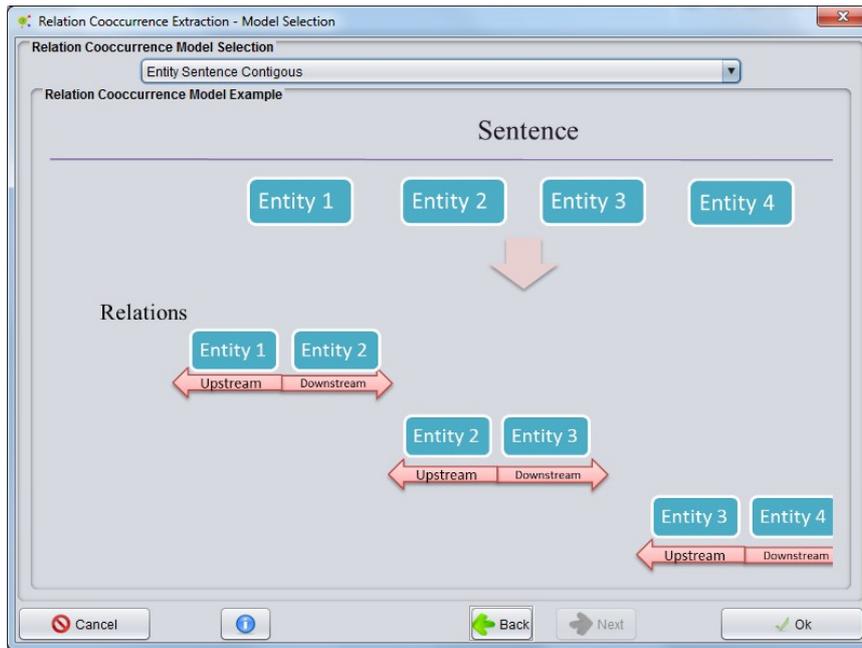
You can perform a new RE co-occurrence process based on entity proximity (i.e. terms in the same sentence) over a corpus. After right clicking on the Corpus data-type and selecting the RE option – Relation Co-occurrence, a GUI to configure the process appears. After this operation is successfully finished, a new RE Process appears on Corpus Processes View.

In this example, we used the E. coli Corpus with NER annotations based on Lexical Resources as above.

In the screenshots the full configuration for this example is shown:



Select model:



Further details can be found:

[http://anote-project.org/wiki/index.php/Corpus\\_Relation\\_Coocurrence\\_Extraction](http://anote-project.org/wiki/index.php/Corpus_Relation_Coocurrence_Extraction)