



User Manual

Arabic Named Entity Extraction

NERSys



Version 1.00

Table of Contents

Coverage of this document	3
Copyright information.....	3
Licensing the software	4
Working with the samples	5
Setting the samples database.....	7
Showing the samples ribbon.....	8
System main window	9
Working with user input	11
Single phrase input.....	11
Saving a phrase samples.....	11
Removing the samples.....	13
System settings	14
Control Pane.....	15
Output customization.....	16
Effect of selected features.....	17
Setting the transliteration system.....	18
Working with files	19
Loading a single file.....	19
Loading multiple files.....	19
Removing text files from the list.....	19
Customizing the system	20
Preparing the text for input.....	20
Typical use cases	21
1. Content analysis.....	21
2. Extracting selected entities to file.....	22
3. Content classification.....	23
4. Listing files of specific classification.....	24

Coverage of this document

This document is the full text user manual for (Arabic Named Entity Extractor) version 1.00.

Copyright and Illegal Usage Policy

Disclaimer of Liability

In preparation of this document, every effort has been made to offer the most current, correct, and clearly expressed information possible. Nevertheless, inadvertent errors in information may occur. In particular but without limiting anything here, Kalmasoft disclaims any responsibility for typographical errors and accuracy of the information that may be contained in this manual. The information and data included herein have been compiled by our staff from a variety of sources, and are subject to change without notice to you. Kalmasoft makes no warranties or representations whatsoever regarding the quality, content, completeness, suitability, adequacy, sequence, accuracy, or timeliness of such information and data. In any situation where the official sent documents of Kalmasoft differ from the text contained in this manual, the official documents take precedence. The information and data made available in this document are provided "as is" without warranties of any kind.

Disclaimer of reliability

Kalmasoft makes no representations or warranties regarding the condition or functionality of this software, its suitability for use, or that this will be uninterrupted or error-free.

Disclaimer of damages

By using Kalmasoft MAPSSeman© Lite (Arabic Named Entity Extractor), you assume all risks associated with the use of this software, including any risk to your computer, software or data being damaged by any virus, software, or any other file which might be transmitted or activated via this software. We shall not in any event be liable for any direct, indirect, punitive, special, incidental, or consequential damages, including, without limitation, lost revenues, or lost profits, arising out of or in any way connected with the use or misuse of the software or lack of information in this manual.

Disclaimer of endorsement

Kalmasoft does not favor one group over another, and any references herein to any country, organizations, specific commercial products, process, or service by trade name, trademark, manufacturer, or otherwise, do not necessarily constitute or imply its endorsement or recommendation by us.

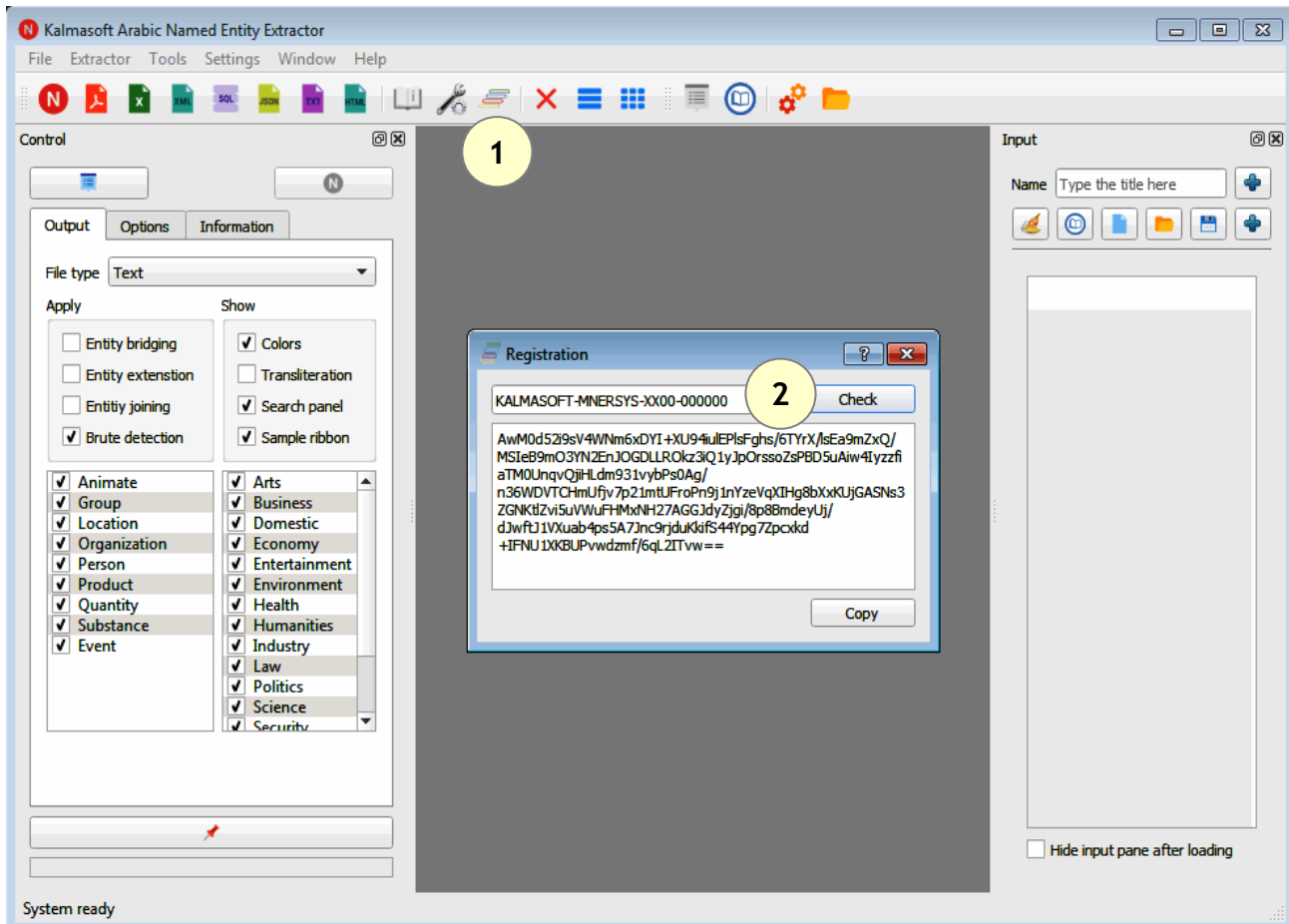
Copyright information

The graphics and contents on this manual are the copyrighted work of Kalmasoft and contain proprietary trademarks and trade names of Kalmasoft. No part of this document can be copied without a prior written consent from Kalmasoft.

Licensing the software

Download the software rar archive and install in the same folder.

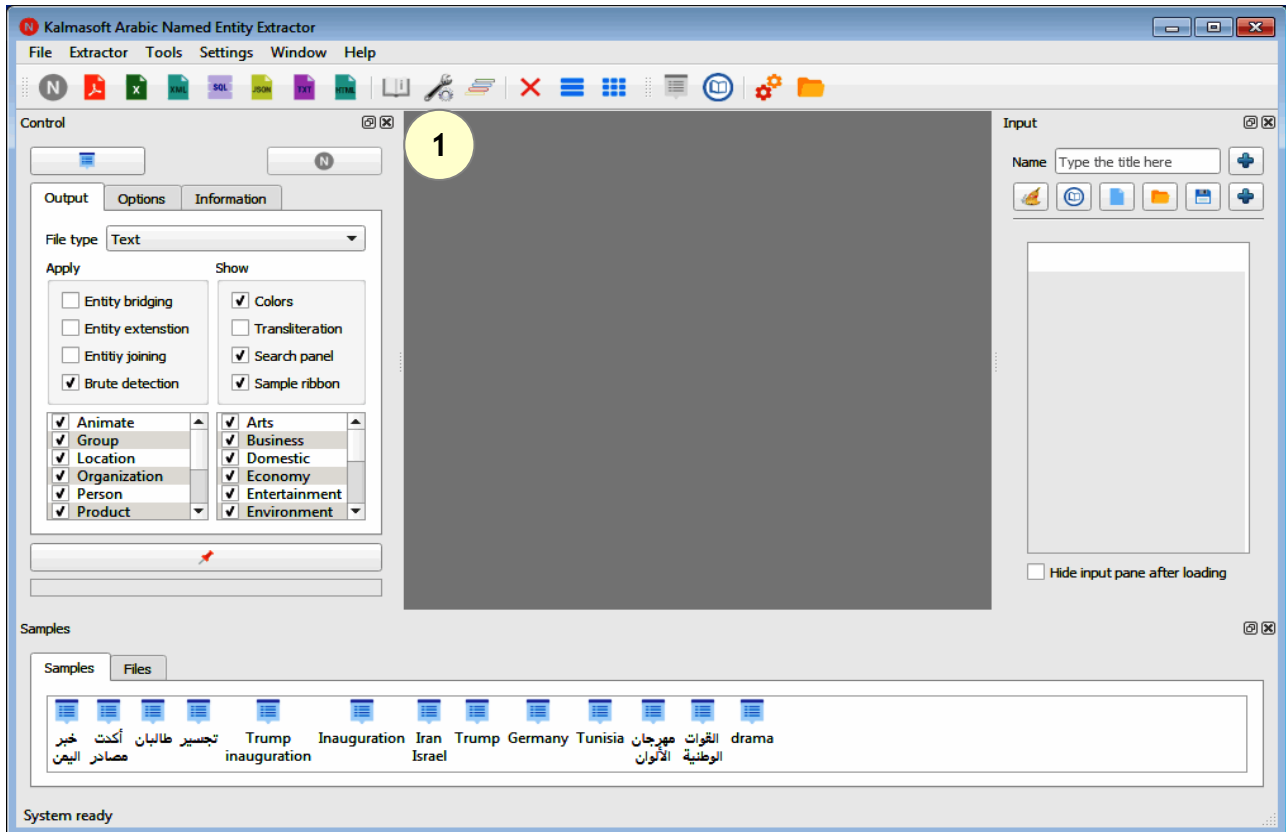
The evaluation copy needs to be registered to use all features.




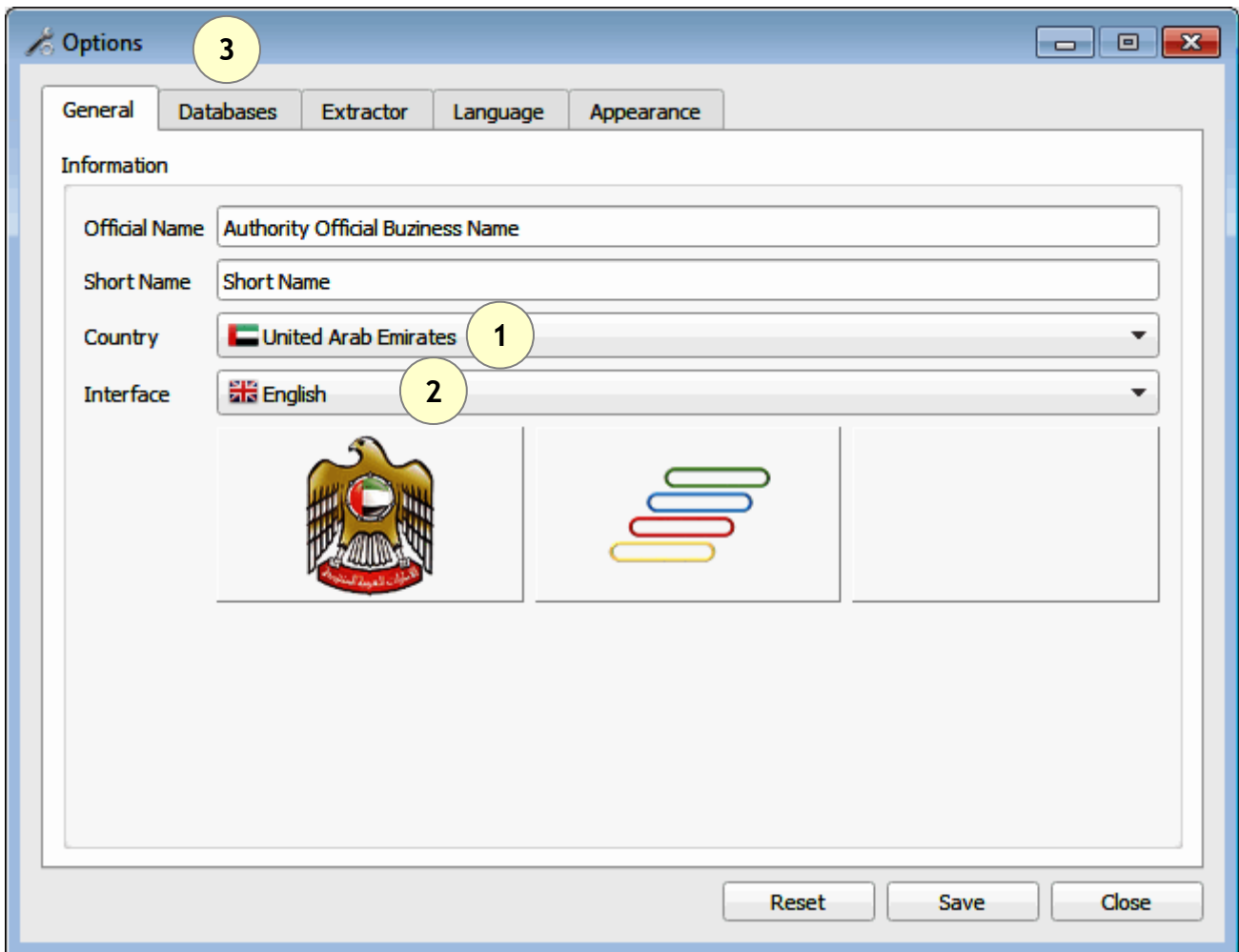
- Enter the registration key (KALMASOFT-MNERSYS-XX00-000000)
- Click (Check) and send the generated key to Kalmsoft to get your registered copy.

Working with the samples

The system comes ready by default with few arbitrary text samples, you need to prepare the system before testing with those samples.

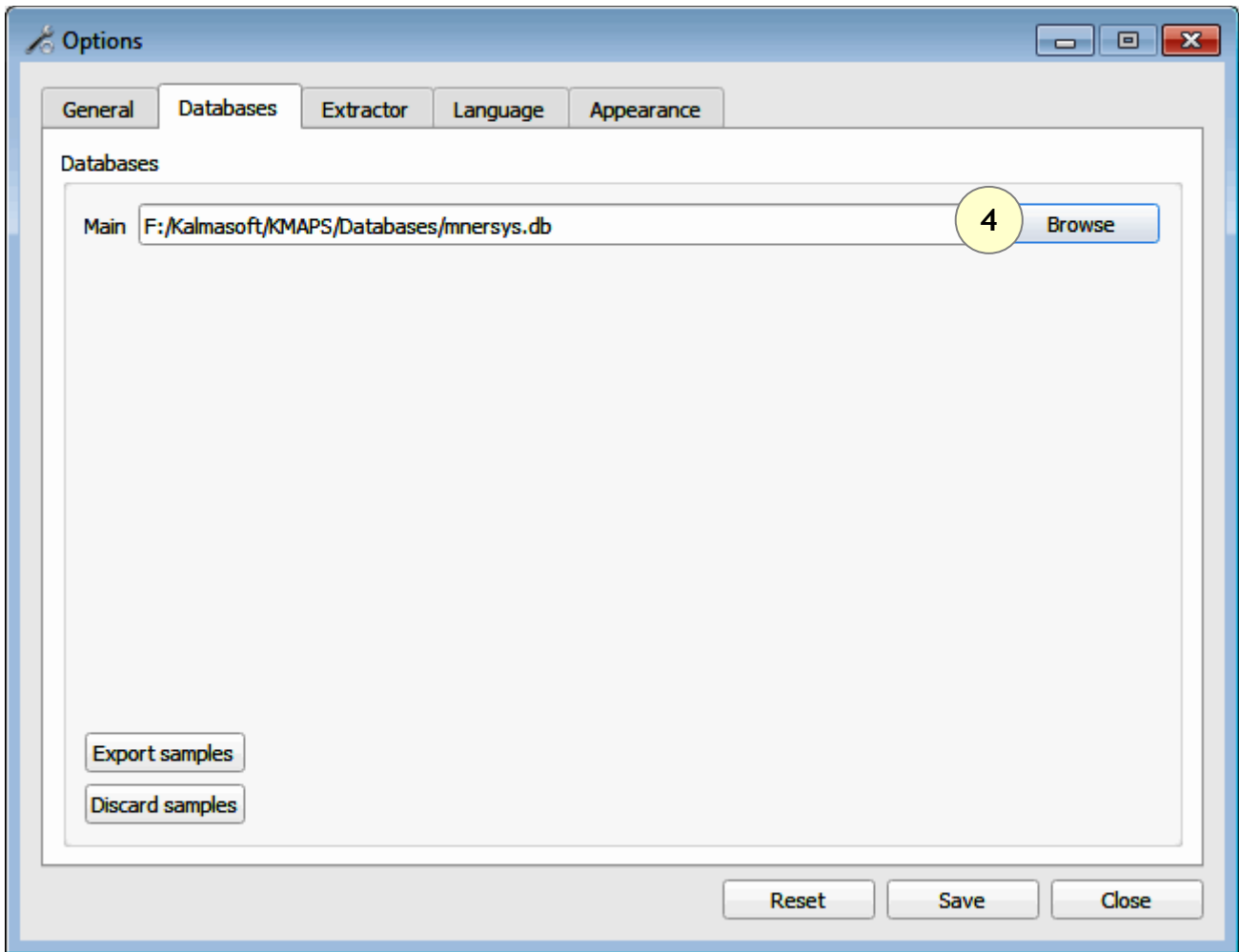


Click on the option button (), the Options dialog will show up.



- Set country and language according to your preferences.
- Switch to the Database tab (3).

Setting the samples database




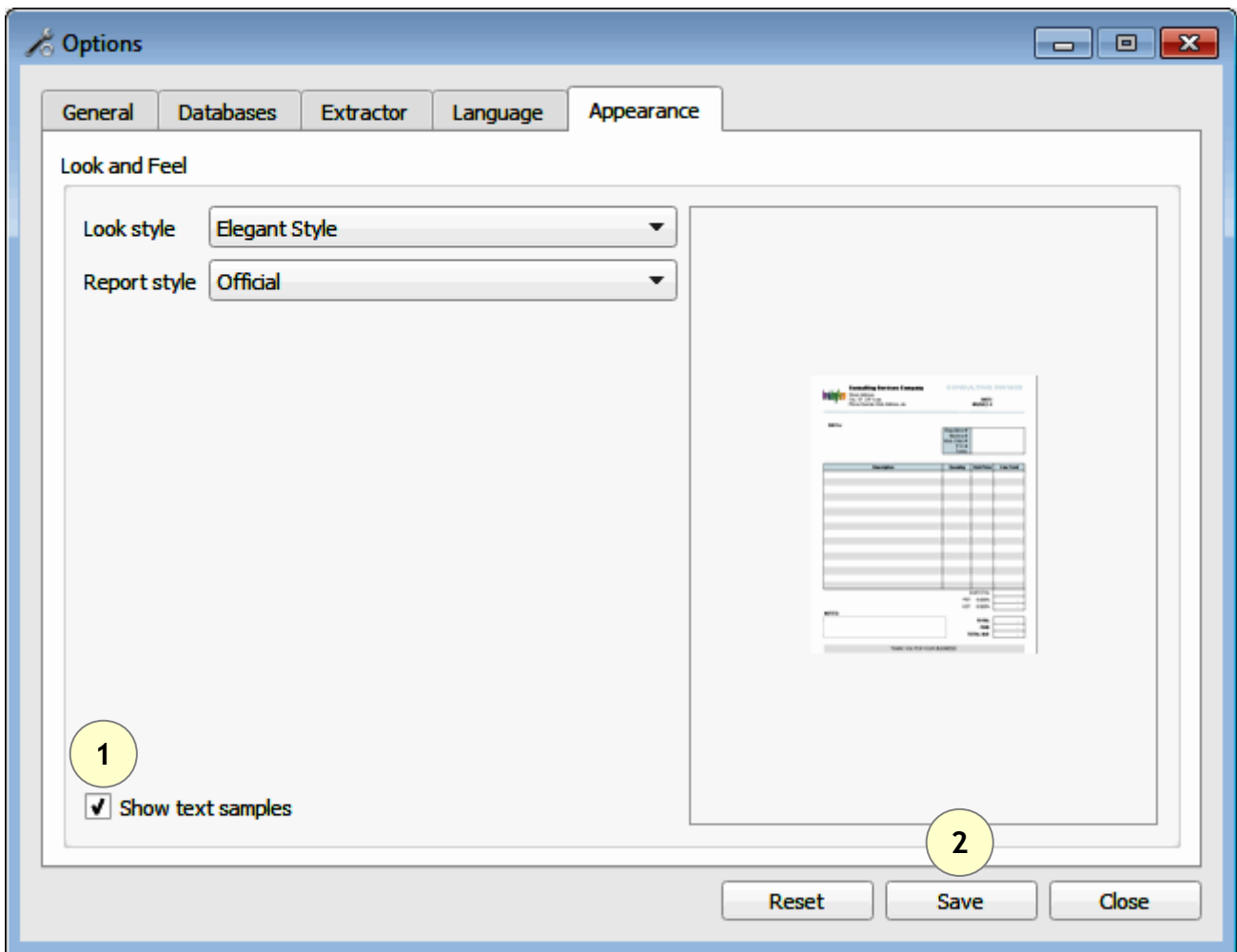
- Set database to **C:/Program Files/Kalmasoft/KMAPS/Databases/mnersys.db**

Depending on your installation folder, the above path may be different but the part (**Kalmasoft/KMAPS/Databases/mnersys.db**) should always remain valid.

- Click (Browse) to set the database as highlighted above.

Showing the samples ribbon

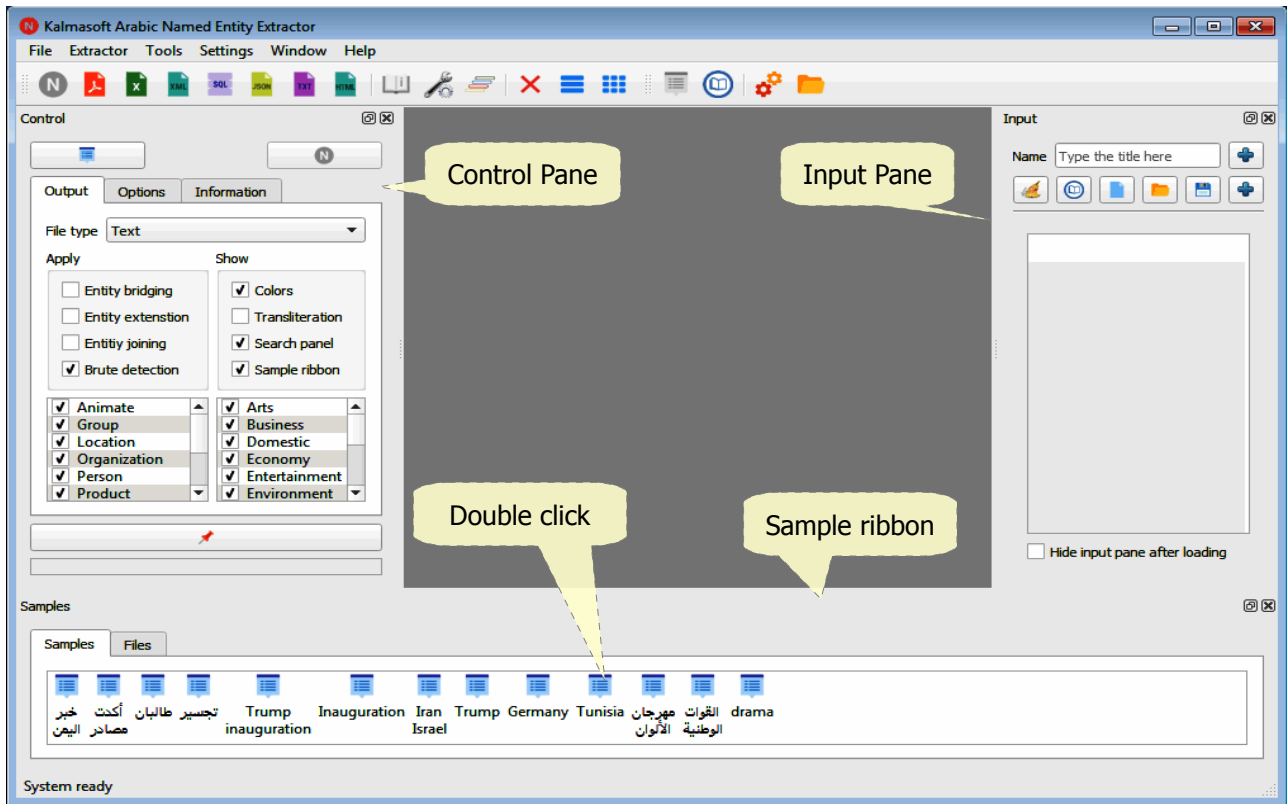
- Click on the Options button () on the menu bar.
- Switch to Appearance tab, check the box (Show text samples)
- Click (Save)



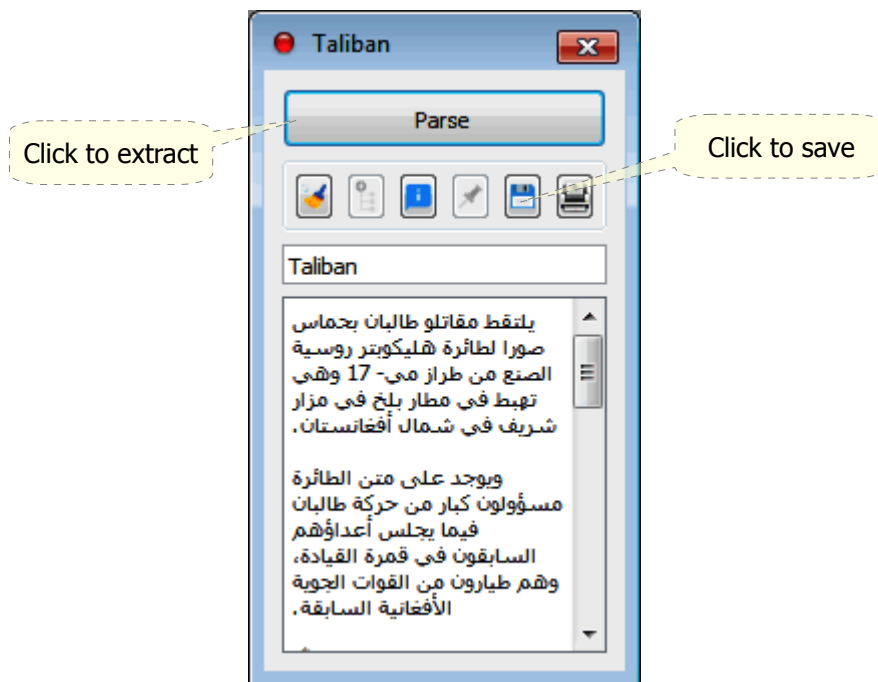
The samples ribbon should appear immediately.

System main window

The system has main window interface with subwindows and panes.



Double click any phrase from the sample ribbon, a phrase card will appear.



The results window will appear shown below.

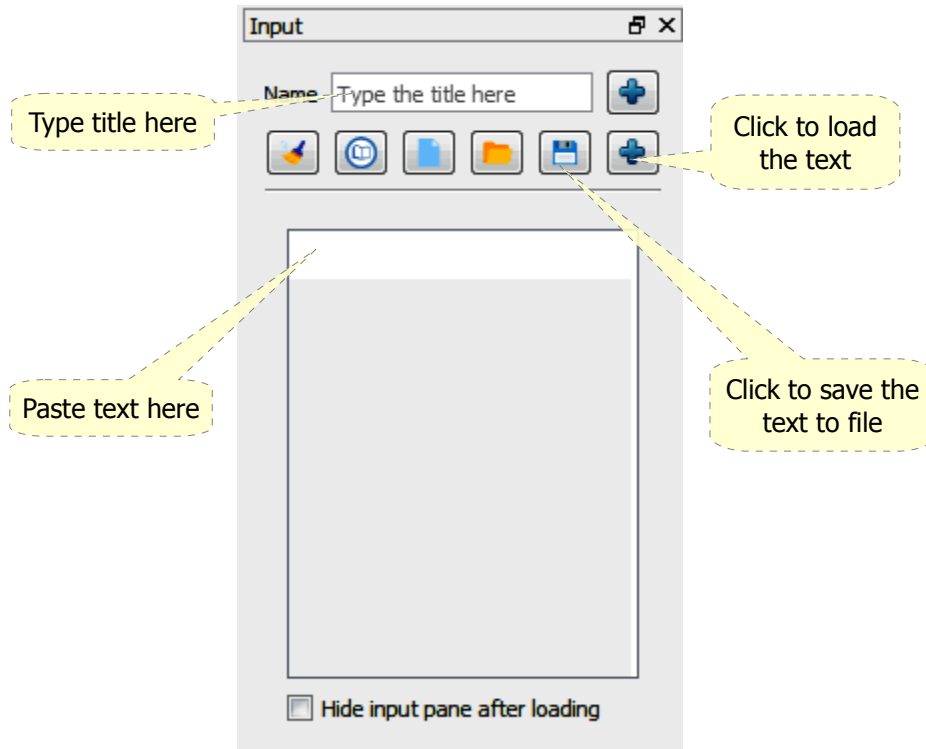


Hover over the text to show further explanation on each extracted entity, the same feature is available if output is exported as HTML file.

Working with user input

Single phrase input

The single phrase input is for testing purposes prior to extraction of a large amount of text, use this input to enter one page of text.

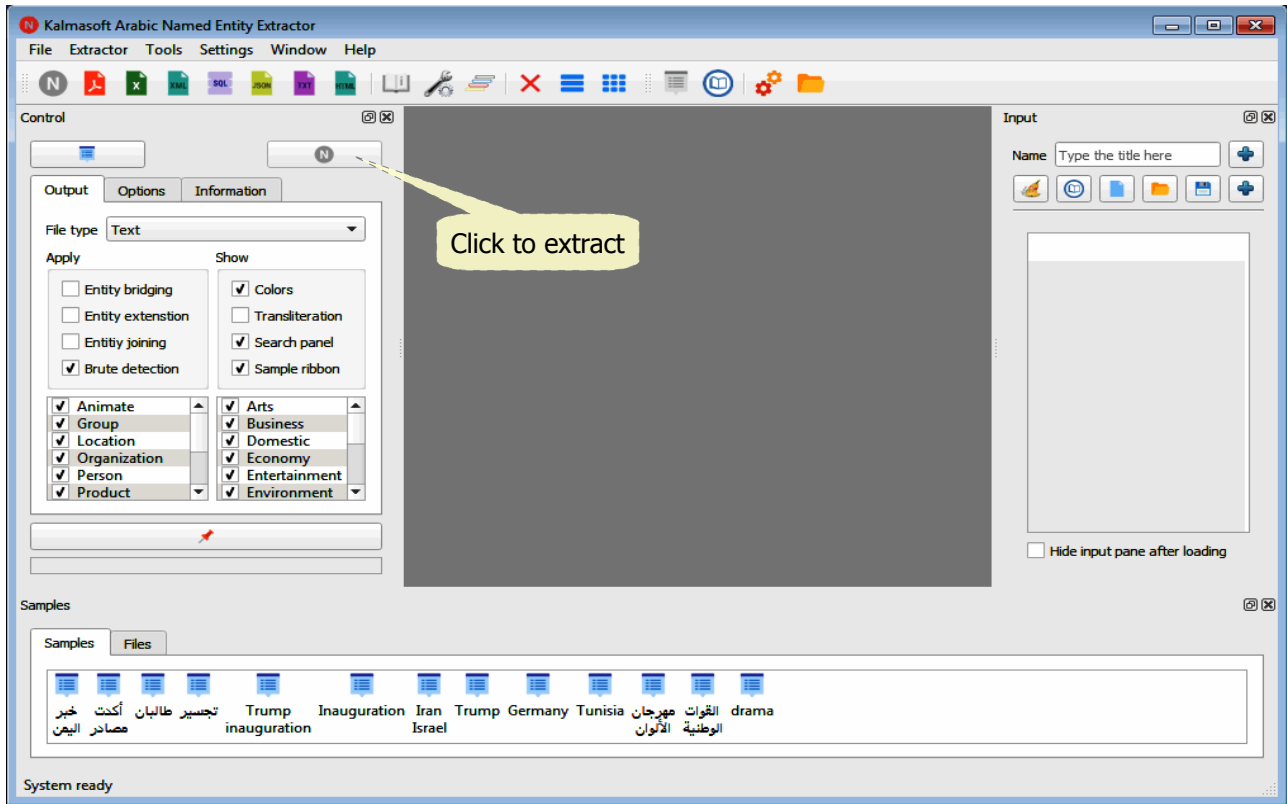


- Paste a text, a title for the phrase is necessary for reference.
- Click the (+) to load, a card will show up in the main area.
- Click the the [Parse] button on top to start extraction, you may manual enter as many phrases using cards this way.
- Click on the (N) button in the Control Pane to start extraction, the result will be displayed on a different window that will show up.

Saving a phrase samples

Once loaded, the single phrase can be saved for later extraction, click the (📄) button to save the text to the internal database, the saved text will show in the sample ribbon at the bottom of the main window.

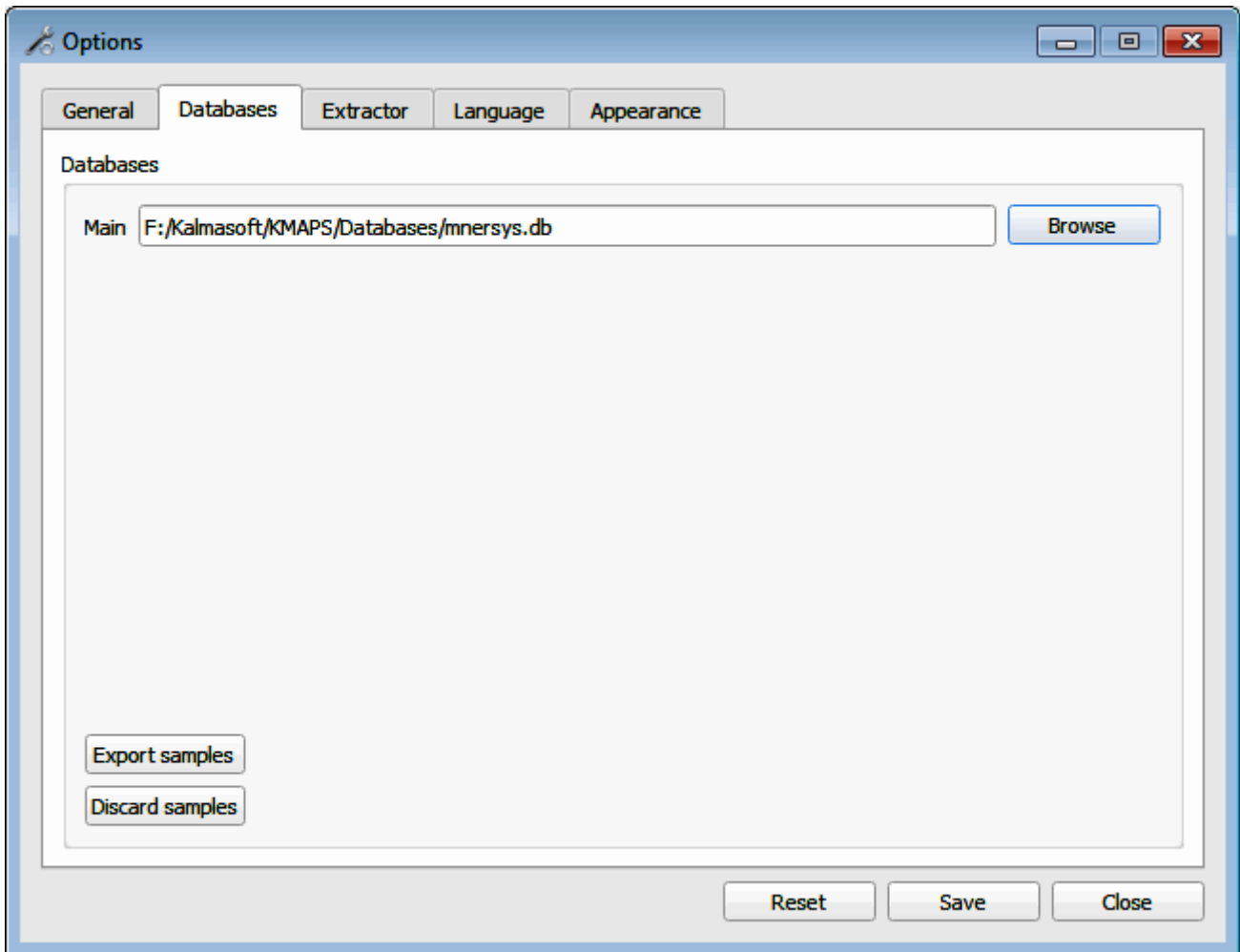
Click on the (N) button in the Control Pane to start extraction, the result will be displayed on a different window that will show up.



Removing the samples

Samples in the ribbon may be deleted completely to spare a space for user's own samples, to remove only few samples, click on text icon then press (Delete) from the keyboard, to remove all samples open the Options menu switch to the (Databases) tab.

- Click the (Discard samples) button, this will remove all samples permanently.
- Click the (Export samples) to keep a copy of the samples before deletion.



System settings

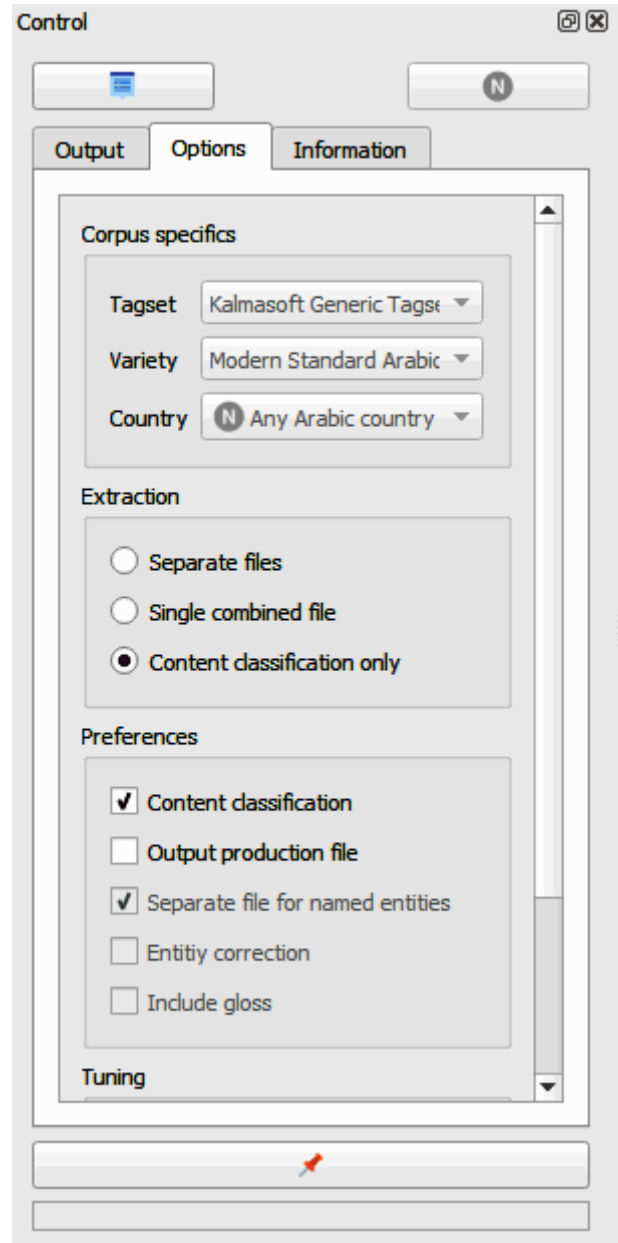
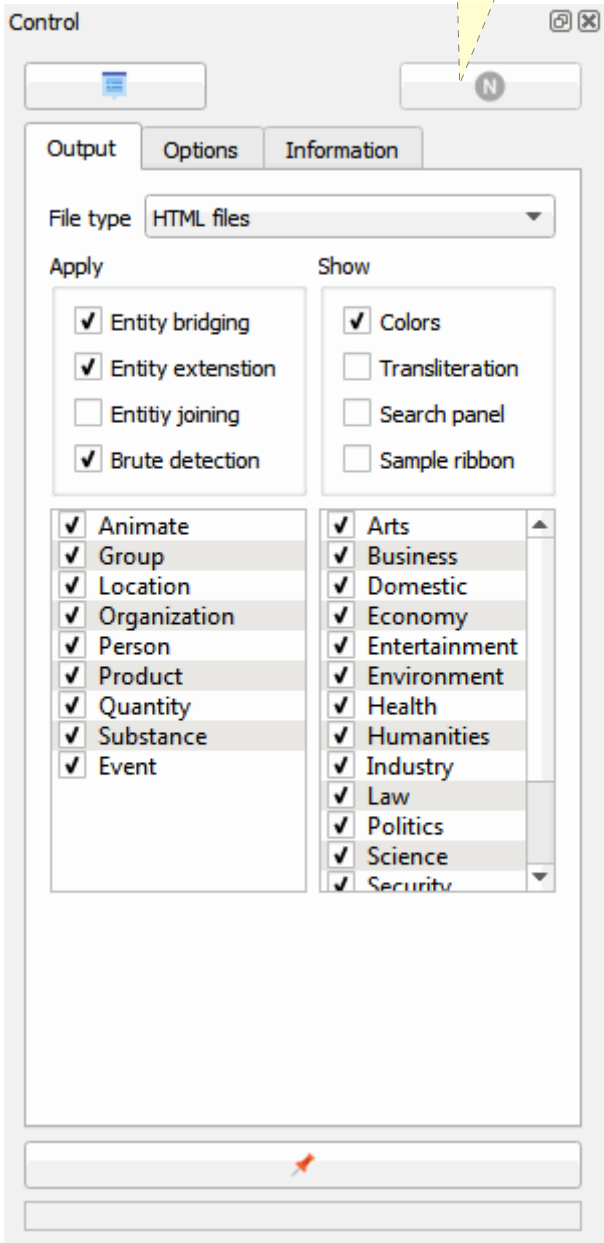


If the extraction process is successfully done, the above screen will show up with different labels and colors. The system does not fully suggest error correction or allow doing correction directly in place from the interface, retype or reload the correct text and run the extraction process again to get rid of the errors messages.

Control Pane

The Control Pane determines options specific to the input and output and also contains the main extraction button.

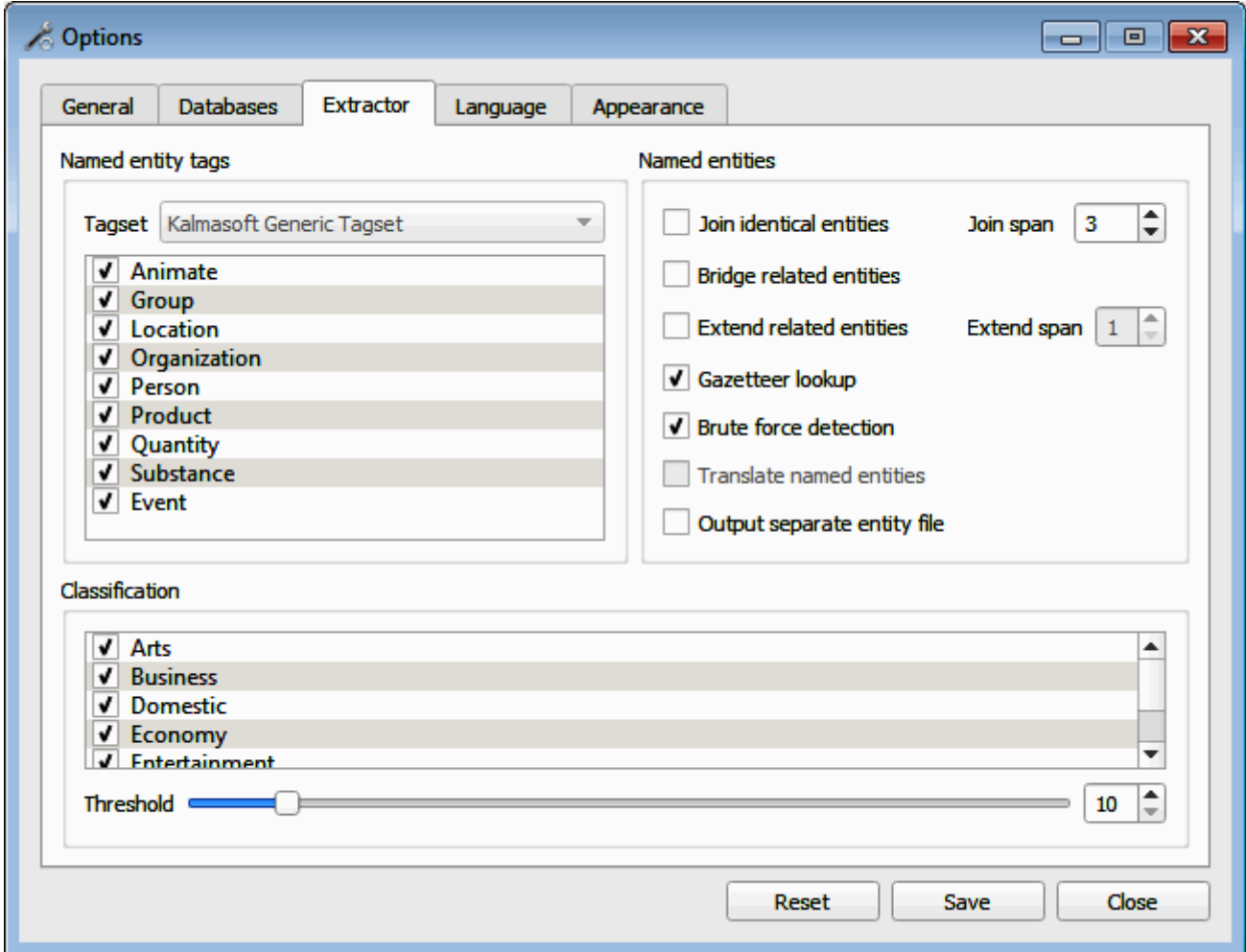
Click to extract



Click the extraction button (**N**) to start the extraction process on the Arabic text, check the other controls as appropriate.

Output customization

System overall behavior can be controlled through this interface, all resources of the system and handling of Arabic text can be set from here.



Effect of selected features

Applying smantic features to the text will have the following effects:

Extension

Extending the same label to words either before, after, or both sides of the currently recognized entity word.

Before extension	After extension
في إطار VNT تظاهرة صفاقس PPL عاصمة للثقافة العربية	في إطار VNT تظاهرة صفاقس PPL عاصمة للثقافة العربية
وأُسند ORG النادي الثقافي العربي	وأُسند ORG النادي الثقافي العربي

Joining

Joining is basically linking two adjacent words or sentences in one semantic unit.

Before joining	After joining
يلتقط GRB مقاتلو GRB طالبان GRB بحماس صورا	يلتقط GRB مقاتلو طالبان GRB بحماس صورا
في LOC شمال LOC أفغانستان	في LOC شمال أفغانستان
رئيسا PER ترامب PER حفل تنصيب VNT	رئيسا VNT حفل تنصيب ترامب PER

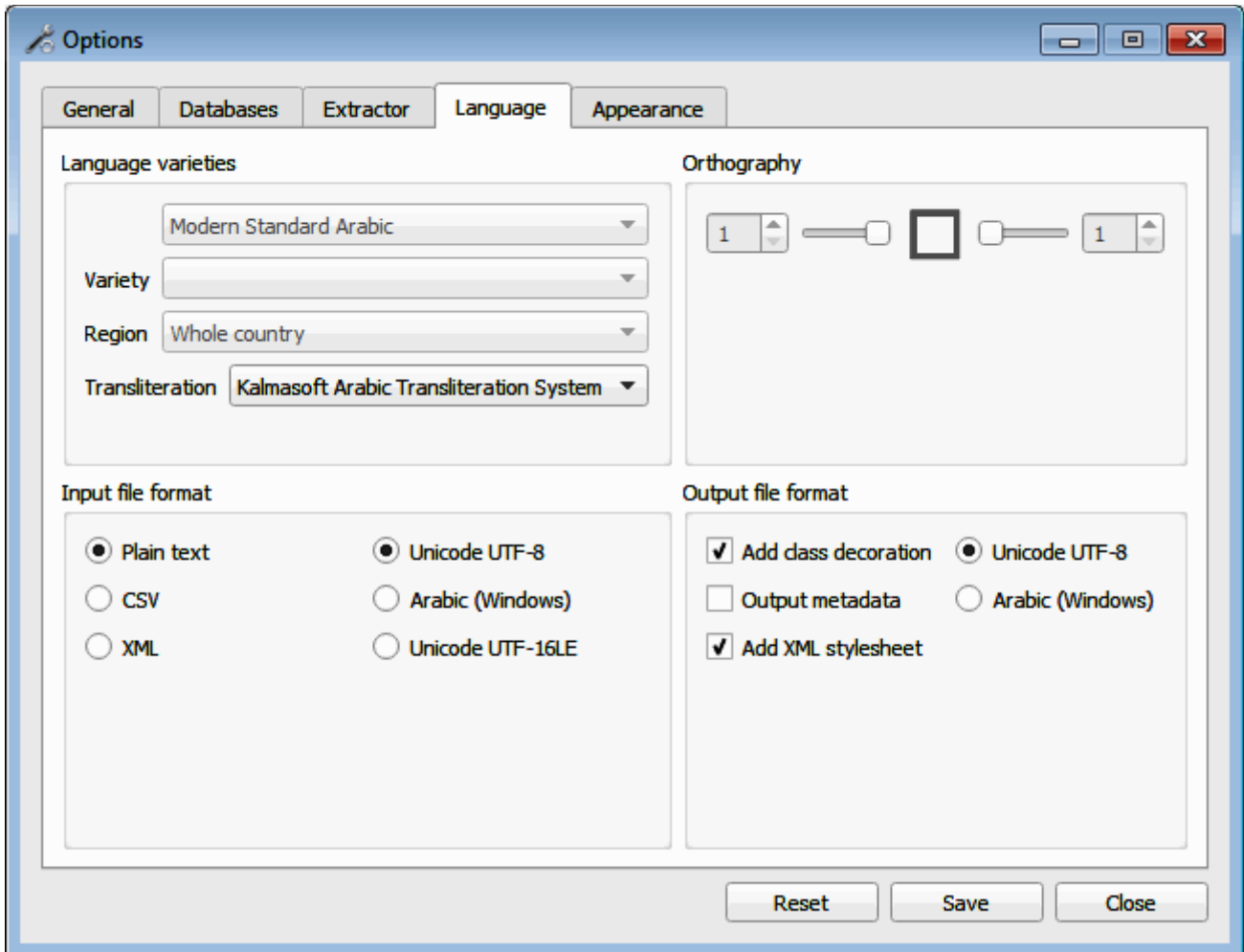
Bridging

Bridging is basically merging a middle word, usually particle, to the rest of two adjacent words.

Before bridging	After bridging
سمعنا برجوع خالنا PER ابن PER علي PER محمد	سمعنا برجوع خالنا PER ابن علي محمد
في LOC لبنان VNT الحرب الأهلية	VNT الحرب الأهلية في لبنان
من DAT جانفي DAT يوم التاسع والعشرين	DAT يوم التاسع والعشرين من جانفي

Setting the transliteration system

The system supports four NER tagging systems, both standard and for academic use, Kalmasoft Arabic Transliteration System (KATS) is the default.




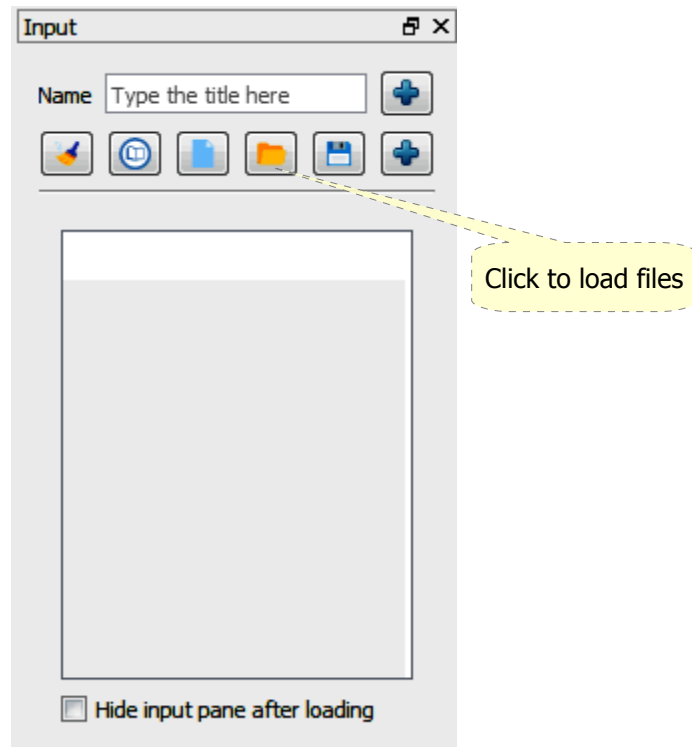
[Add class decoration] feature will prepend the classification label to the file name for easy sorting as shown in the example below, the string “Output_” is added by default to all output files.

Input file name	Output file name
sampleFile-4-20210712.txt	HEALTH_sampleFile-4-20210712.txt
sampleFile-5-20200410.txt	SPORT_sampleFile-5-20200410 .txt

Working with files


Loading a single file

A single Arabic text file can be loaded from the Input Pane, the text file must be a UTF8 encoded text file without any headers. Click () button and browse to the text file, the file will be added to the ribbon ready for loading at any time.




Loading multiple files

Multiple files of Arabic text can be loaded for batch extraction using the Input Pane, the text files must be encoded in UTF8 without any headers.

- Click on button () to browse the files folder.
- Select all files to be added to the waiting list ready for loading at any time you decide to process them.

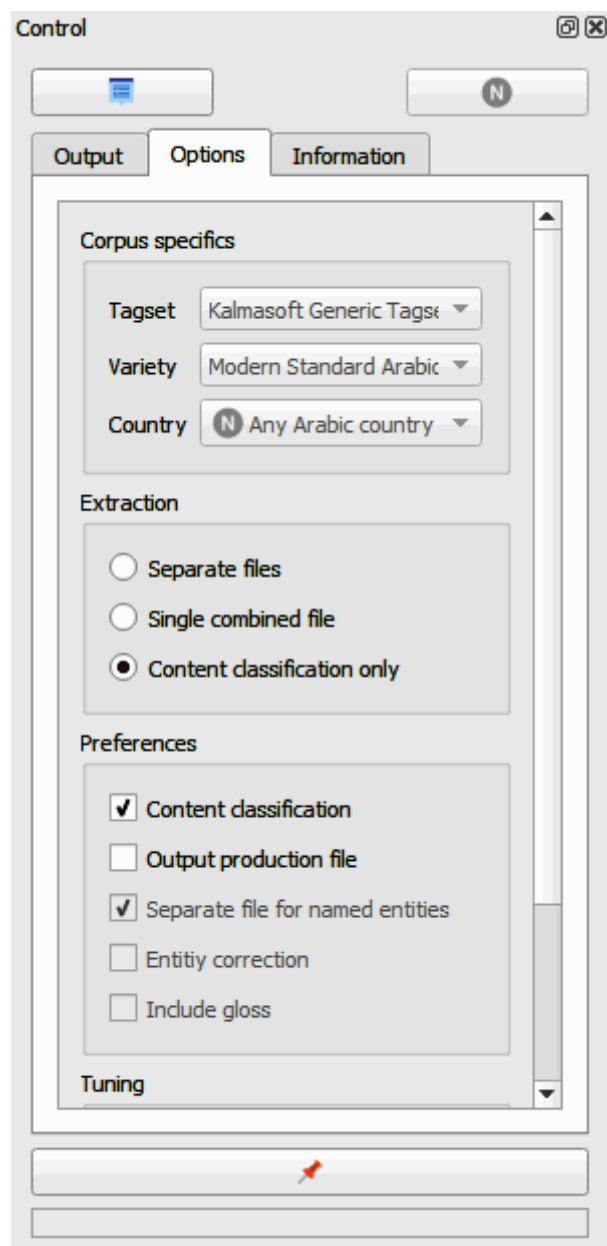
Removing text files from the list

On clicking the file you may either load it or remove it from the list by clicking on the () button. This will only remove the file from the list not deleting the original final.

Customizing the system

Preparing the text for input

The system is very sensitive to the input data, it makes several tests to before the extraction process actually begins, these tests are of orthographic nature in general, they include checking for properties such as length of the text, the system also does some other covert arrangements including the removal of unnecessary characters such as white spaces.



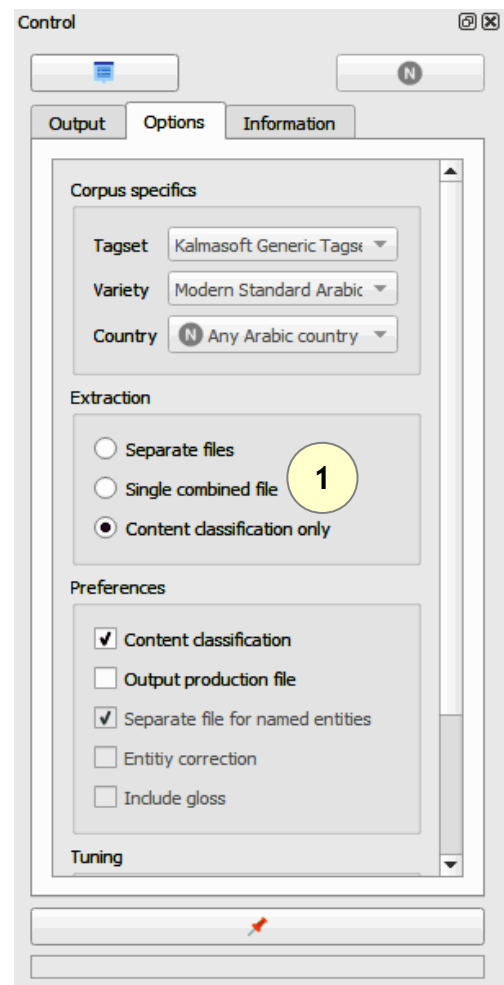
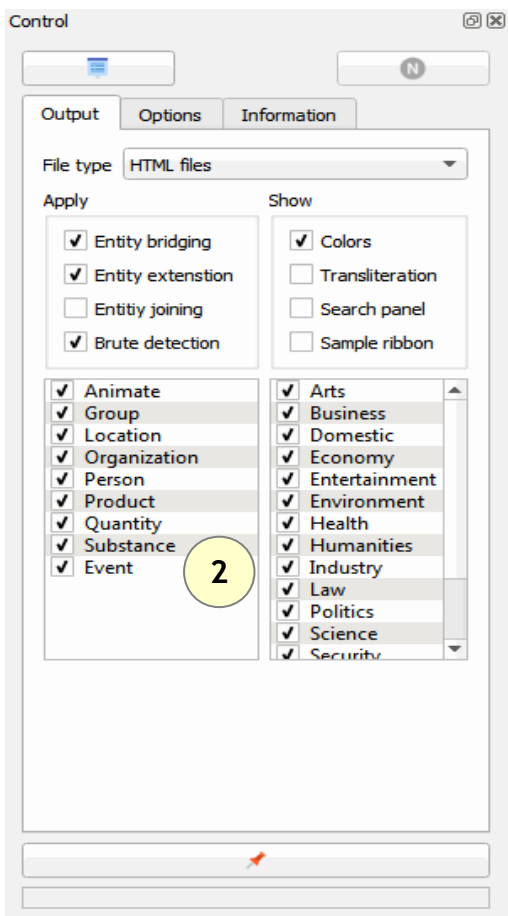
Typical use cases

With batch files NERSys can be used in many useful ways, below are three typical use cases.

1. Content analysis

To analyze text or file content including charts, statistics, and listing of entity names from a single or group of files.

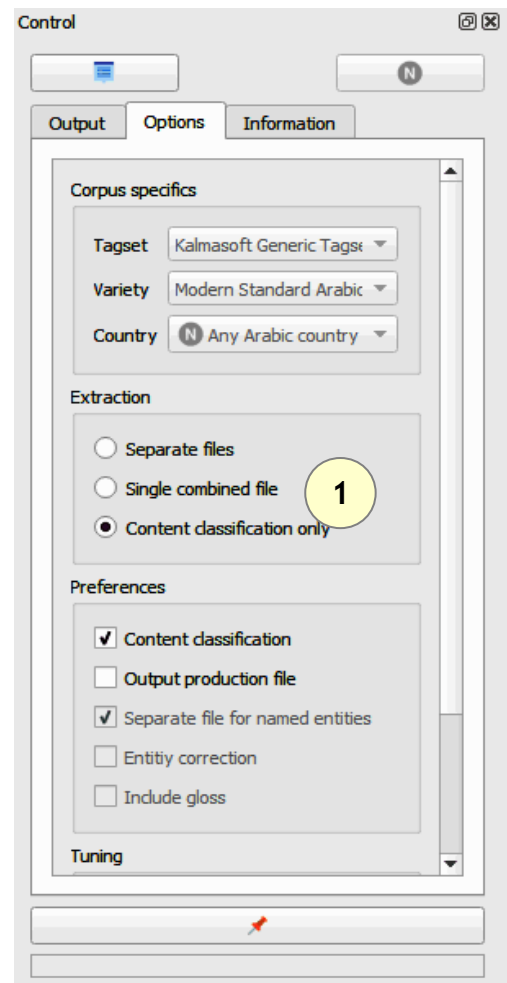
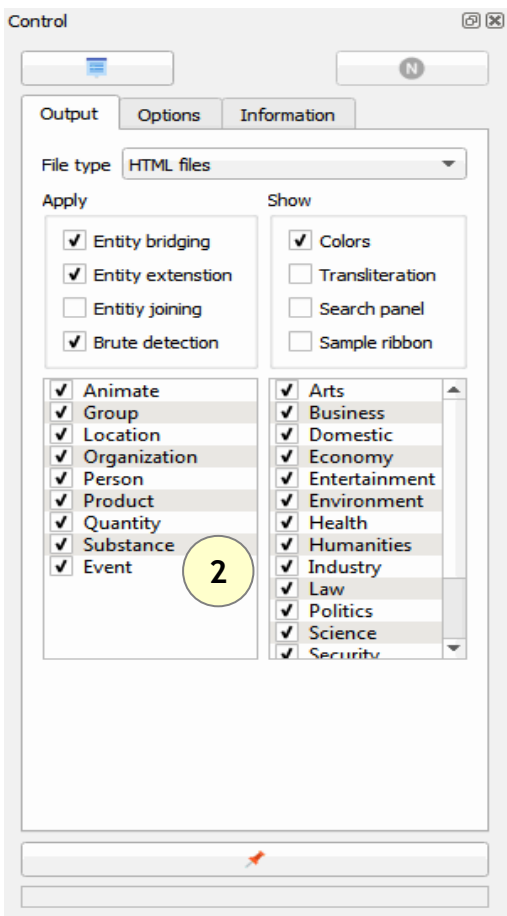
1. Set extraction to [Separate files] and make sure [Content classification] option is checked
2. Select output file format [File type], HTML will support interactive view of contents
3. Load source files using (📁) from the Input Pane
4. Start extraction by clicking (N)
5. The result files will be output to the same folder with slightly altered names



2. Extracting selected entities to file

To extract some specific entity names e.g. personal names, locations etc. from a group of files.

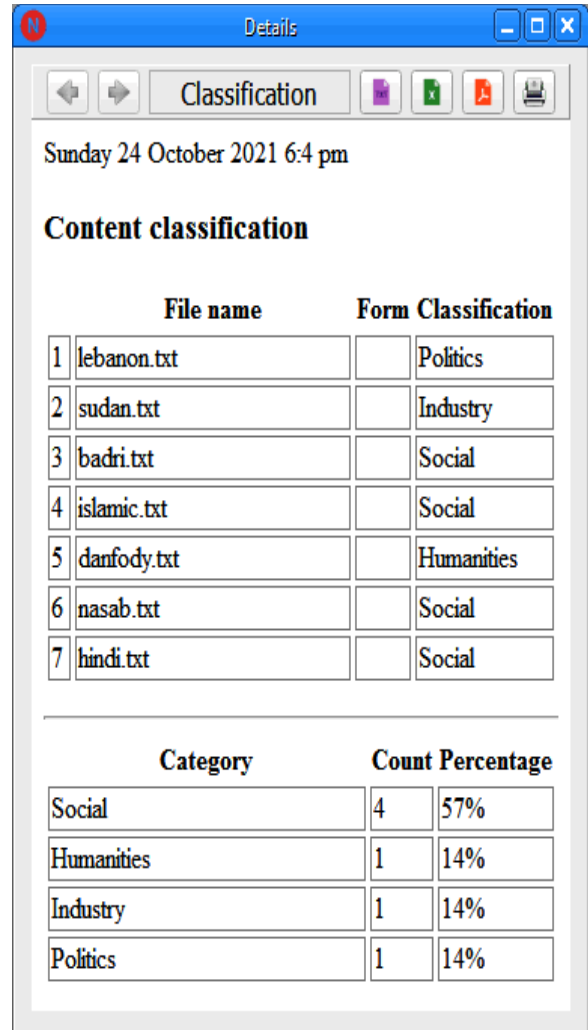
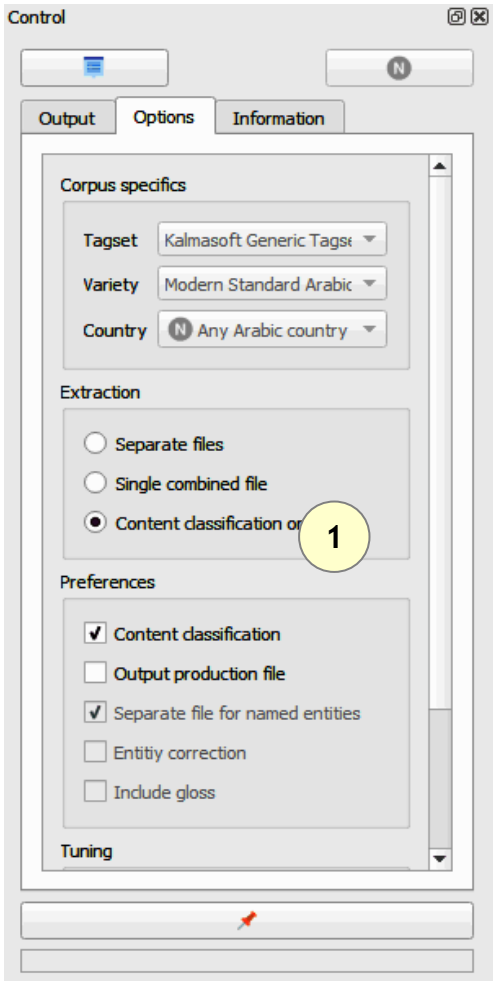
1. Set extraction to [Single combined file]
2. Select entities to be extracted and output file format [File type]
3. Load source files using (📁) from the Input Pane
4. Start extraction by clicking (N)



3. Content classification

To classify text “content categorization” follow the following steps.

1. Set extraction to [Content classification only]
2. Load source files using (📁) from the Input Pane
3. Start extraction by clicking (N)



4. Listing files of specific classification

To get a list of files with text classified to specific domain e.g. health, politics etc. from a group of files.

1. Make sure [Content classification] option is checked
2. Select domains to be classified
3. Load source files using (📁) from the Input Pane
4. Start classification by clicking (N)

