



# **User Manual**

**Arabic Named Entity Extraction**

**NERSys**



Version 1.00

## Table of Contents

<b>Coverage of this document.....</b>	<b>3</b>
Copyright information.....	3
<b>Installation instructions.....</b>	<b>4</b>
<b>Uninstalling the system.....</b>	<b>5</b>
<b>System interface.....</b>	<b>6</b>
Control Pane.....	7
<b>Input Pane.....</b>	<b>8</b>
Entering text.....	8
Single phrase input.....	9
Loading text.....	10
Loading single file from the Input Pane.....	11
Working with the samples.....	13
Entering text samples.....	17
Loading files from the sample ribbon.....	17
Saving user samples.....	18
Saving text to file.....	18
<b>System output.....</b>	<b>19</b>
Sorting and filtering.....	20
Output customization.....	21
Effect of selected features.....	22
Setting the input and output transliteration systems.....	23
Output file format and encoding.....	24
Supported output file formats.....	24
Supported input file encoding.....	24
<b>Working with batch files.....</b>	<b>25</b>
<b>Customizing the system.....</b>	<b>26</b>
Preparing the text for input.....	26
Removing text samples.....	27
Removing text files from the list.....	28
System look and feel.....	29
<b>Text classification.....</b>	<b>30</b>
<b>Typical use cases.....</b>	<b>31</b>
Content analysis.....	31
Extracting selected entities to file.....	32
Content classification.....	33
Listing files classified to specific domain.....	34
<b>Registration.....</b>	<b>35</b>
<b>Frequently asked questions.....</b>	<b>36</b>
Input Interface issues.....	36
Setting issues.....	36
Display issues.....	36
Filtering and sorting issues.....	37
<b>Tips and tricks.....</b>	<b>37</b>

## Coverage of this document

This document is the full text user manual for MAPSSeman© Lite (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.

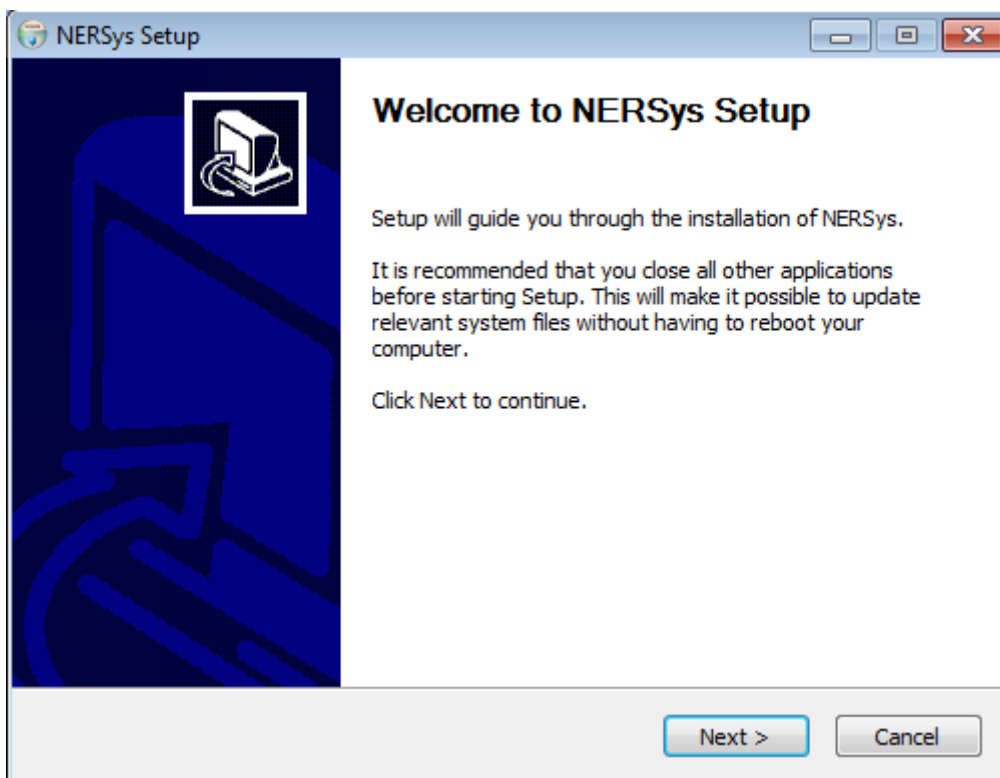
## Copyright information

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.

## Installation instructions

MAPSSeman© Lite (Arabic Named Entity Extractor) comes packaged in an installable file, extract the archive , double click the file (NERSys100.exe) and follow the on screen instructions.

The system does not add any entries to the Start menu “Start” or entries to the system registry, so it can be easily removed by deleting its folder only, the comes with an uninstaller tool for this purpose described in the section Uninstalling the system on page (5) how to remove the system from your system.

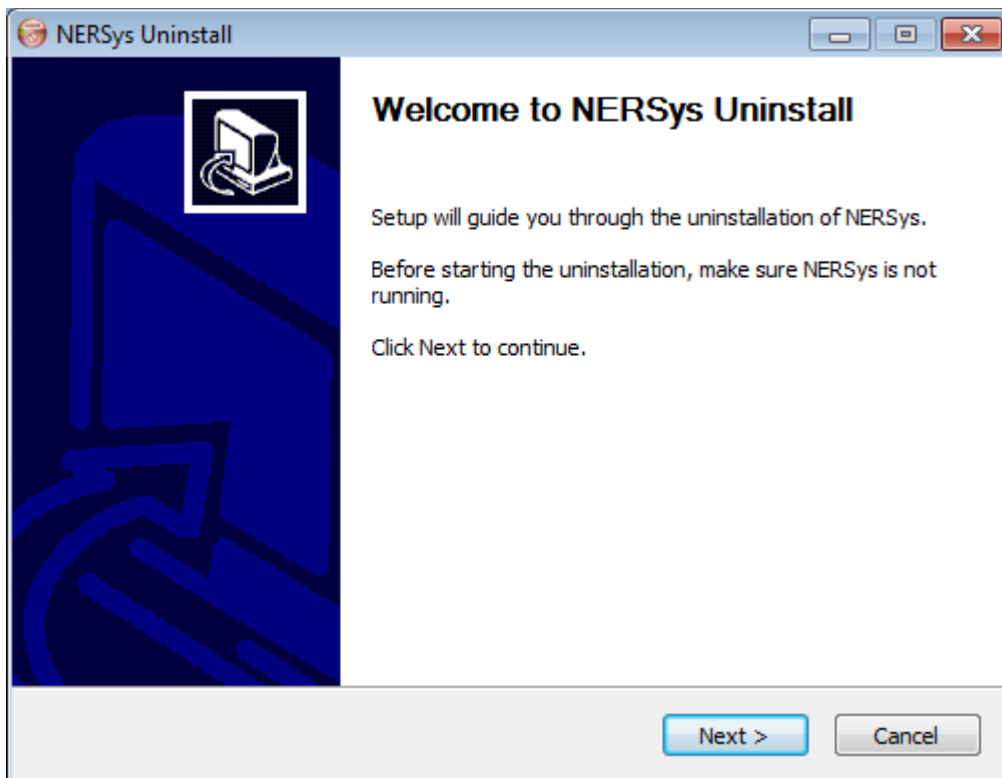


After installation, click (OK), Extractor system is now ready to use.

## Uninstalling the system

To remove the system, go to the system folder and run the uninstalling utility (MNERSysUninstall.exe) by double-clicking, you can do so through the following steps:

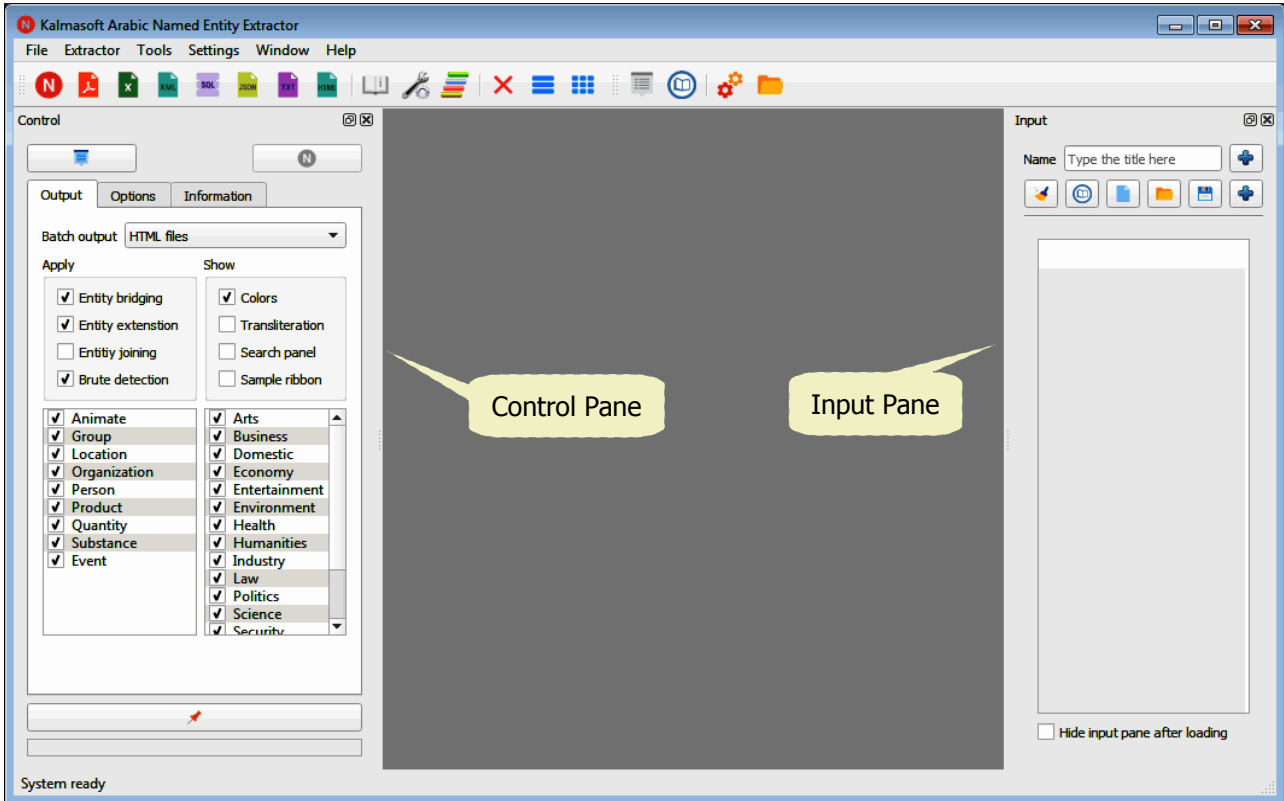
- Right click on the system icon on the desktop to display the properties popup menu and select (Properties) at the bottom of the list.
- Click on the button (Find Target ...) to go to the system folder, the contents of the folder may be similar to what is shown in
- Run the uninstaller system by double-clicking on the icon (Uninstall.exe).



You will need to manually remove the system's icon on the desktop.

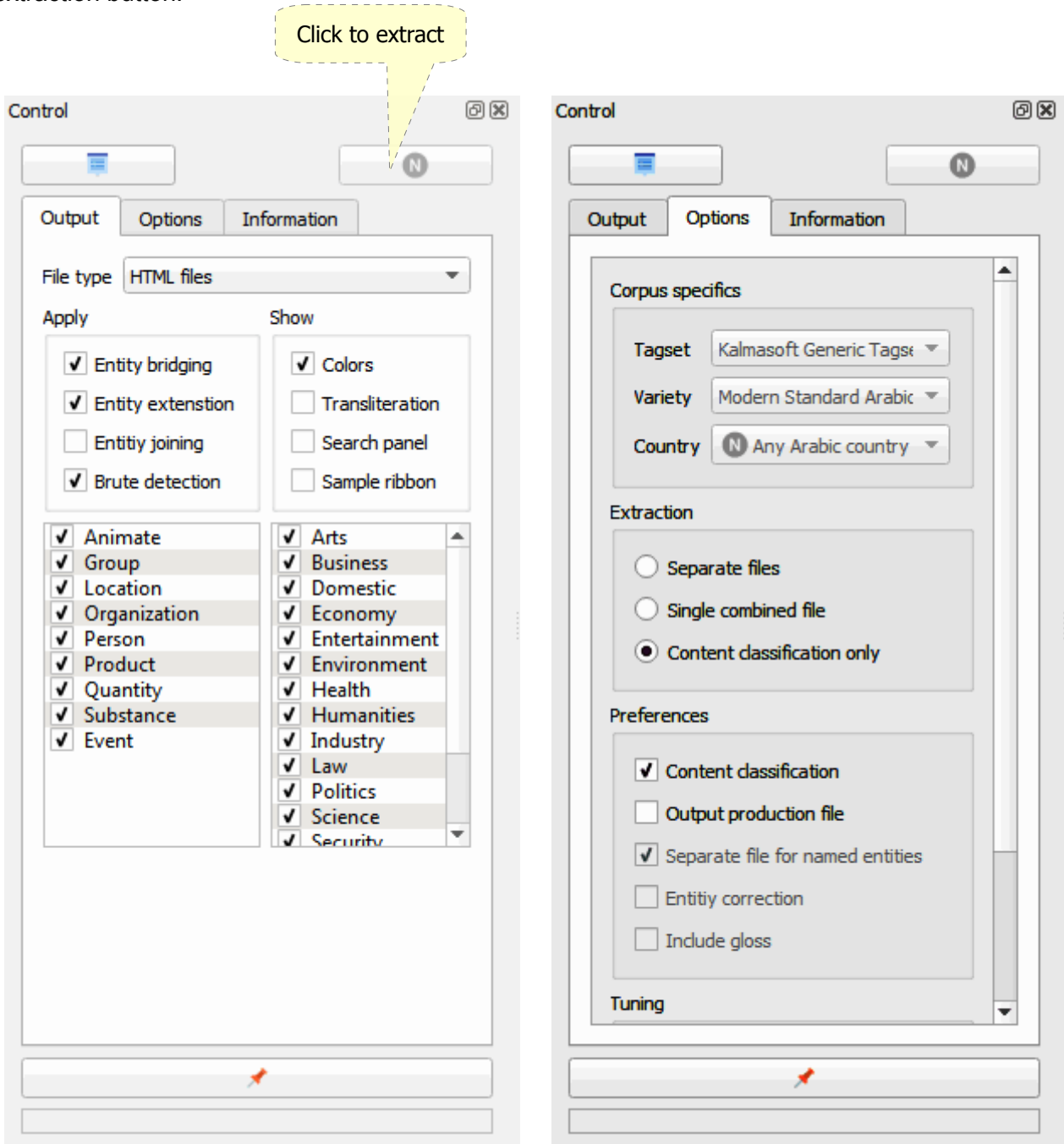
## System interface


Arabic Named Entity Extractor has main window interface with subwindows and panes, Input Pane to enter Arabic text, Control Pane to control the system behavior.



## Control Pane

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

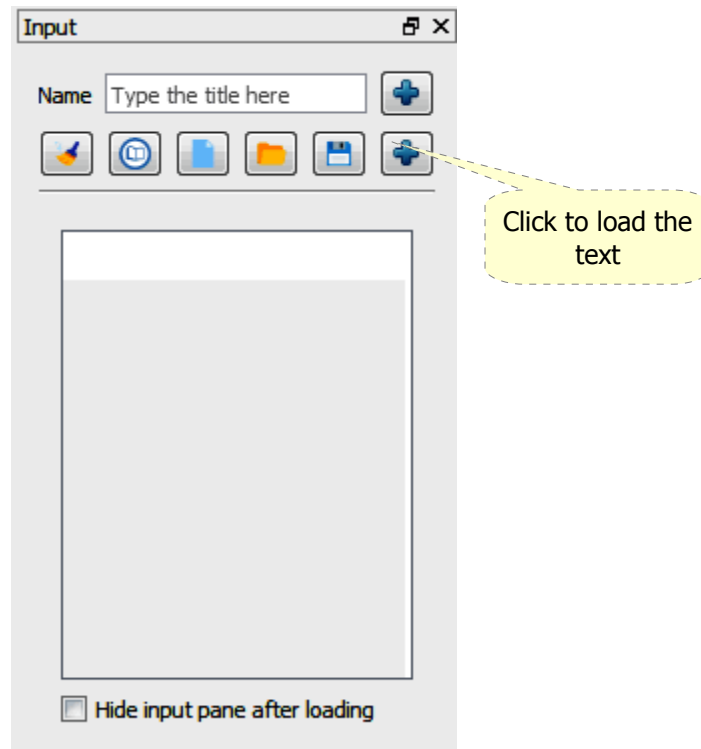



Click the extraction button (  ) to start the extraction process on the Arabic text, check the other controls as appropriate.

## Input Pane

### *Entering text*

Use the Input Pane for an average of 1000 words; text can be input in several ways either by directly using the keyboard, pasting, or dragging and dropping contents from another application window.

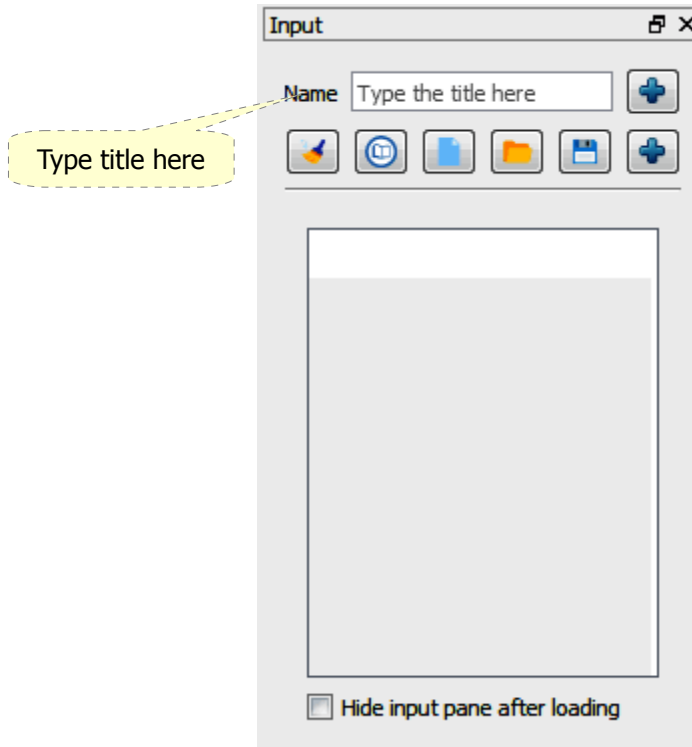


To clear all the text click on the (  ) button at the bottom, text can be separated from each other in several ways.



## 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 in the following way:

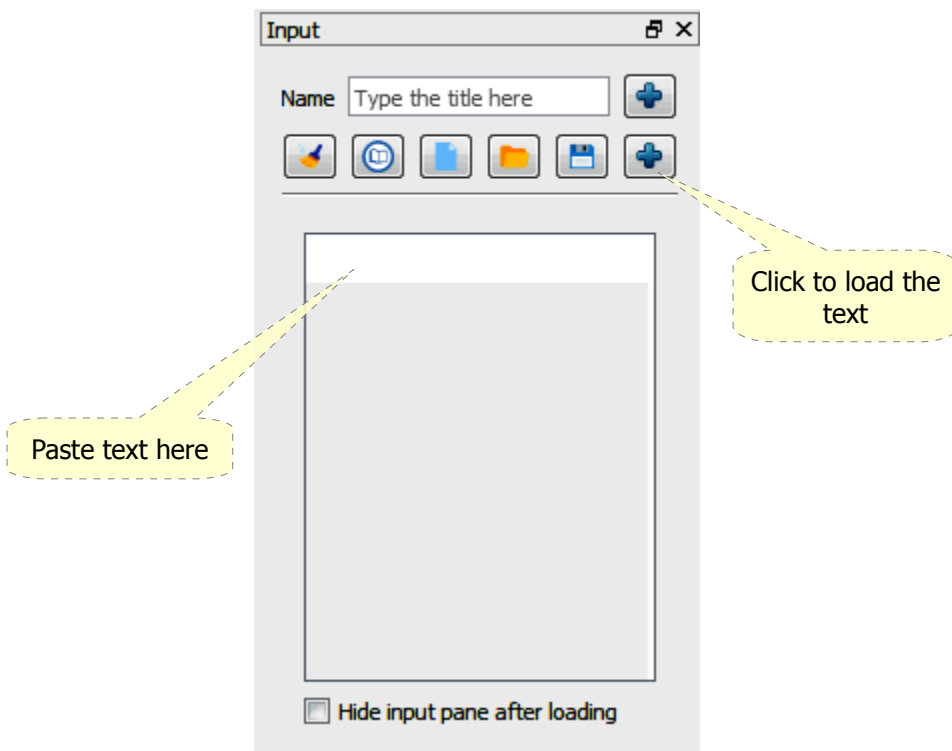


- Paste a text then click the (+) to load, a card will show up in the main area.
- Click the the [Extract] button on top to start extraction, you may enter as many texts using cards as needed this manual way.




### Loading text

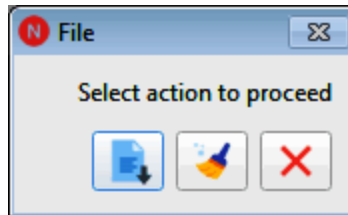
Direct paste of text is also available, click on the (+) button to load the text when done:




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.

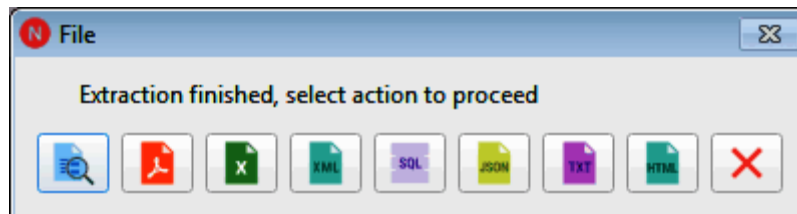
### ***Loading single file from the Input Pane***

You can load a file of Arabic text using the same Input Pane above, 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 any time you decide to use it.



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

After extraction done you can then output the results to any of the file formats shown below for later use, or just view the results for checking.



Extractor Viewer

چ ت ا ب ا ن ا و ا ء  
ص ش س ز ر د د خ ح  
م ل ك ق ف غ ط ض  
د ر ؤ خ ج د ب  
ز د ث ب ن ش ك ك ف  
ه ث ر ع

All

All

Clear

Output production file

Person - 0 Place - 0 Organizati Group - 1 Event - 0

مقاتلو طالبان  
Tag: E1GRP  
Type: GRP  
Category: GENR  
Subcategory:

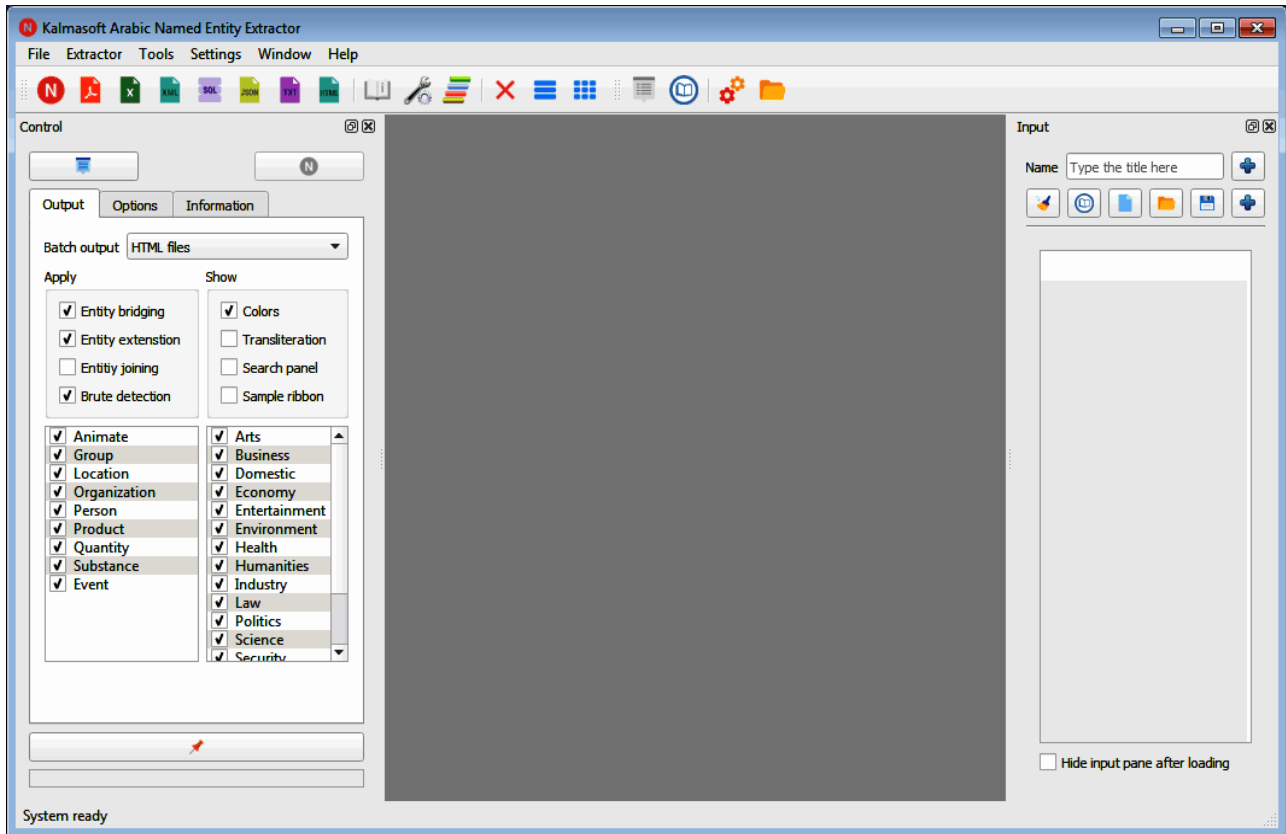
Expand Expand Expand Collapse Expand

17 - مي طراز من الصنع لطائرة هليكوبتر روسية صوراً بحماس مقاتلو طالبان يلتقط


مقاتلو طالبان GRP بحماس صوراً PRO لطائرة هليكوبتر روسية الصنع من طراز  
مي - 17 PER وهي تهبط في FAC مطار بلخ في FAC مزار شريف في  
شمال أفغانستان LOC . ويوجد على متن PRO الطائرة مسؤولون كبار من  
حركة طالبان GRP فيما يجلس أعداؤهم ORD السابقون في قمرة القيادة ، وهم طيارون  
من GRP القوات الجوية الأفغانية ORD السابقة . واصطحبني مولوي  
عبدالله منصور PER ، قائد عناصر GRP طالبان المسؤول عن FAC المطار ، في  
جولة VNT أطلعني فيها على الأسطول الذي بات يسيطر عليه الآن ، ويشمل  
PRO مروحيات VNT هجومية وطائرات مقاتلة أهدتها GRP القوات الدولية  
ORG للحكومة السابقة . وفي ظل الحكومة القديمة ، كانت PRO الطائرات تُستخدم غالباً  
VNT لاستهداف حركة طالبان ، وليس من الواضح كيف سيتم استخدامها الآن بعد  
VNT توقف الحرب . ويقول PER منصور : " إذا احتجنا إلى تلك PRO الطائرات في  
المستقبل PER فهي هنا " . كان QTY عشرات الطيارين قد فروا من LOC البلاد خوفاً

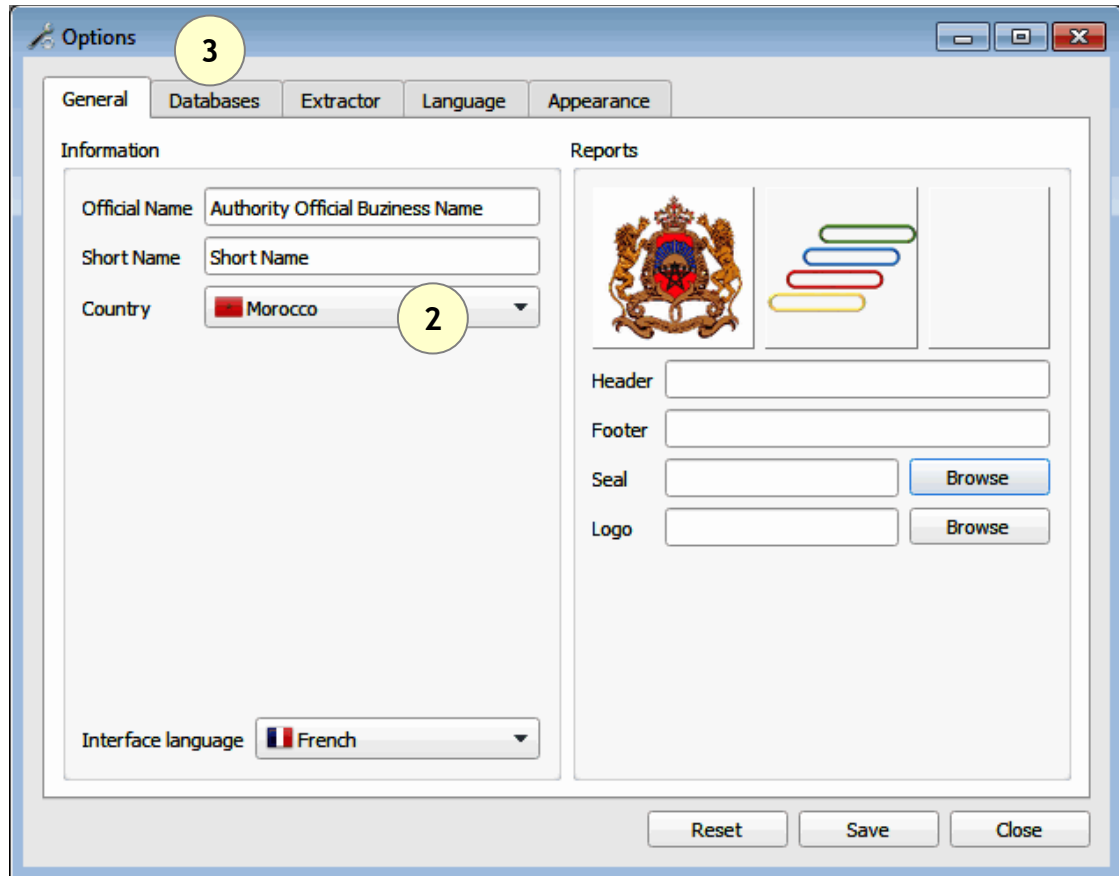
## Working with the samples

Arabic Named Entity Extractor comes ready with few arbitrary text samples, you need to prepare the system before testing with those samples.

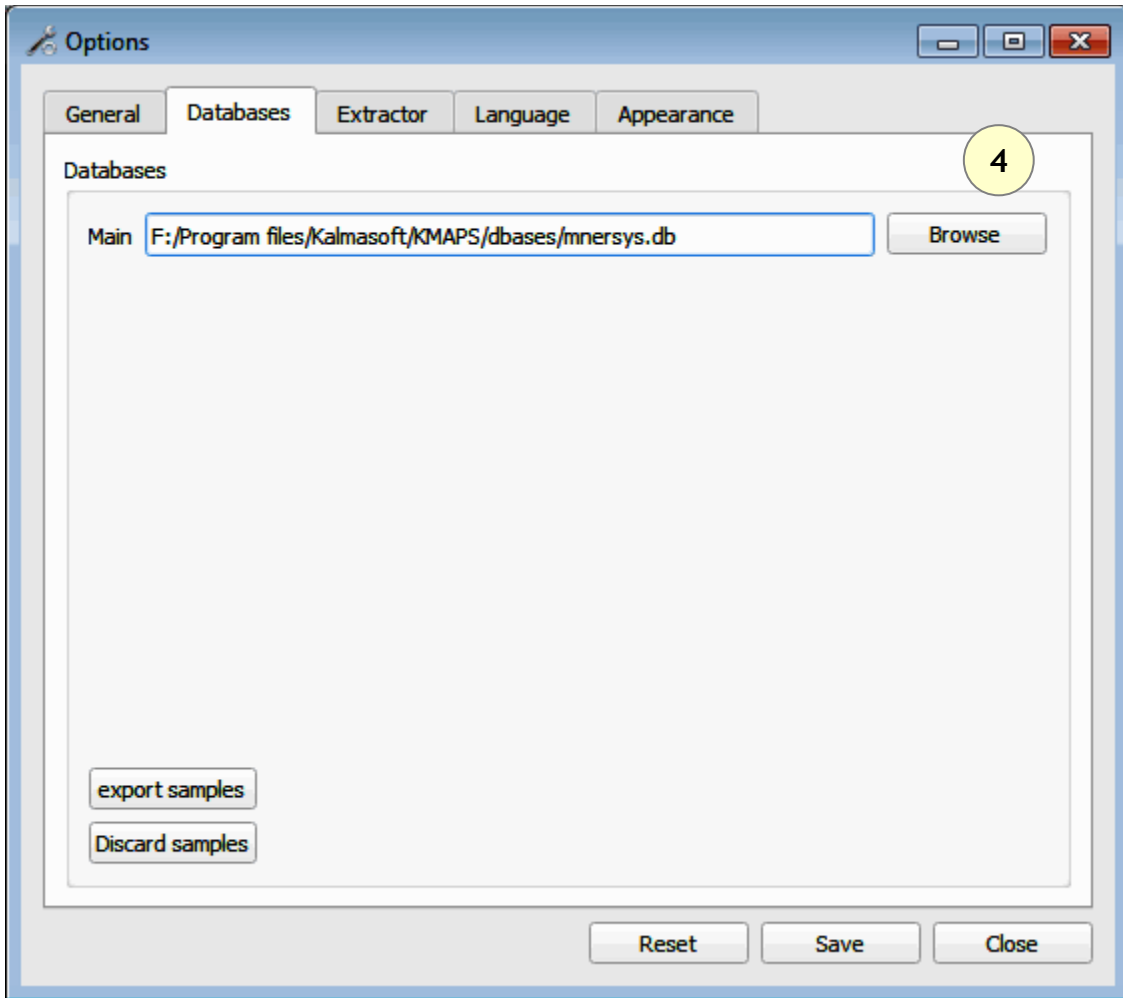


1

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

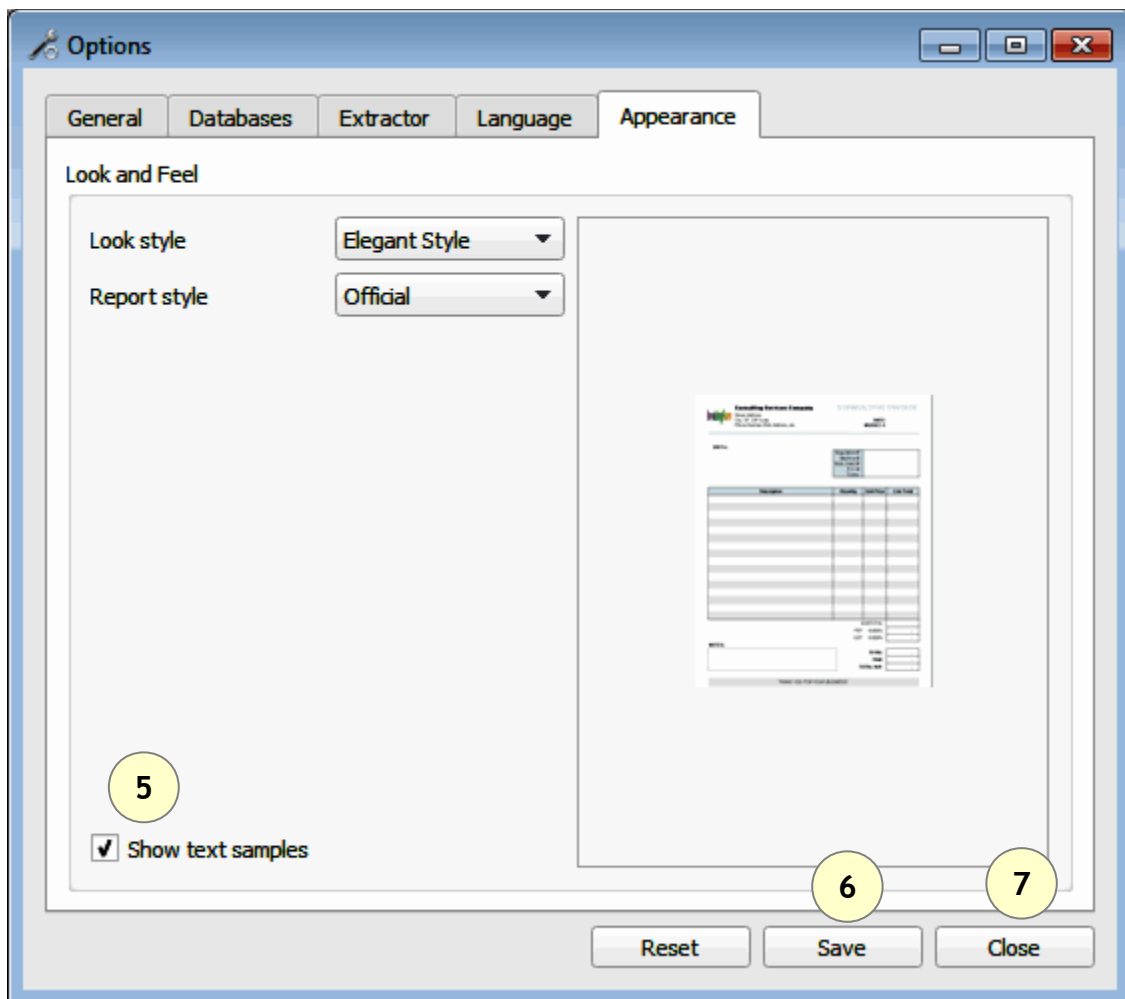


- Set country and language according to your preferences.
- Switch to the Database tab (3).
- make necessary changes as shown above.



- 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.

Switch to the (Appearance) tab to activate the samples.

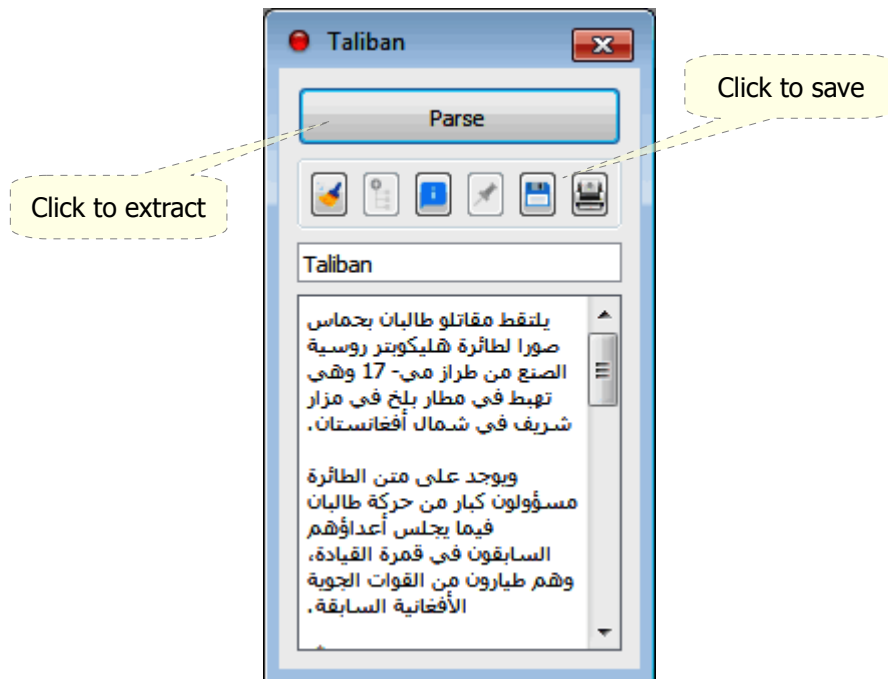
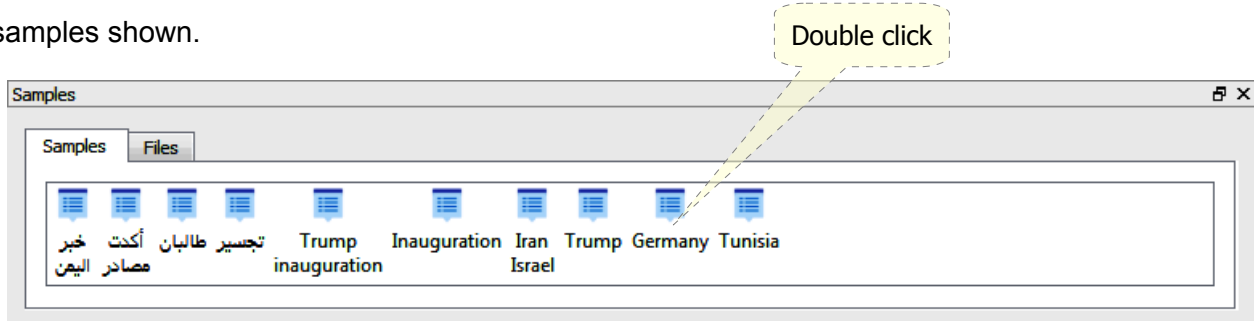


- Check the [Show text samples]
- Click (Save) to save your new settings then click (Close)
- Restart the system, next time you run the system text samples will appear at the bottom of the main window.



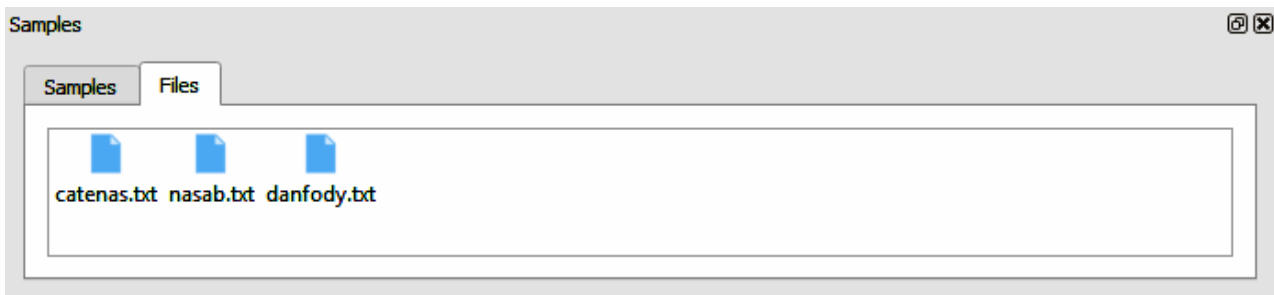
## Entering text samples

Entering text samples is straightforward, double click on the sample icon in the ribbon, a text card shows up, click the the button to start extraction, you may repeat the same procedure for all samples shown.




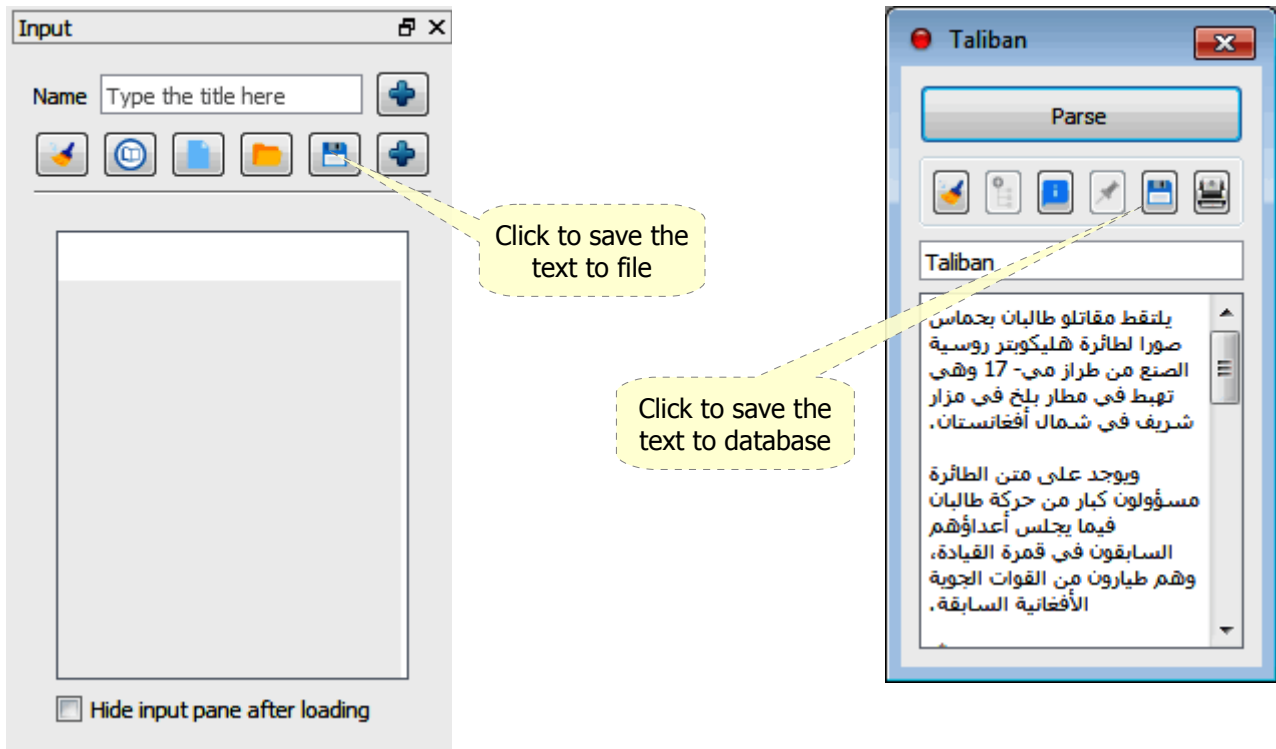
## Loading files from the sample ribbon

Same way as single phrase samples, double click on the file name to load all text.




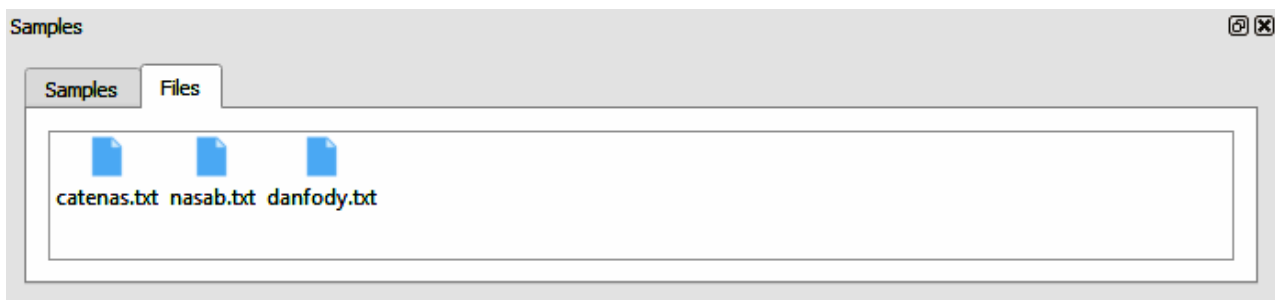
## Saving user samples

Now you have the system set for working with samples, once loaded, the single phrase entered using the Input Pane 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.



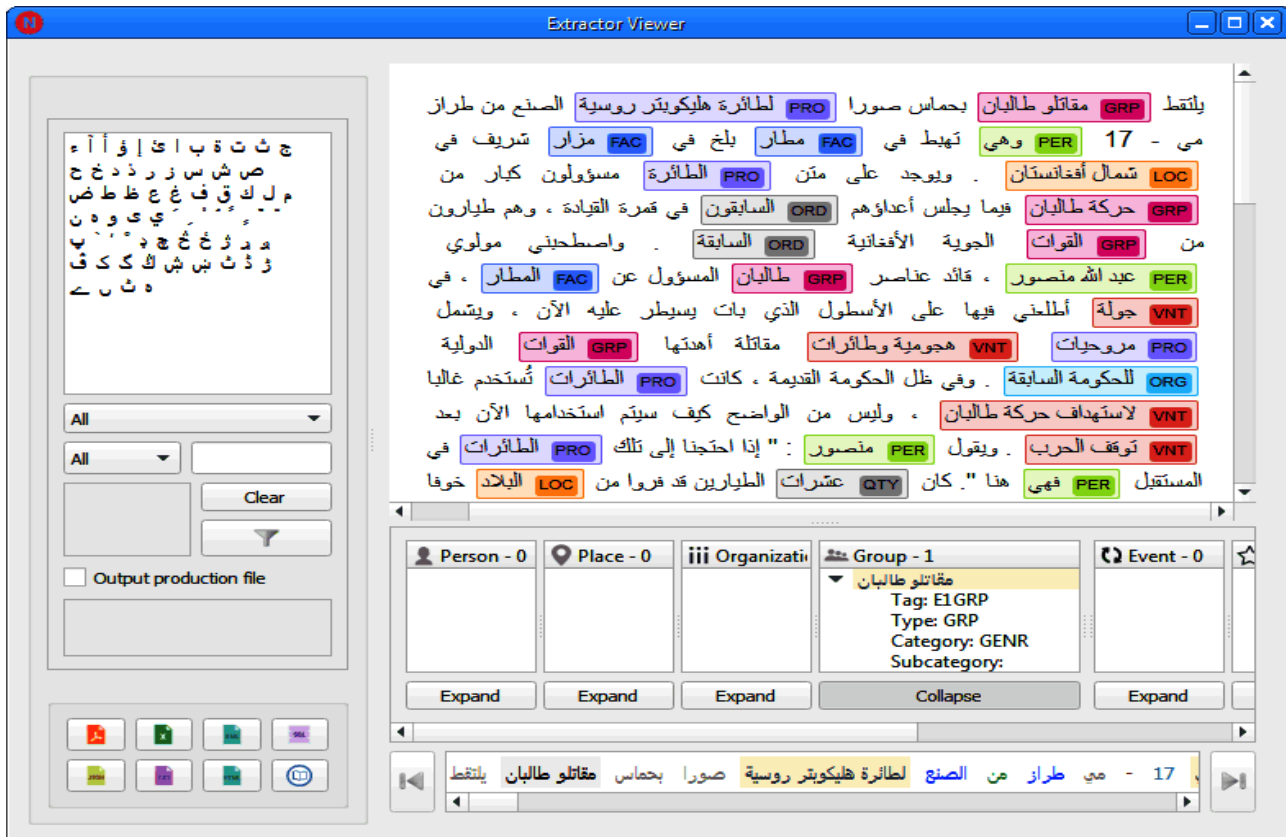
## Saving text to file

Once loaded, your text can be saved the same way to a file for later use; click on the (  ) button to save, the file will be saved to your file system with a link added to the sample ribbon.



## System output


The Output Interface shows all information about the extraction process done, shown below, is the default setting, the system behaviour can be changed as illustrated in paragraph Customizing the system on page (26).






If the extraction process is successfully done on the text, the above screen will show with different labels.

The system does not fully suggest error correction or allow doing correction in place directly from the interface, retype or reload the correct text and run the extraction process again to get rid of the errors messages.

## Sorting and filtering

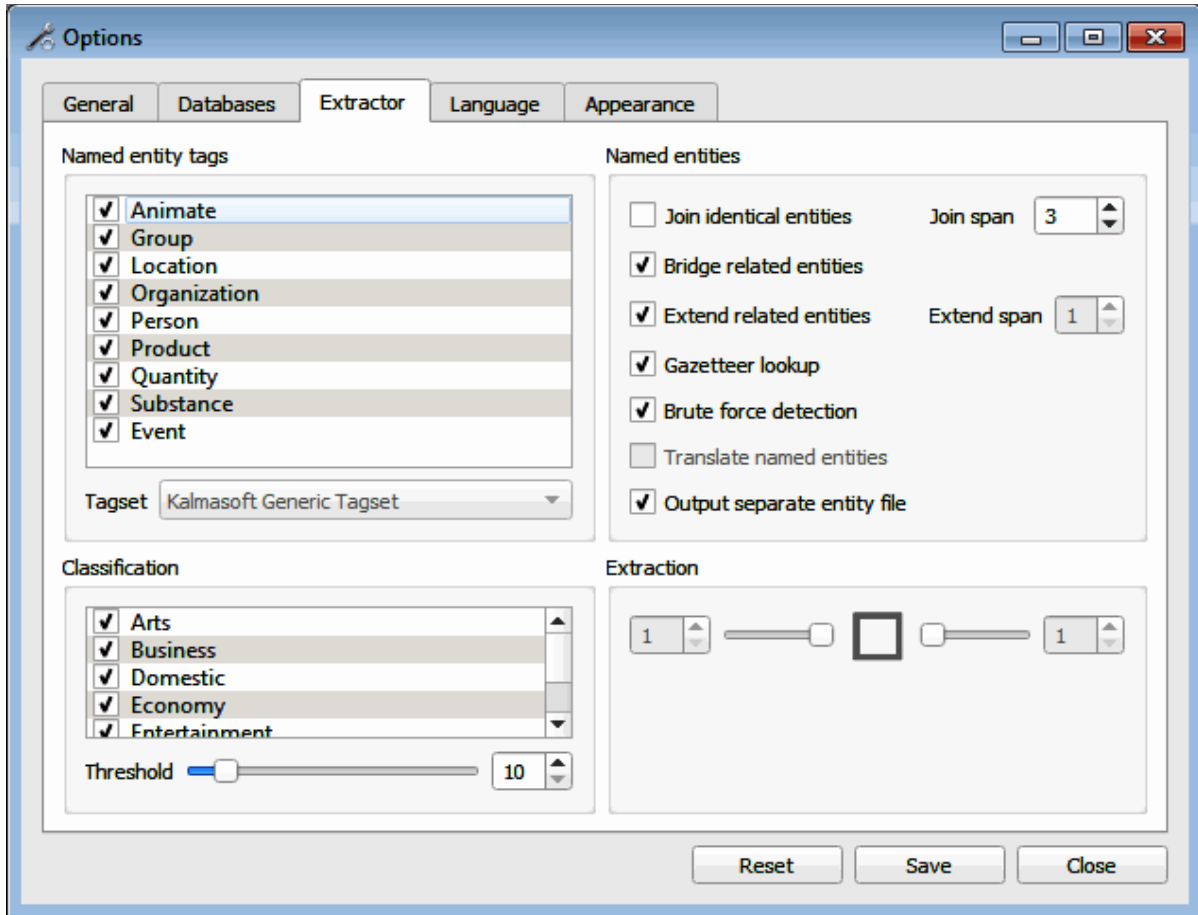
To filter the extraction results click on the filter (  ) button, you must specify the filter criteria to display only the matching records, available criteria are listed in the table below.

Criteria	Details
Contains	To display the words contain a specific text or symbols
Starts with	To display the words begin with a specific text or symbol
Ends with	To display the words end with a specific text or symbol

A larger text will be displayed in the preview area, when you are done setting the filtering criteria, click on (  ) button to start filtering the results, click the (  ) to clear the filter then click (  ) again restore the original view.

## Output customization

The Setting Interface is where Arabic Named Entity Extractor overall behavior can be controlled through this interface, 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 <b>VNT</b> تظاهرة صفاقرس <b>PPL</b> عاصمة للثقافة العربية	في إطار <b>VNT</b> تظاهرة صفاقرس <b>PPL</b> عاصمة للثقافة العربية
وأسند <b>ORG</b> النادي الثقافي العربي	وأسند <b>ORG</b> النادي الثقافي العربي

### Joining

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

Before joining	After joining
يلتقط <b>GRP</b> مقاتلو <b>GRP</b> طالبان <b>GRP</b> بحماس صورا	يلتقط <b>GRP</b> مقاتلو طالبان <b>GRP</b> بحماس صورا
في <b>LOC</b> شمال <b>LOC</b> أفغانستان	في <b>LOC</b> شمال أفغانستان
رئيسا <b>PER</b> ترامب <b>VNT</b> حفل تنصيب	رئيسا <b>VNT</b> حفل تنصيب ترامب

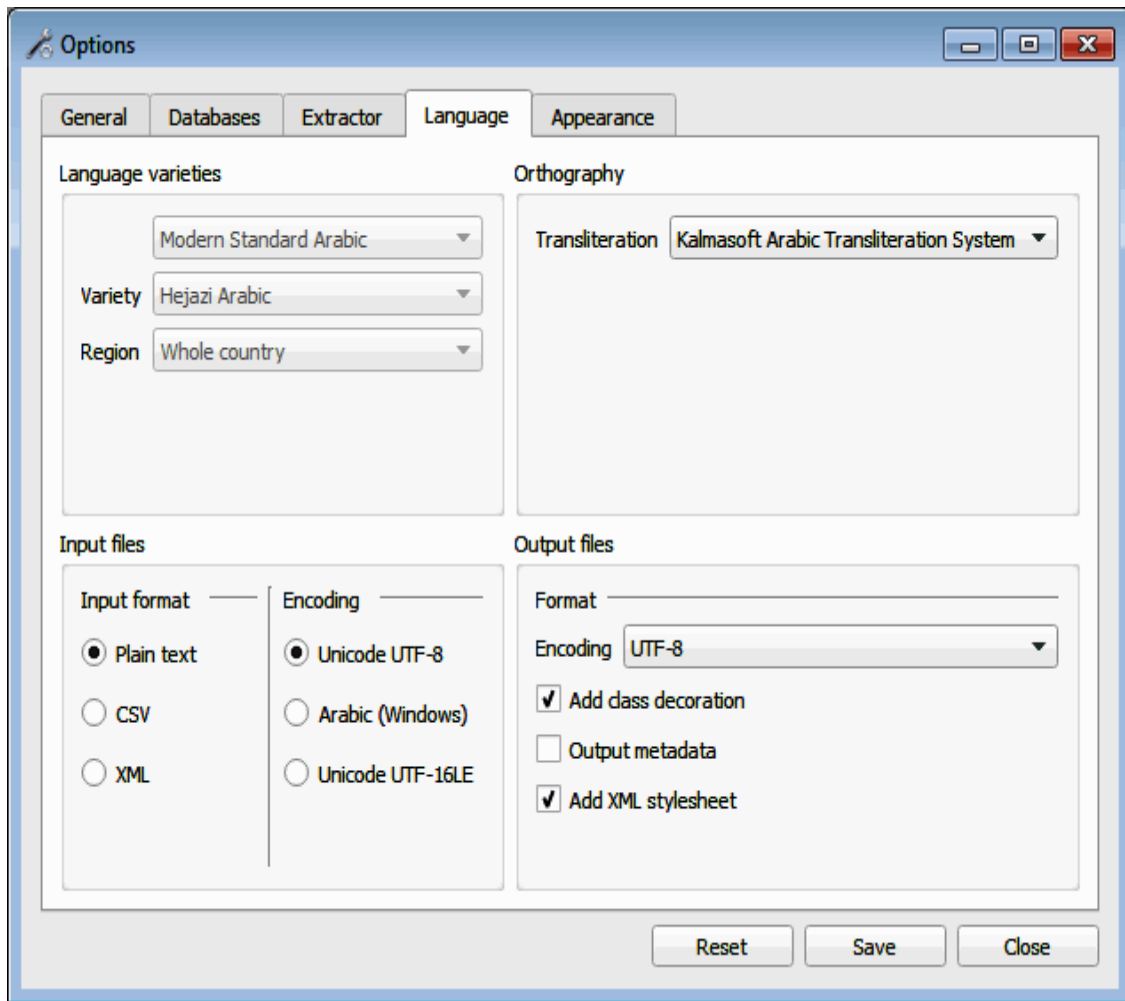
### Bridging

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

Before bridging	After bridging
سمعنا يرجوع خالنا <b>PER</b> ابن <b>PER</b> علي <b>PER</b> محمد	سمعنا يرجوع خالنا <b>PER</b> ابن علي محمد
في <b>LOC</b> لبنان <b>VNT</b> الحرب الأهلية	<b>VNT</b> الحرب الأهلية في لبنان
من <b>DAT</b> جانفي <b>DAT</b> يوم التاسع والعشرين	<b>DAT</b> يوم التاسع والعشرين من جانفي

## Setting the input and output transliteration systems

The system supports some 4 NER tagging systems, both standard and for academic use.



[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

## **Output file format and encoding**

Output file encoding is set to default but you may need to customize another setting for saving the results of extraction, The system supports Unicode by default in both input and output so make sure this encoding is used to ensure that all output details are preserved, the remaining options shown on the image below are self-explanatory.

## **Supported output file formats**


Format	Full name	Details
Text	Plain text	Processing purposes
XLSX	Excel sheet	Microsoft Excel
SQL	MySQL dump format	MySQL database
PDF	Adobe Portable Document Format	Display and print purposes
JSON	Java Simple Object Notation	Processing purposes
XML	Extended Markup Language	Processing purposes
HTML	Hypertext Markup Language	Viewing purposes

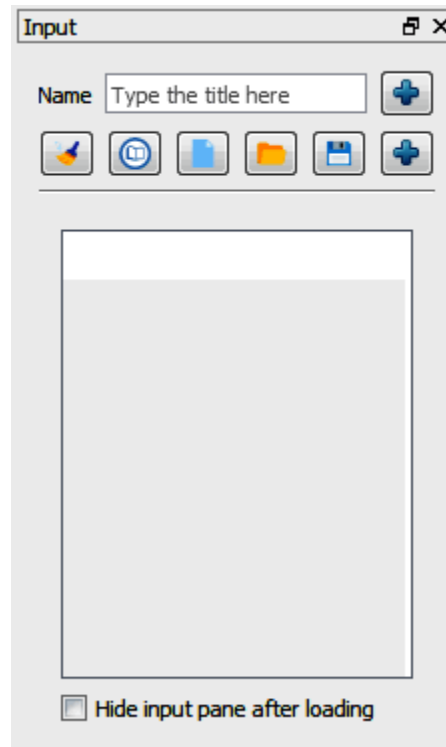
## **Supported input file encoding**

Encoding	Uses	Example of applications
UTF-8	for most multilingual applications	Universal
Arabic (Windows)	use with specific applications	Microsoft applications
UTF-16BE	use with most applications	Apple Mac OS X



## Working with batch files

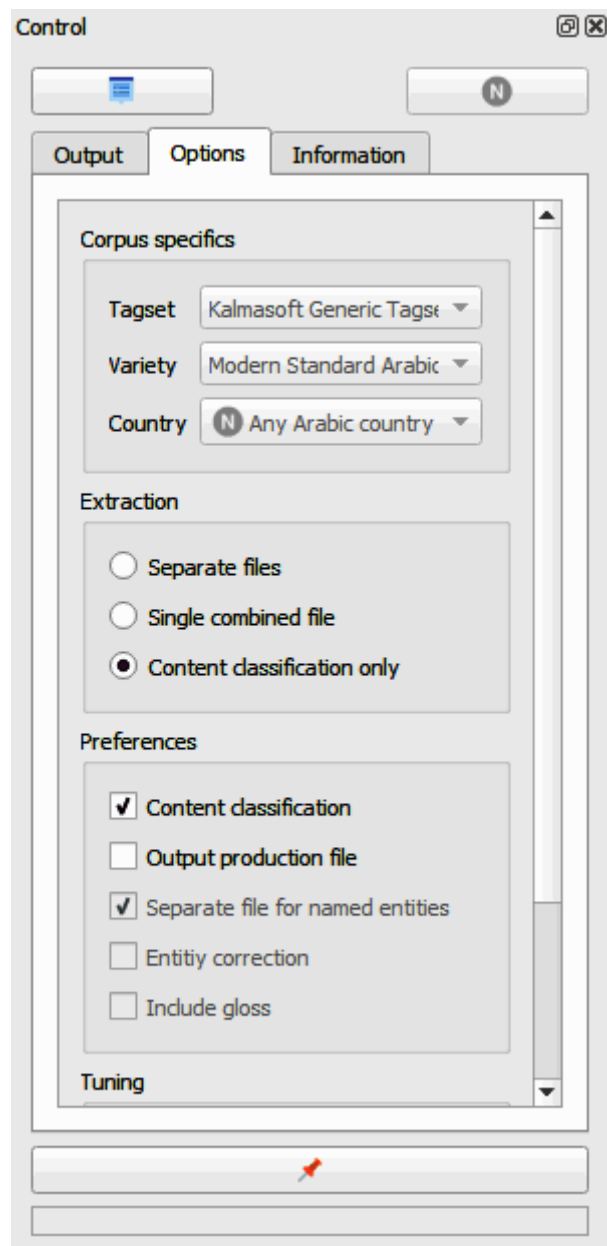
You can load multiple files of Arabic text using the same Input Pane above, the text files must be encoded in UTF8 without any headers. Click (  ) button and browse to the files folder, the files will be added to the waiting list ready for loading at any time you decide to process them.



## Customizing the system

### *Preparing the text for input*

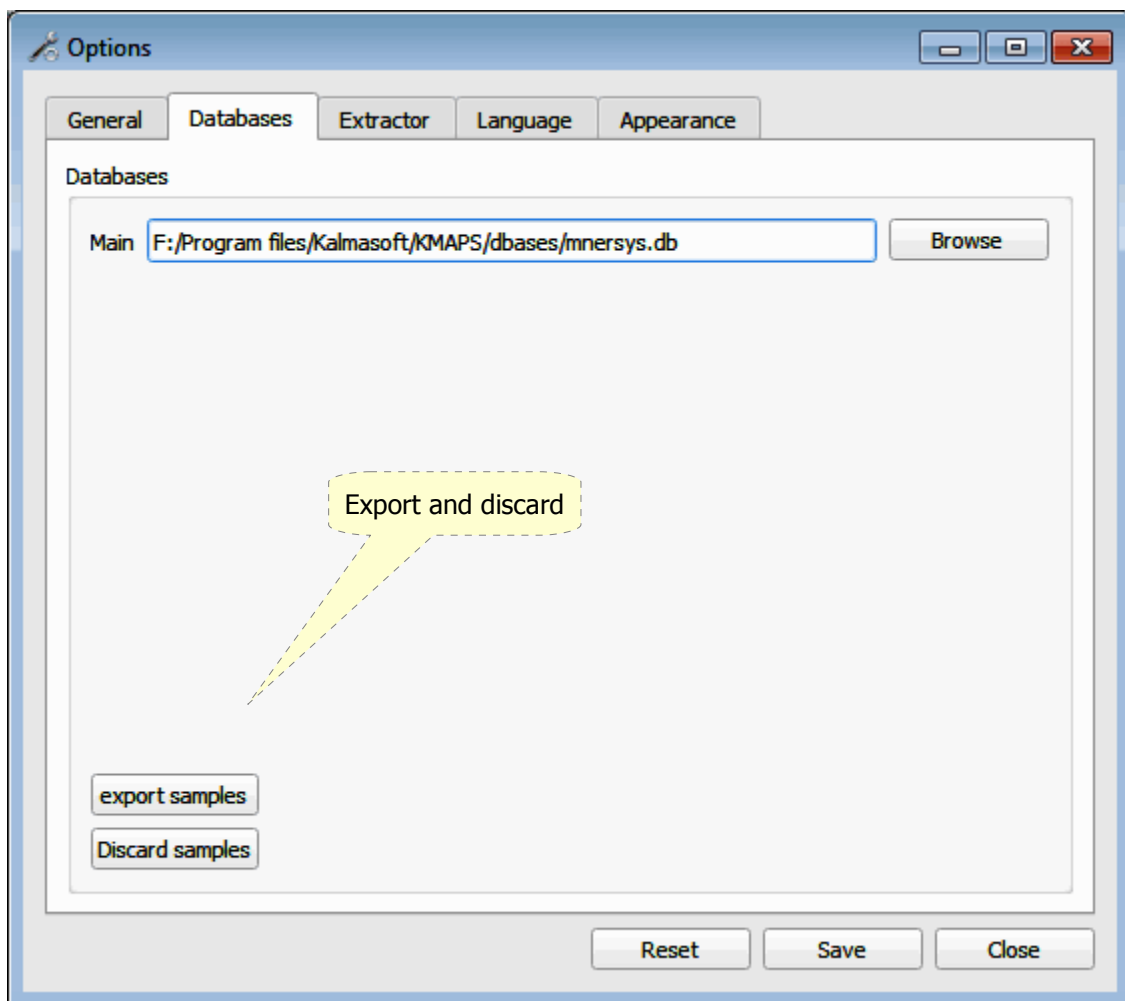
Arabic Named Entity Extractor is very sensitive to the input data, it makes several tests to before the extraction process begins, these tests are of orthographic aspects in general, they include checking for properties such as length of the text, the system also does some other covert arrangements including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




## Removing text samples

Samples in the ribbon may be deleted completely to spare a space for user's own samples.

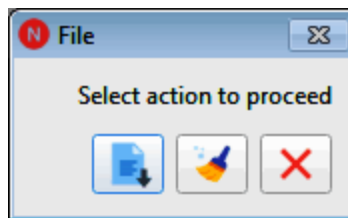
- To remove only few samples, click on text icon then press (Delete) from the keyboard.
- Click the (Discard samples) button, this will remove all samples permanently.
- Click the (Export samples) to keep a copy of the samples before deletion.



## 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.

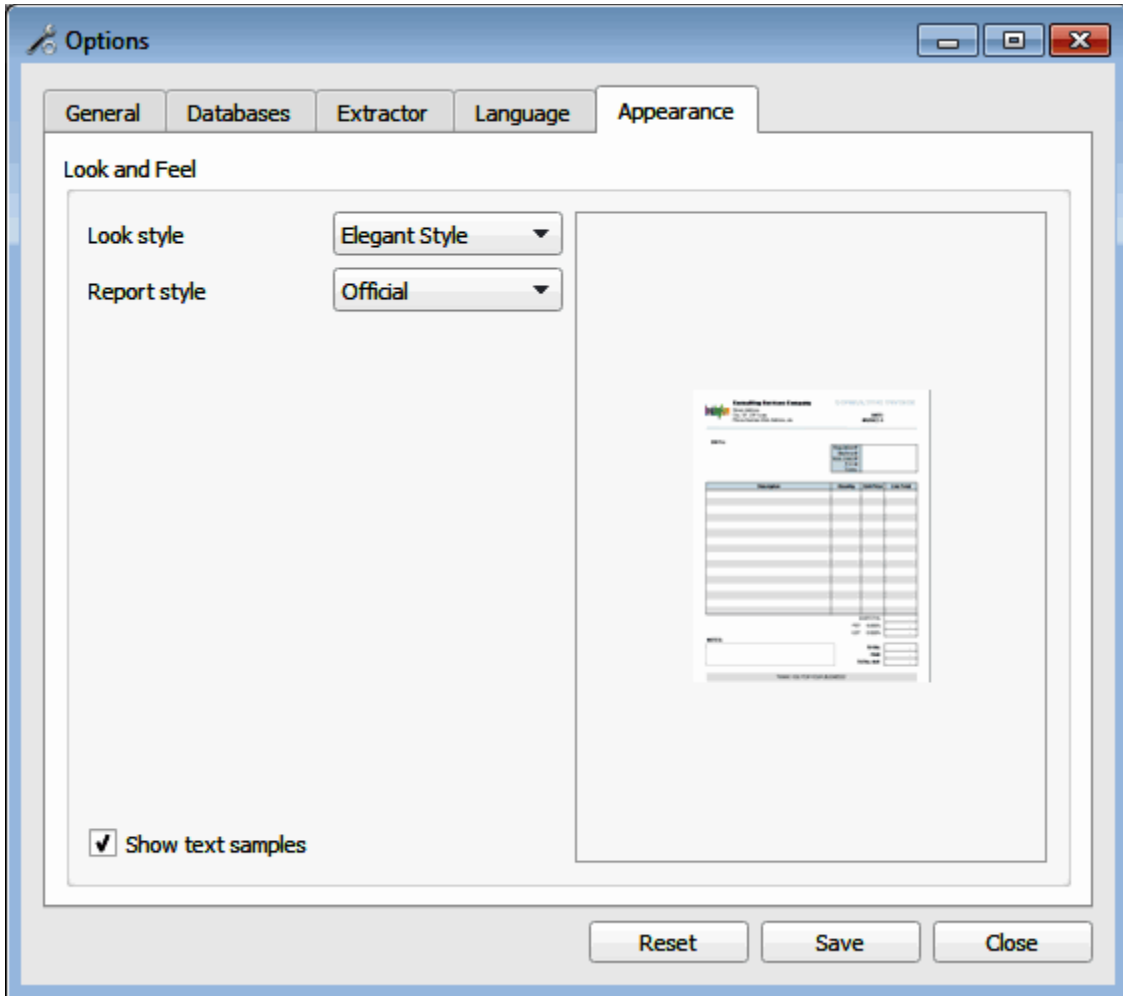
Double click the file



This will only remove the file from the list not deleting the original final.

## System look and feel

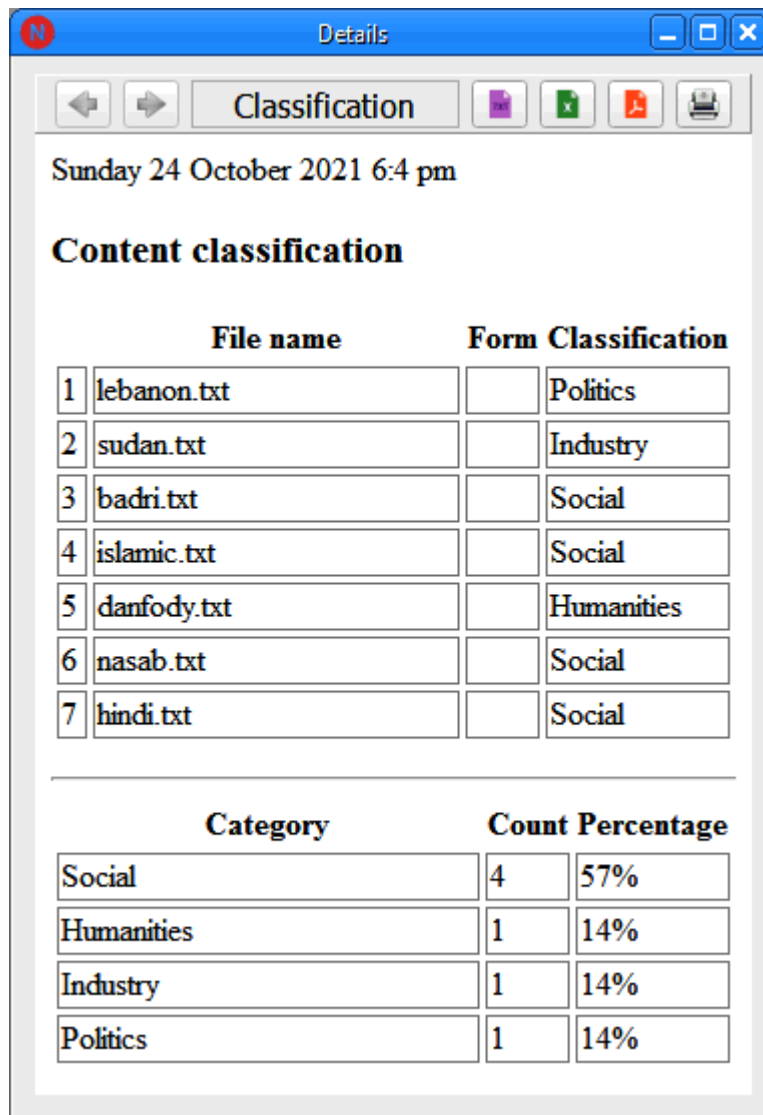
Three interface styles are available, select the style for both the interface and the reports then click (Save), the changes will not take effect until the next run of the system.



## Text classification

You can get a list of semantically classified input files after setting the [Content classification] option from the Control Pane.

Use the [Threshold] option to control the percentage limit of content affecting the classification.



The screenshot shows a window titled 'Details' with a 'Classification' tab. The window displays the date and time 'Sunday 24 October 2021 6:4 pm'. Below this, the section 'Content classification' contains two tables. The first table lists seven files with their corresponding classifications. The second table summarizes the counts and percentages for each classification category.

	File name	Form Classification
1	lebanon.txt	Politics
2	sudan.txt	Industry
3	badri.txt	Social
4	islamic.txt	Social
5	danfody.txt	Humanities
6	nasab.txt	Social
7	hindi.txt	Social

Category	Count	Percentage
Social	4	57%
Humanities	1	14%
Industry	1	14%
Politics	1	14%

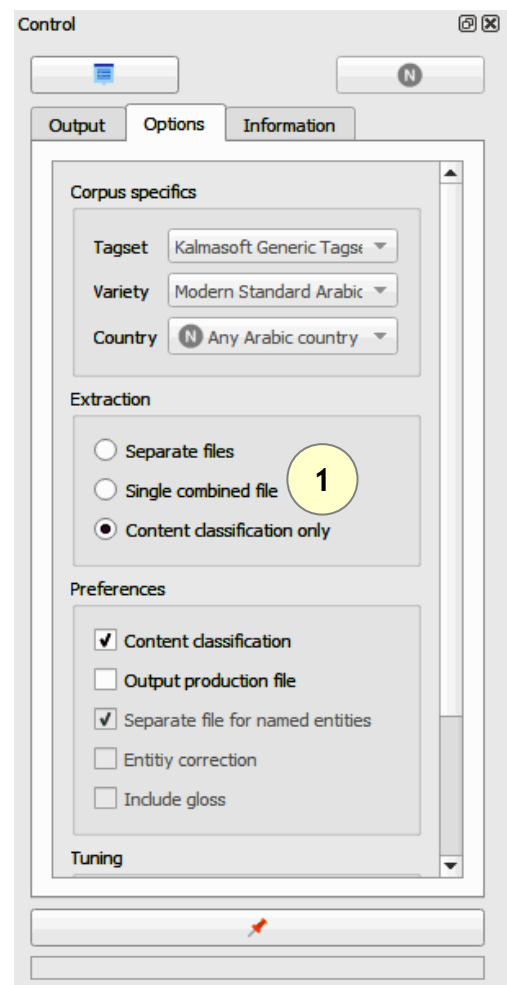
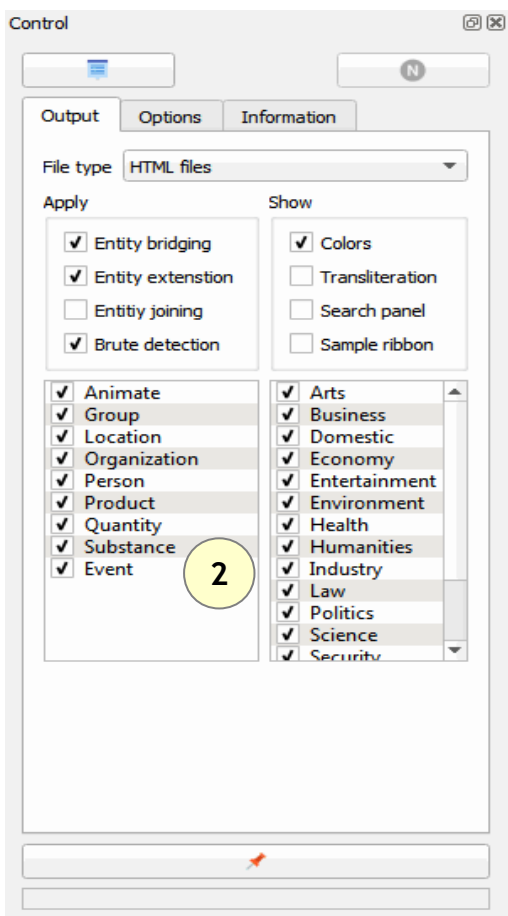
## Typical use cases

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

### Content analysis

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

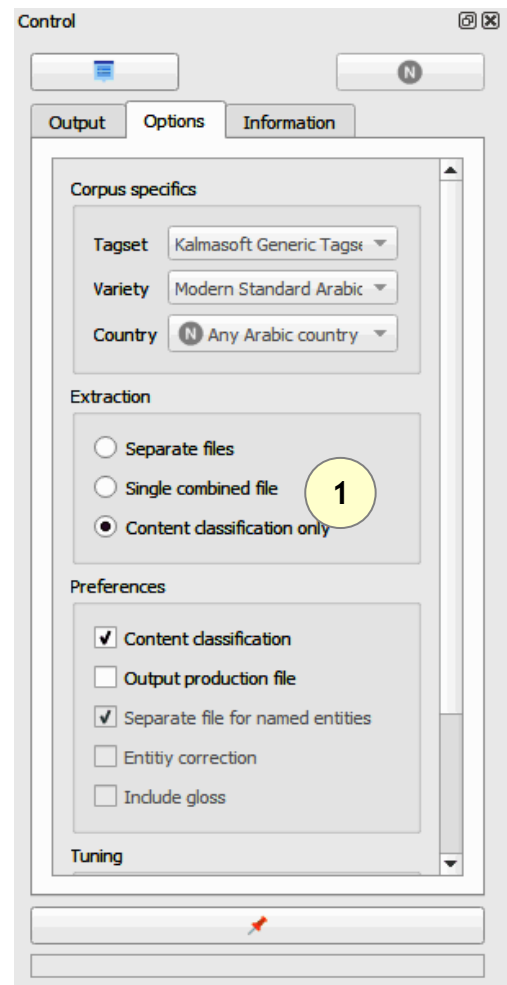
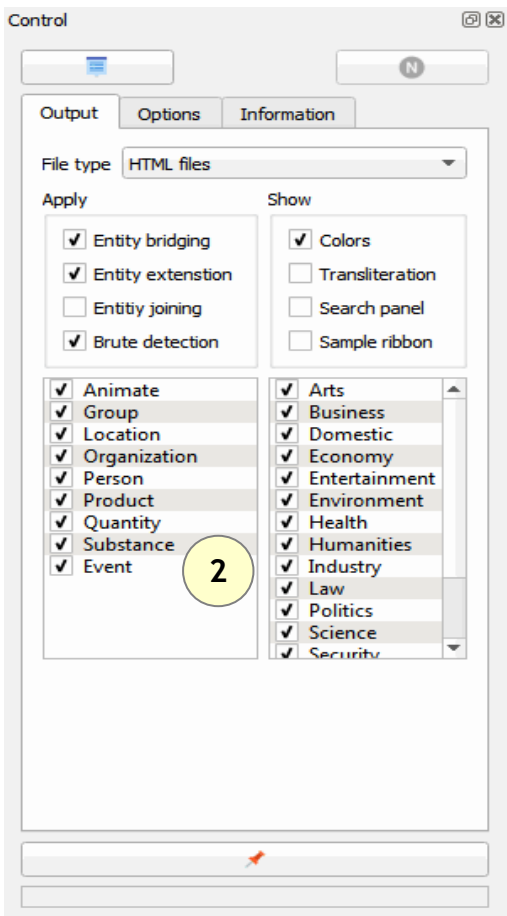
1. Set extraction to [Separate files] and make sure [Content classification] option is checked
2. Select output file format [File type], HTML will support interactive view of contents
3. Load source files using (📁) from the Input Pane
4. Start extraction by clicking (N)
5. The result files will be output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



## 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)

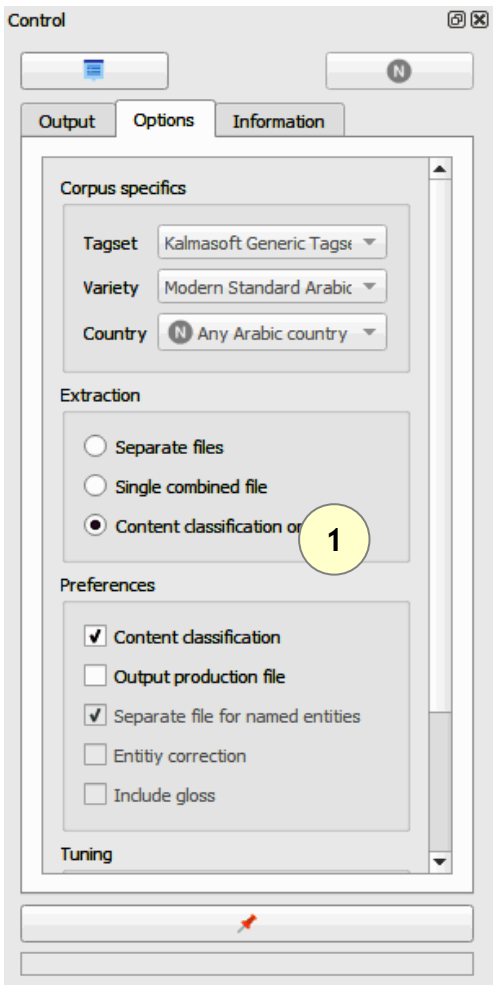




## Content classification

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

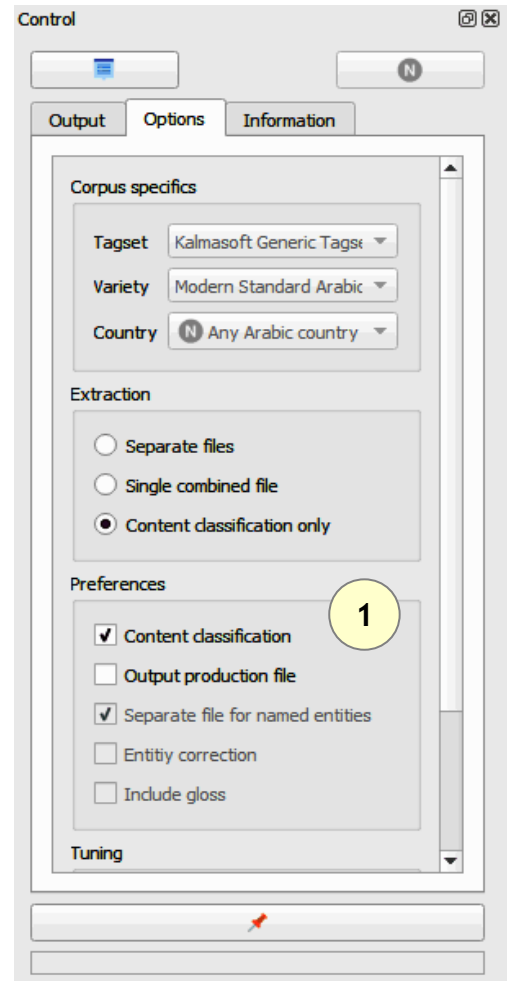
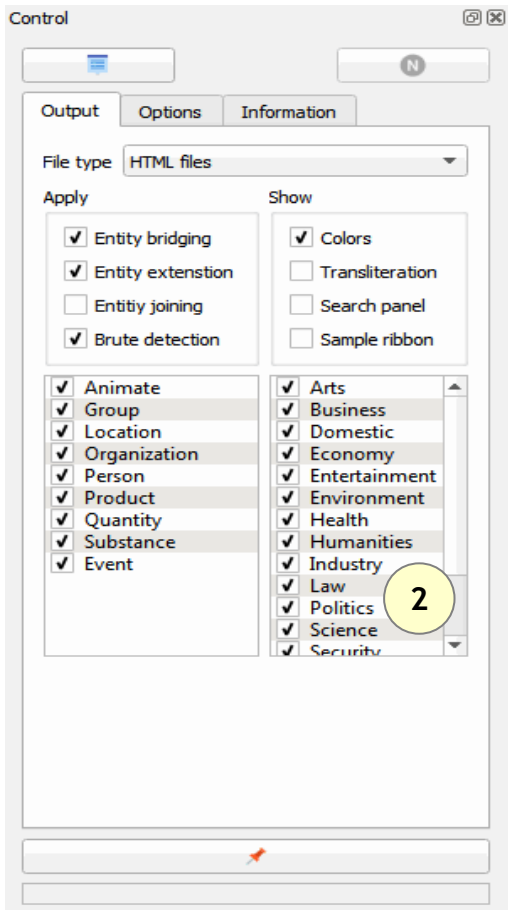
1. Set extraction to [Content classification only]
2. Load source files using (  ) from the Input Pane
3. Start extraction by clicking (  )



## Listing files classified to specific domain

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

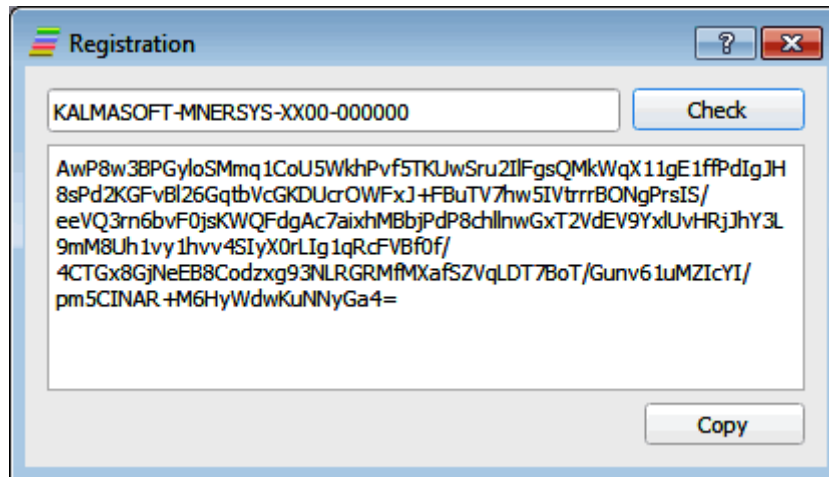
1. Make sure [Content classification] option is checked
2. Select domains to be classified
3. Load source files using (📁) from the Input Pane
4. Start classification by clicking (N)



## Registration

You need to register the evaluation copy before you are able to use all features.

- Open the (Help) menu
- Click (Registration)
- the following screen emerges



- Enter the main key (KALMASOFT-MNERSYS-XX00-000000)
- Click (Check) to generate your local key
- Click copy and send the generated code to Kalmasoft by email to get your registered copy

## Frequently asked questions

### Input Interface issues

#### Can't load files

- You may be trying to load an unsupported file type, check the list of supported file types; if your file is not supported the easiest way is to copy the text and paste them to the Input Pane; you may also save the file as txt from the native application.

#### System halts during load

- You are trying to load a huge file that could possibly render the system unresponsive, interrupt the system, divide the file, and try again.

### Setting issues

#### My settings don't work

- Some settings may only apply under certain circumstances, for instance, the [Use Gazetteers] is useless if no database available.
- Some settings may look contradicting for the first time, this is not the case, however, if you think that you have messed things up close the system and start again; if you have already saved any custom settings restore the defaults and try again with a fresh settings.

#### My new settings do not seem to affect the results

- Some specific settings may not be available in your current version e.g. language varieties and regional dialects, these will be introduced in the future.

### Display issues

#### Interface language looks "Greek"

- Make sure that you select the right language for your interface, if it still looks strange check if you have the right fonts installed in your system.
- If you still can't figure out what's on the screen reset the system setting from the options menu.

#### Extracted entities don't show up

- Make sure you selected the right input source, you will be notified if the source is empty.

## Filtering and sorting issues

### Filtering doesn't work

- Check the criteria you are trying to apply.
- Make sure that the results are not filtered already, clear the current filter text and apply your new filter.
- Perhaps no matches for the filter you set.

## Tips and tricks

The following are few notes to speed up the work and add up for a professional use of the system:

- Hide either the Input Panes or the Control Pane to spare a space for large view of extracted entities the result display area, or save to XML for better and easier look.