

## TERMINUS 2.0 USER'S GUIDE

Terminus is a web application for corpus and terminology management. It allows creating and managing terminological projects and selecting data categories. It integrates the complete work process of individual and/or team terminographic work: textual corpus search, compilation and analysis, term extraction, glossary and project management, database creation and maintenance, and dictionary edition.

Terminus contains different articulated modules:

**Projects:** to create a terminological project.

**Sources:** to manage the sources used in a terminological project.

**Conceptual structure:** to create a concept tree in order to structure the terms of a given glossary.

**Documents:** to upload text files in order to build a extraction corpus.

**Corpus:** to select the documents to be included in the corpus.

**Corpus analysis:** to make a statistical corpus exploitation. Word frequency, concordances, n-grams and association measures can be extracted. Terminus also allows term extraction from LSP corpus.

**Glossaries:** to create and manage the glossaries of a project, and to import glossaries.

**Terms:** to introduce, consult and export terminological data within a glossary.

## TABLE OF CONTENTS

1. Log in into Terminus _____	2
2. Basic management _____	3
3. <i>Conceptual structure</i> module _____	5
4. <i>Documents</i> module _____	8
5. <i>Corpus</i> module _____	12
6. <i>Analysis</i> module _____	14
7. <i>Glossaries</i> module _____	24
8. <i>Terms</i> module _____	25

## 1. Log in into Terminus

In order to access Terminus, registered users must introduce their log-in data (username and password) and select the interface language.

erminus 2.0

Home | User's guide | User details | Projects | Sources | Contact us | Exit

User: Natalia Seghezzi  
Group: Workshop1

**Conceptual structure**  
[Concept tree](#)

**Documents**  
[Upload a document](#)  
[Upload a group of documents](#)  
[File manager](#)

**Corpus**  
[New corpus](#)  
[Corpus manager](#)

**Analysis**  
[New analysis](#)  
[Saved analysis manager](#)  
[Validated analysis manager](#)

**Glossaries**  
[Glossary manager](#)  
[Import a glossary](#)

**Terms**  
[New term](#)  
[Term search and export](#)

**Terminus** is a web application for corpus and terminology management. It allows creating and managing terminological projects and selecting data categories. It integrates the complete work process of individual and/or team terminographic work: textual corpus search, compilation and analysis, term extraction, glossary and project management, database creation and maintenance, and dictionary edition. Terminus contains different articulated modules:

- **Projects:** for creating a terminological project.
- **Sources:** for managing the sources used in a terminological project.
- **Conceptual structure:** for creating a concept tree in order to structure the terms of a given glossary.
- **Documents:** for uploading text files which will constitute the working corpus.
- **Corpus:** for selecting documents to build a corpus.
- **Corpus analysis:** for statistical corpus exploitation. Word frequency, concordances, n-grams and association measures can be extracted. It also for term extraction from LSP corpus.
- **Glossaries:** for creating and managing the glossaries of a project.
- **Terms:** for introducing terminological data into a glossary, retrieve them and export them.

UNIVERSITAT POMPEU FABRA iuliaTerm

Once on the system, all the work options offered by Terminus, distributed in several modules, can be seen on the left part of the screen: *Conceptual Structure*, *Documents*, *Corpus*, *Analysis*, *Glossaries* and *Terms*. In addition, the basic management options can be seen on the upper part of the screen: *Home*, *User's guide*, *User details*, *Projects*, *Sources*, *Contact us* and *Exit*. Next some details about the main functions of the different modules are provided.

## 2. Basic management

**erminus 2.0**

Home | User's guide | User details | Projects | Sources | Contact us | Exit

User: Natalia Seghezzi  
Group: Workshop1

**Conceptual structure**  
Concept tree

**Documents**  
Upload a document  
Upload a group of documents  
File manager

**Corpus**  
New corpus  
Corpus manager

**Analysis**  
New analysis  
Saved analysis manager  
Validated analysis manager

**Glossaries**  
Glossary manager  
Import a glossary

**Terms**  
New term  
Term search and export

**Terminus** is a web application for corpus and terminology management. It allows creating and managing terminological projects and selecting data categories. It integrates the complete work process of individual and/or team terminographic work: textual corpus search, compilation and analysis, term extraction, glossary and project management, database creation and maintenance, and dictionary edition. Terminus contains different articulated modules:

- **Projects:** for creating a terminological project.
- **Sources:** for managing the sources used in a terminological project.
- **Conceptual structure:** for creating a concept tree in order to structure the terms of a given glossary.
- **Documents:** for uploading text files which will constitute the working corpus.
- **Corpus:** for selecting documents to build a corpus.
- **Corpus analysis:** for statistical corpus exploitation. Word frequency, concordances, n-grams and association measures can be extracted. It also for term extraction from LSP corpus.
- **Glossaries:** for creating and managing the glossaries of a project.
- **Terms:** for introducing terminological data into a glossary, retrieve them and export them.

UNIVERSITAT POMPEU FABRA iulaTerm

Terminus basic management options are the following:

*Home:* Link to Terminus home page.

*User's guide:* Link to Terminus guideline.

*User details:* To modify the user's name, surname, password and/or e-mail address.

**erminus 2.0**

Home | User's guide | User details | Projects | Sources | Contact us | Exit

User: Natalia Seghezzi  
Group: Workshop1

**Conceptual structure**  
Concept tree

**Documents**  
Upload a document  
Upload a group of documents  
File manager

**Corpus**  
New corpus  
Corpus manager

**Analysis**  
New analysis  
Saved analysis manager  
Validated analysis manager

**Glossaries**  
Glossary manager  
Import a glossary

**Terms**  
New term  
Term search and export

**Update user details**

Update your details here.

Log-in name: Natalia Seghezzi

Name and surname(s): Natalia Seghezzi

New password:

Re-enter your password:

New e-mail address: natalia.seghezzi@upf.edu

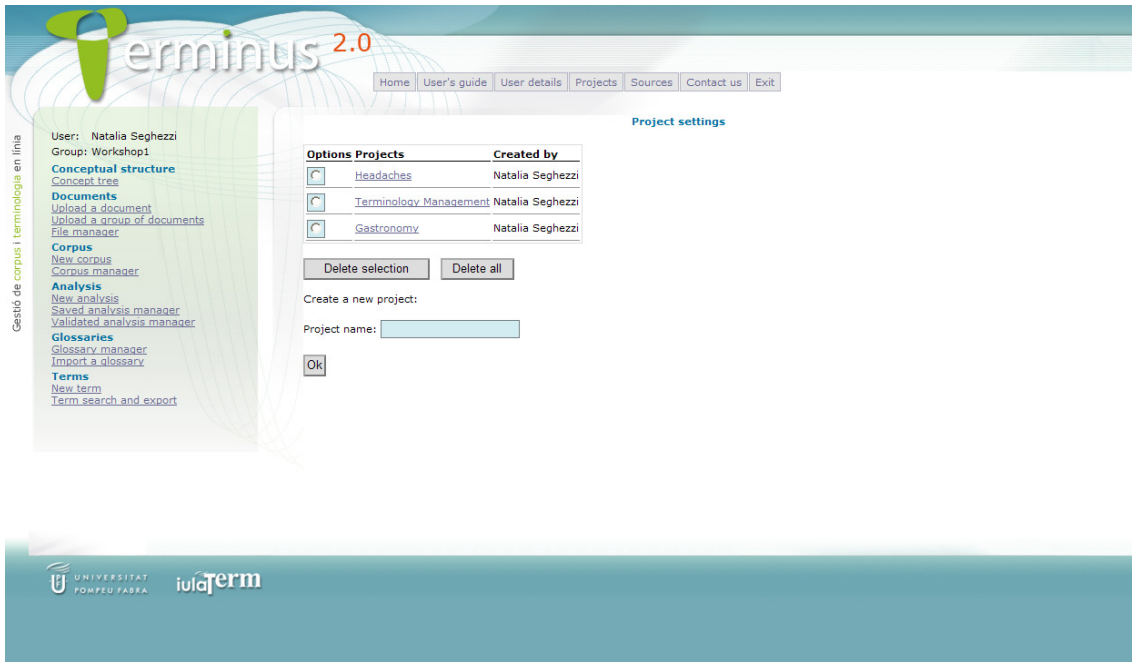
Please re-enter your new e-mail address: natalia.seghezzi@upf.edu

To prevent accidental changes, please enter your old password:

Ok Reset form

UNIVERSITAT POMPEU FABRA iulaTerm

*Projects:* Terminus allows creating, modifying and eliminating projects. Please notice that each work team may have more than one project.



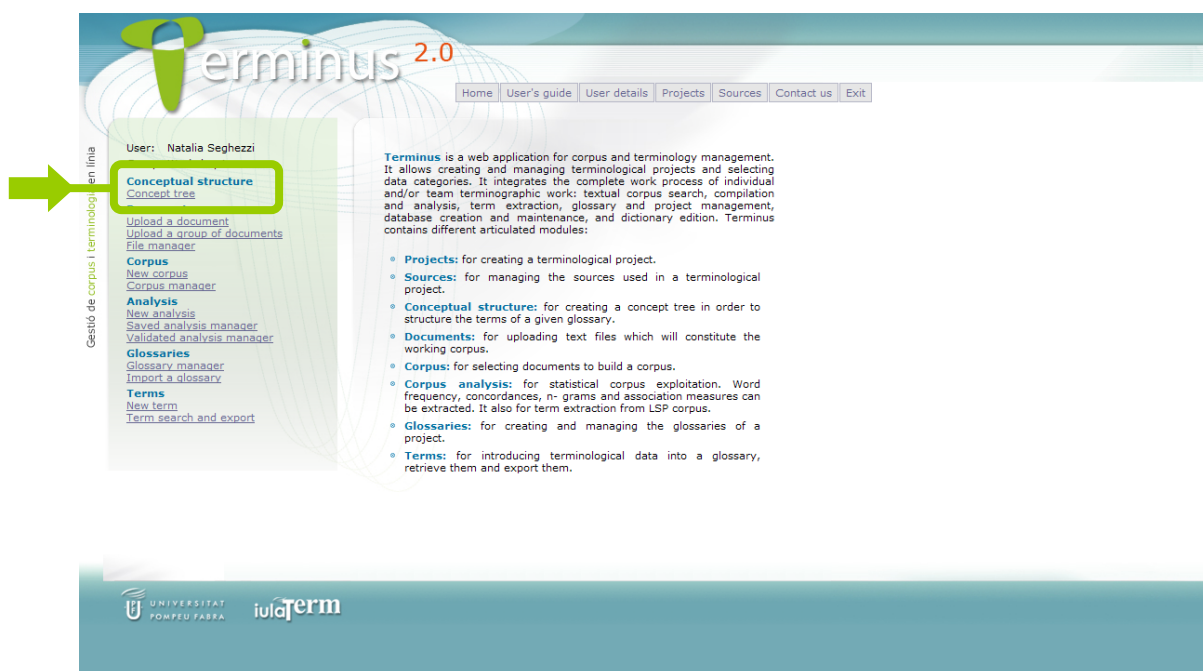
**Sources:** The sources for terms, contexts and definitions can be managed, i.e. edited or deleted. It is important to remember that any change in a source is automatically updated everywhere in the database.



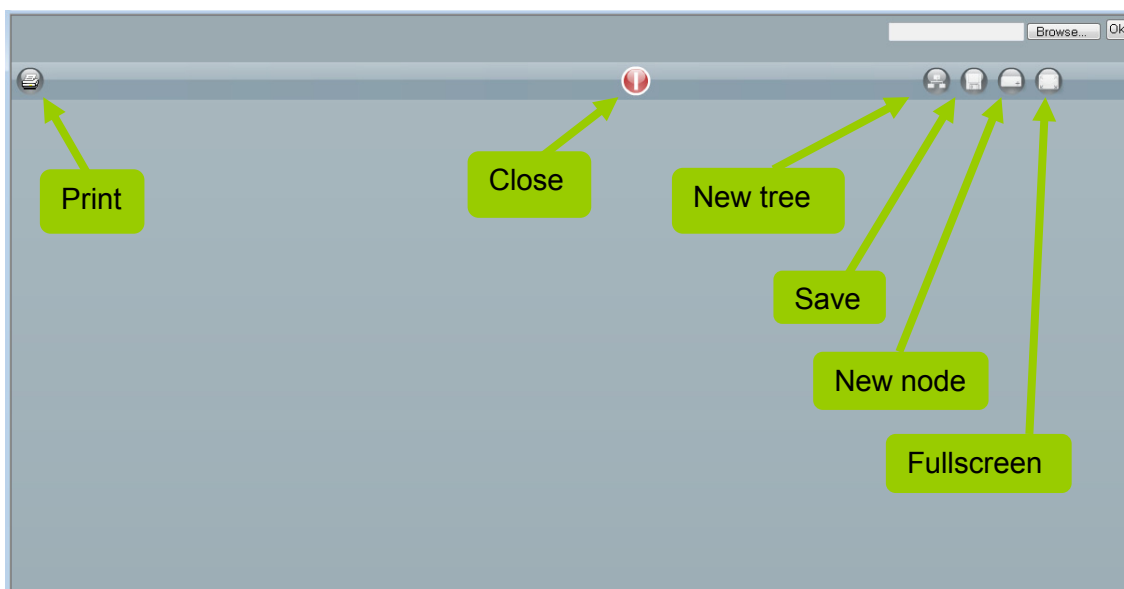
**Contact us:** To send comments or bug reports to the webmaster.

**Exit:** To log out.

### 3. Conceptual structure module



The *Conceptual structure* module offers the possibility to create a concept tree in order to organize the terms of a given glossary.



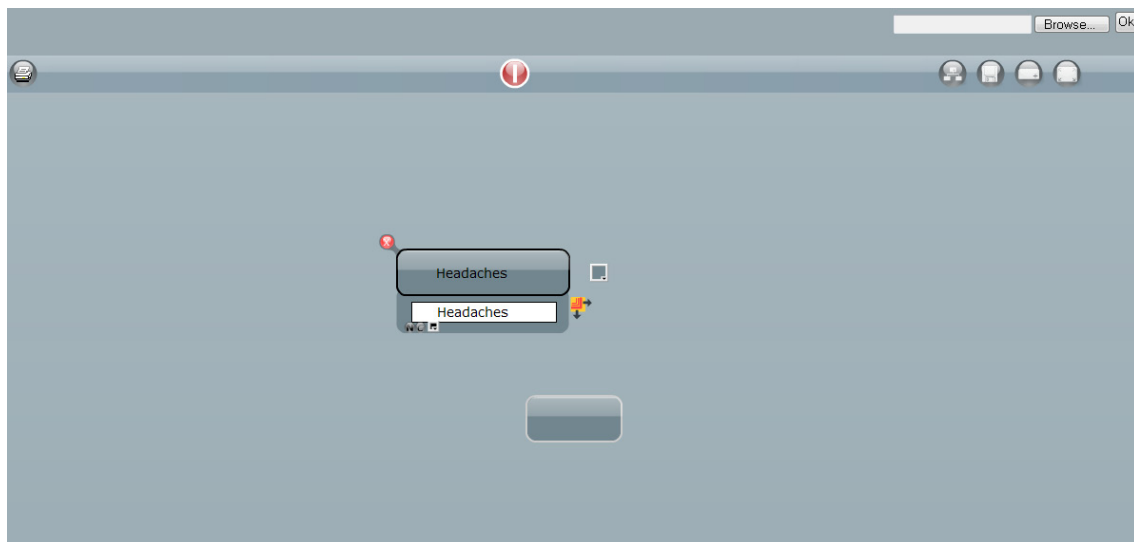
Click on *Concept tree* to access the tool click on *Concept tree* to access the tool (the system may ask you to install Adobe Flash Player in order to open the Concept tree tool). To have a full screen format, click on *Fullscreen* (fourth icon on the upper right corner). In order to start creating your concept tree, click first on *New node* (third icon on the upper right corner).

You can create as many nodes as you want for each tree. Please notice that, each time you create a new node, it appears in the centre of the screen; thus, when several

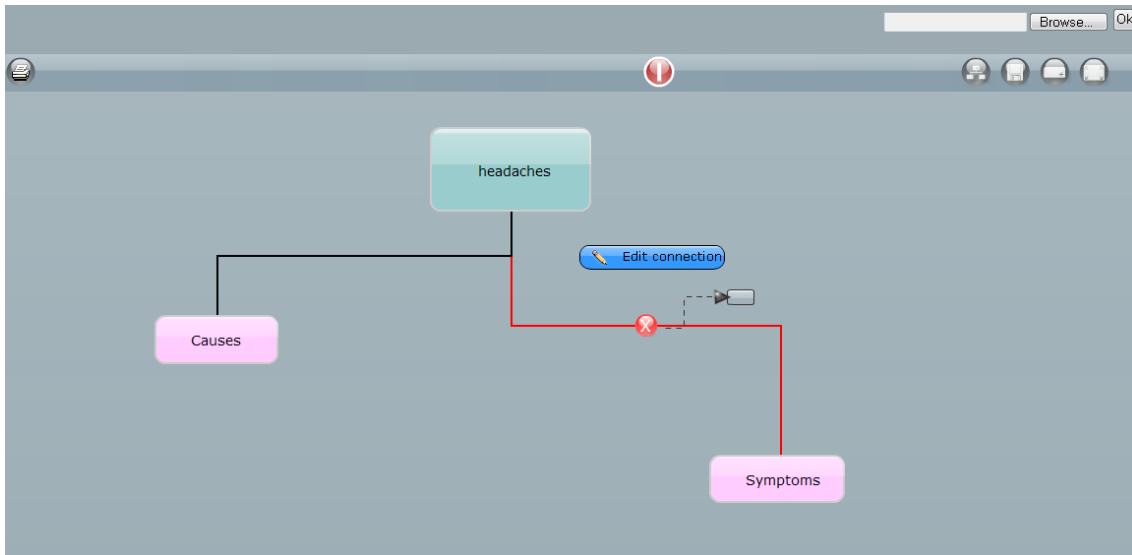
nodes are inserted one after the other they overlap and it is necessary to move them along the screen in order to differentiate them. To move the nodes, click on them and simply drag them along the screen. If you want to connect them with a line, put the two nodes one next to the other and the line will be drawn automatically.

Important: if you click on *New tree* (first icon on the upper right corner), the screen will go blank and all the content will be lost; that is why it is highly recommended that you save the tree as soon as you start working. To save the concept tree on your computer, click on *Save* (floppy disk icon on the upper right corner).

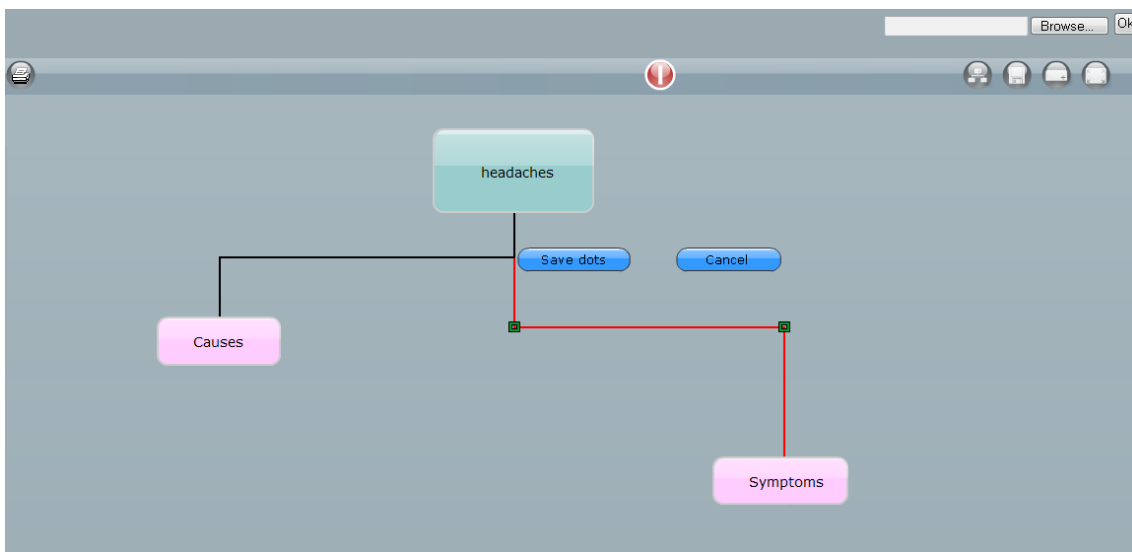
Click twice on a node for all the functions to appear in order to modify it. To introduce text, write in the blank box below the node and then press the return (Enter) key. If you want to return within the same node, press Caps lock and the return (Enter) key at the same time. In order to change the font style (bold, italics and colour), use the specific icons on the lower left part. To change the colour of the node, click on the right hand square. To delete a node, click on the X red icon on the upper left part.



The lines connecting the nodes can also be modified. First you must click on the line for the editing function to appear (the line will change into red colour; the discontinuous arrow allows extending the line to a non-connected node).



To remove a connecting line click on the X red icon. To modify a line click on *Edit connection*; when this is done two square inflection dots appear in order to move the line; when the line and node are placed in the desired position, click on *Save dots* and finally click on the red line for it to turn black again.



Please note that concept trees created with Terminus editor can only be open with Terminus' concept tree editor. In order to see a saved concept tree, open a new white screen (*New tree*), click on Browse (or the corresponding button, depending on the language of your browser), look for the file on your computer and press OK.

The concept tree created with Terminus editor can be printed by clicking on *Print* (printer icon on the upper left corner). Finally, the red icon in the top middle of the screen allows to exit Terminus' *Conceptual structure* module.

#### 4. Documents module

The *Documents* module offers three possibilities: uploading a document onto the system from an URL or a file stored in your computer, uploading several documents together (in one single step) either by means of a zip file containing several files, a file containing several URLs or an Internet search, and managing all of a user's documents.

The screenshot shows the Terminus 2.0 web application interface. The user is logged in as Natalia Seghezzi, Group: Workshop1. The main navigation menu on the left includes: Conceptual structure, Documents (highlighted with a green box and an arrow), Corpus, Analysis, Glossaries, and Terms. The Documents menu has sub-items: Upload a document, Upload a group of documents, and File manager. The main content area displays a description of Terminus and a list of modules: Projects, Sources, Conceptual structure, Documents, Corpus, Corpus analysis, Glossaries, and Terms.

- *Upload a document*: To upload a single document you must give it a name (to identify it later), and specify its origin and location (URL or desktop file), source (for example bibliographic reference) and language.

The screenshot shows the 'Upload a document' form in the Terminus 2.0 web application. The form fields are: Name (Primary headaches), URL (http://www.bmj.com/highw), Local file (with a Browse... button), Source (The British Medical Journal (BMJ) http://www.bmj.com/), and Language (English). There is an 'Ok' button at the bottom of the form.



Once the document is loaded, a fragment of it is shown on the screen. It is possible to save the document in your desktop in TXT format.

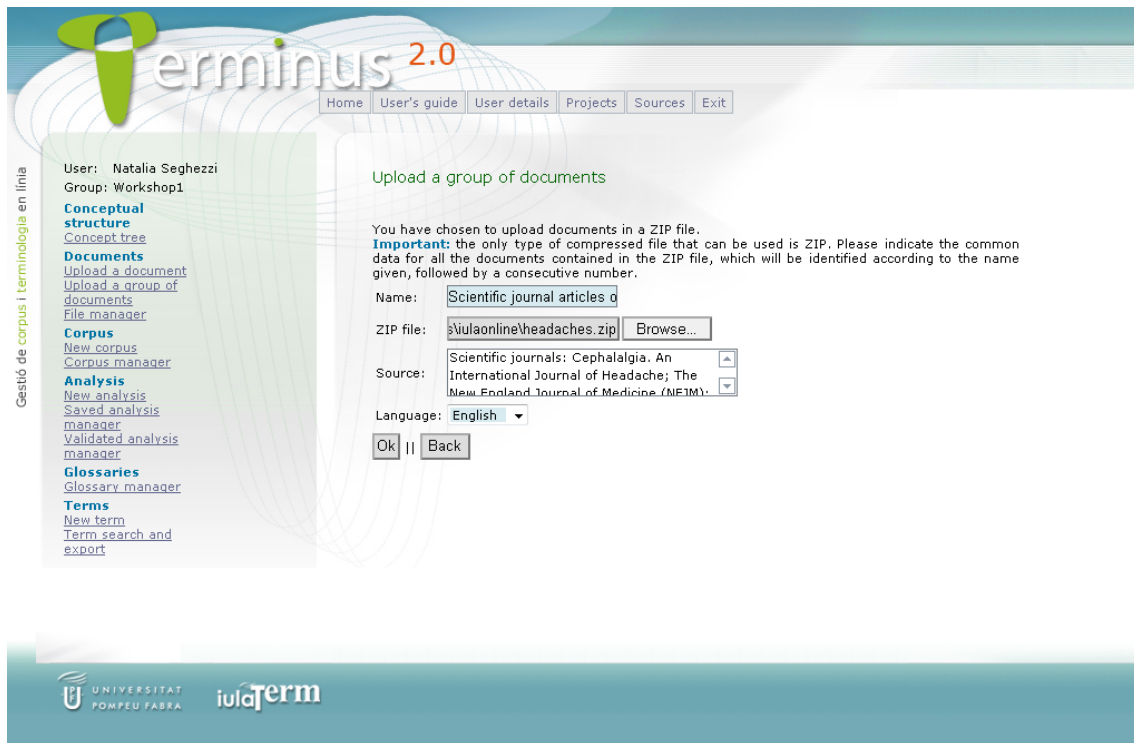


- *Upload a group of documents*: To upload a new group of documents, select one of the three available options (zip file, URL or Internet search).



If the zip file option is chosen, Terminus will ask the user to specify some common data for all the documents in the group (identification name, source and language) and look

for the file to upload. Please notice that, in the present version of Terminus 2.0, it is not possible to upload other than ZIP files (other formats such as RAR files are not accepted) and it is important that they contain only individual files, i.e. that the ZIP file does not contain any folders.



If the URL option is chosen, Terminus will also ask the user to specify some common data for all the documents in the group (identification name, source and language) and the file to upload. Please notice that the file must be a TXT file and contain one URL per line.



If you choose to get the documents from the Internet, Terminus will ask you to specify an identification name for the group of documents and the following search fields: the query expression or term to be searched for, the maximum number of documents, the file type (PDF, DOC, PS or indifferent), and the language. There is also the possibility to apply further restrictions to the search by indicating a particular Internet domain, and to include some units that should or should not appear in the texts yielded by the search (in the first case up to 5 units can be included, and in the second up to 100 units; in both cases the units can be written one per line, or one after the other separated by comma, or semicolon).

**Terminus 2.0**

Home | User's guide | User details | Projects | Sources | Contact us | Exit

User: Natalia Seghezzi  
Group: Workshop1

**Upload a group of documents**

You have chosen to search for the documents on the Internet. Please indicate the common data for all the documents resulting from the search, which will be identified according to the name given, followed by a consecutive number.

Name:  Query expression:   
(Mandatory field)

Number of documents to download:  Document type:

Language:  Domain:   
(For example: upf.edu)

Terms that should co-occur with the query expression:  Terms that should NOT co-occur with the query expression:

Ok | Back

- **File manager:** Users can manage the documents which have been saved in Terminus. It is possible to view the documents (by clicking on the document name), edit or delete them. You can also download texts in TXT format and save them onto your desktop.

**Terminus 2.0**

Home | User's guide | User details | Projects | Sources | Contact us | Exit

User: Natalia Seghezzi  
Group: Workshop1

**File manager**

Options	Name	Language	Source
<input type="checkbox"/>	Primary headaches	English	The British Medical Journal (BMJ) <a href="http://www.bmj.com/">http://www.bmj.com/</a>
<input type="checkbox"/>	Headache from Internet1	English	<a href="http://www.wikihow.com/Get-Rid-of-a-Headache">http://www.wikihow.com/Get-Rid-of-a-Headache</a>
<input type="checkbox"/>	Headache from Internet2	English	<a href="http://www.uptodate.com/contents/patient-information-headache-causes-and-diagnosis-in-adults">http://www.uptodate.com/contents/patient-information-headache-causes-and-diagnosis-in-adults</a>
<input type="checkbox"/>	Headache from Internet3	English	<a href="http://www.medicinenet.com/migraine_headache/article.htm">http://www.medicinenet.com/migraine_headache/article.htm</a>
<input type="checkbox"/>	Headache from Internet4	English	<a href="http://www.mayoclinic.com/health/migraine-headache/DS00120">http://www.mayoclinic.com/health/migraine-headache/DS00120</a>
<input type="checkbox"/>	Headache from Internet5	English	<a href="http://www.healthscout.com/ency/68/96/main.html">http://www.healthscout.com/ency/68/96/main.html</a>
<input type="checkbox"/>	Headache from Internet6	English	<a href="http://www.medicinanet.com/headache/article.htm">http://www.medicinanet.com/headache/article.htm</a>
<input type="checkbox"/>	Headache from Internet7	English	<a href="http://en.wikipedia.org/wiki/Migraine">http://en.wikipedia.org/wiki/Migraine</a>
<input type="checkbox"/>	Headache from Internet8	English	<a href="http://www.webmd.com/migraines-headaches/guide/migraines-headaches-diagnosis-tests">http://www.webmd.com/migraines-headaches/guide/migraines-headaches-diagnosis-tests</a>
<input type="checkbox"/>	Headache from Internet9	English	<a href="http://headaches.about.com/">http://headaches.about.com/</a>
<input type="checkbox"/>	Headache from Internet10	English	<a href="http://www.mayoclinic.com/health/headaches/HE00001">http://www.mayoclinic.com/health/headaches/HE00001</a>
<input type="checkbox"/>	Scientific journal articles on headaches1	English	Scientific journals: Cephalalgia, An International Journal of Headache; The New England Journal of Medicine (NEJM); The British Medical Journal (BMJ)

Delete selection | Delete all

## 5. Corpus module

Based on the documents uploaded in a previous step (see *Documents* module), the *Corpus* module allows creating and managing specialized corpus.

The screenshot shows the erminius 2.0 web application interface. The user is Natalia Seghezzi, Group: Workshop1. The main menu includes: Conceptual structure, Documents, Corpus, Analysis, Glossaries, and Terms. The **Corpus** module is highlighted with a green box and a green arrow. The Corpus module options are: New corpus, Corpus manager, and Analysis. The Analysis module options are: New analysis, Saved analysis manager, and Validated analysis manager. The Glossaries module options are: Glossary manager and Import a glossary. The Terms module options are: New term and Term search and export. The main content area contains a description of Terminus and a list of modules: Projects, Sources, Conceptual structure, Documents, Corpus, Corpus analysis, Glossaries, and Terms.

The *Corpus* module contains the following two options:

- *New corpus*: In order to create a new corpus, give the corpus a name and select the documents. It is possible to create as many corpus as needed. At the moment of creation, users may select as many documents as necessary.

The screenshot shows the erminius 2.0 web application interface with the 'New corpus' form. The user is Natalia Seghezzi, Group: Workshop1. The main menu includes: Conceptual structure, Documents, Corpus, Analysis, Glossaries, and Terms. The **Corpus** module is highlighted with a green box and a green arrow. The Corpus module options are: New corpus, Corpus manager, and Analysis. The Analysis module options are: New analysis, Saved analysis manager, and Validated analysis manager. The Glossaries module options are: Glossary manager and Import a glossary. The Terms module options are: New term and Term search and export. The main content area contains the 'New corpus' form. The form has two steps: 1. Enter the name of the new corpus: Name: Headaches. 2. Select the documents for the corpus:  Select documents whose name has at least partial coincidence with the name of the corpus. The table below shows the selected documents:

Options	Name	Language
<input checked="" type="checkbox"/>	Primary headaches	English
<input checked="" type="checkbox"/>	Headache_from_Internet1	English
<input checked="" type="checkbox"/>	Headache_from_Internet2	English
<input checked="" type="checkbox"/>	Headache_from_Internet3	English

Buttons: Ok, Select all, Reset form.

- *Corpus manager*: By clicking on the corpus name, the corpus can be viewed, edited or deleted.

The screenshot displays the erminis 2.0 web interface. At the top, there is a navigation menu with links for Home, User's guide, User details, Projects, Sources, Contact us, and Exit. The main content area is divided into several sections:

- User Information:** User: Natalia Seghezzi, Group: Workshop1
- Conceptual structure:** Concept tree
- Documents:** Upload a document, Upload a group of documents, File manager
- Corpus:** New corpus, Corpus manager
- Analysis:** New analysis, Saved analysis manager, Validated analysis manager
- Glossaries:** Glossary manager, Import a glossary
- Terms:** New term, Term search and export


The **Corpus manager** section is highlighted, showing a table with the following data:

Options	Name
<input type="checkbox"/>	Headaches
<input type="checkbox"/>	Terminology
<input type="checkbox"/>	Gastronomy

Below the table are two buttons: "Delete selection" and "Delete all".

At the bottom of the page, there is a footer with the logos for UNIVERSITAT POMPEU FABRA and iulqTerm.

## 6. Analysis module



The screenshot shows the Terminus 2.0 web application interface. The top navigation bar includes links for Home, User's guide, User details, Projects, Sources, Contact us, and Exit. The user information section displays: User: Natalia Seghezzi, Group: Workshop1. The main menu is organized into several categories: Conceptual structure (Concept tree), Documents (Upload a document, Upload a group of documents, File manager), Corpus (New corpus, Corpus manager), Analysis (New analysis, Saved analysis manager, Validated analysis manager), Glossaries (Glossary manager, Import a glossary), Terms (New term, Term search and export). A green box highlights the 'Analysis' section, and a green arrow points to it from the left. Below the menu, there is a brief description of Terminus and a list of modules: Projects, Sources, Conceptual structure, Documents, Corpus, Corpus analysis, Glossaries, and Terms.

The *Analysis* module includes four main types of statistical corpus analyses: n-gram extraction, association measures, keyword in context and term extraction. Click on *New analysis*, select the desired analysis method and the corpus to be explored.



The screenshot shows the 'Corpus analysis' step 1 of 2 in the Terminus 2.0 web application. The user information remains the same. The main menu is visible on the left. The 'Corpus analysis' section is active, showing two steps: 1. Choose analysis method and 2. Select the corpus to analyse. Under step 1, there are four radio button options: N-gram extraction, Association measures, Keyword in context, and Term extraction. Under step 2, there are three radio button options: Headaches, Terminology, and Gastronomy. An 'Ok' button is located below the options.

*N-gram extraction*: Terminus extracts n-grams (sequences of  $n$  words) from the corpus according to the analysis options selected by the user. The options include the possibility to select the number of grams ( $n$ ), the minimum frequency of the n-gram in

the corpus, a form that must be contained in the n-gram, ignore the numbers in the corpus, include a stoplist with grammatical words (that is why the language of the corpus must be selected), select the maximum number of results and sort results by frequency or alphabetic order.

The result of this analysis is a list of n-grams with their absolute and relative frequency.

No.	Units	Absolute frequency	Relative frequency
1) <input type="checkbox"/>	tension-type headache	62	0.00211193
2) <input type="checkbox"/>	mayo clinic	40	0.00136254
3) <input type="checkbox"/>	migraine headaches	37	0.00126035
4) <input type="checkbox"/>	without aura	37	0.00126035
5) <input type="checkbox"/>	migraine headache	34	0.00115816
6) <input type="checkbox"/>	headache subtypes	32	0.00109003
7) <input type="checkbox"/>	migraine without	30	0.00102190
8) <input type="checkbox"/>	tension headaches	24	0.00081752
9) <input type="checkbox"/>	cluster headaches	24	0.00081752
10) <input type="checkbox"/>	blood vessels	18	0.00061314
11) <input type="checkbox"/>	chronic daily	14	0.00047689
12) <input type="checkbox"/>	headache subtype	12	0.00040876
13) <input type="checkbox"/>	accessed feb	12	0.00040876
14) <input type="checkbox"/>	primary headaches	11	0.00037470
15) <input type="checkbox"/>	met criteria	11	0.00037470
16) <input type="checkbox"/>	patient information	11	0.00037470
17) <input type="checkbox"/>	some people	11	0.00037470
18) <input type="checkbox"/>	blood pressure	10	0.00034063
19) <input type="checkbox"/>	medical education	10	0.00034063
20) <input type="checkbox"/>	secondary headaches	10	0.00034063
21) <input type="checkbox"/>	clinical features	10	0.00034063
22) <input type="checkbox"/>	migraine attacks	10	0.00034063
23) <input type="checkbox"/>	accessed march	10	0.00034063

From this list the term candidates can be validated (as will be seen later in this section).

**Association measures:** Terminus allows the extraction of a list of bigrams which present some degree of association. This is done by analysing all the bigrams in the corpus or those bigrams indicated by the user containing a specific unit. The association measure can be obtained with three different measures (t-score, chi square and mutual information), and according to the analysis options selected by the user. The options include the possibility to ignore the numbers in the corpus, include a stoplist with grammatical words (that is why the language of the corpus must be selected), select the maximum number of results and sort results by weighting or alphabetic order.

The screenshot displays the Terminus 2.0 web interface. At the top, the logo 'erminus 2.0' is visible. A navigation menu includes 'Home', 'User's guide', 'User details', 'Projects', 'Sources', 'Contact us', and 'Exit'. The user information shows 'User: Natalia Seghezzi' and 'Group: Workshop1'. The main content area is titled 'Corpus analysis' and is 'Step 2 of 2'. It features a 'Choose analysis options' section with an 'Analysis name' input field. The 'Association measures' section includes options for calculation type (all bigrams or specific unit), minimum frequency (set to 3), and whether to ignore numbers in the corpus (checked). Under 'Measures', 'Mutual Information' is selected. Other options include language selection, a checked 'Include stoplist' option, and a maximum number of results set to 600. 'Ok', 'Reset form', and 'Back' buttons are at the bottom. A vertical sidebar on the left lists various management functions like 'Conceptual structure', 'Documents', 'Corpus', 'Analysis', 'Glossaries', and 'Terms'. The footer contains the logos for 'UNIVERSITAT POMPEU FABRA' and 'iulTerm'.

The result of this analysis is a list of bigrams indicating the weighting obtained according to the measure chosen. From this list the term candidates can be validated (as will be seen later in this section).



Terminus 2.0

Home User's guide User details Projects Sources Contact us Exit

User: Natalia Seghezzi  
Group: Workshop1

**Corpus analysis**  
Step 2 of 2  
Association measures

Validate selected candidates

No.	Units	Mutual Information
1) <input type="checkbox"/>	stewart wf	13.2056
2) <input type="checkbox"/>	cortical spreading	13.2056
3) <input type="checkbox"/>	cutrer fm	13.2056
4) <input type="checkbox"/>	rasmussen bk	13.2056
5) <input type="checkbox"/>	mutually exclusive	13.2056
6) <input type="checkbox"/>	diabetes digestion	12.8837
7) <input type="checkbox"/>	unsourced statements	12.8837
8) <input type="checkbox"/>	lpton rh	12.6207
9) <input type="checkbox"/>	brown reclusa	12.6207
10) <input type="checkbox"/>	facebook twitter	12.6207
11) <input type="checkbox"/>	united states	12.3983
12) <input type="checkbox"/>	full profile	12.2987
13) <input type="checkbox"/>	kids hearing	12.2987
14) <input type="checkbox"/>	close window	12.2056
15) <input type="checkbox"/>	black widow	12.2056

UNIVERSITAT POMPEU FABRA iulqTerm

**Keyword in context:** It is possible to consult the contexts of occurrence of a particular unit (which can be simple or syntagmatic, i.e. a phrase) by extracting concordances. To search for key words in context the term to be sought must be indicated.

Terminus 2.0

Home User's guide User details Projects Sources Exit

User: Natalia Seghezzi  
Group: Workshop1

**Keyword in context**  
Step 2 of 2

Choose analysis options  
Analysis name: concordances\_headache

Query expression: headache

Size of the context window: 15 tokens

Ok Reset form Back

UNIVERSITAT POMPEU FABRA iulqTerm

The result of this analysis is a list of contexts of occurrence (concordances) of the unit searched for in the corpus.

**erminusus 2.0**

Home | User's guide | User details | Projects | Sources | Exit

User: Natalia Seghezzi  
Group: Workshop1

**Conceptual structure**  
Concept tree

**Documents**  
Upload a document  
Upload a group of documents  
File manager

**Corpus**  
New corpus  
Corpus manager

**Analysis**  
New analysis  
Saved analysis manager  
Validated analysis manager

**Glossaries**  
Glossary manager

**Terms**  
New term  
Term search and export

**Corpus analysis**  
Step 2 of 2  
Keyword in context  
Back

Query expression: headache

1)	<a href="#">Primary headaches</a>	magnitude, impact, and stability of primary	<a href="#">headache</a>	subtypes 30 year prospective swiss cohort study open access kathleen r merikangas senior investigator 1
2)	<a href="#">Primary headaches</a>	abstract objective to determine the prevalence, impact, and stability of different subtypes of	<a href="#">headache</a>	in a 30 year prospective follow-up study of a general population sample. design prospective
3)	<a href="#">Primary headaches</a>	, interviewed seven times across 30 years of follow-up. main outcome measures prevalence of	<a href="#">headache</a>	stability of the predominant subtype of
4)	<a href="#">Primary headaches</a>	stability of the predominant subtype of	<a href="#">headache</a>	over time; and age of onset, severity, impact, family history,
5)	<a href="#">Primary headaches</a>	severity, impact, family history, use of healthcare services, and drugs for	<a href="#">headache</a>	subtypes. results the average one year prevalences of subtypes of
6)	<a href="#">Primary headaches</a>	subtypes. results the average one year prevalences of subtypes of	<a href="#">headache</a>	were 0. 9% ( femalemale ratio of 2. 8 ) for migraine with
7)	<a href="#">Primary headaches</a>	, and 11. 5% ( femalemale ratio of 1. 2 ) for tension-type	<a href="#">headache</a>	cumulative 30 year prevalences of
8)	<a href="#">Primary headaches</a>	cumulative 30 year prevalences of	<a href="#">headache</a>	subtypes were 3. 0% for migraine with aura, 36. 0% for migraine
9)	<a href="#">Primary headaches</a>	, 36. 0% for migraine without aura, and 29. 3% for tension-type	<a href="#">headache</a>	despite the high prevalence of migraine without aura, most cases were transient and only
10)	<a href="#">Primary headaches</a>	the follow-up period. 69% of participants with migraine and 58% of those with tension-type	<a href="#">headache</a>	manifested the same predominant subtype over time. however, the prospective stability of the

UNIVERSITAT POMPEU FABRA | iulaterm

**Term extraction:** Terminus allows a semiautomatic term extraction which the user will have to validate later. This analysis has two options: training the term extractor in a specific domain by incorporating an electronic dictionary containing terms of the same field, or applying the term extractor to a textual corpus. In the latter case it must be noticed that the quality of the corpus (i.e. its degree of specialization and pertinence to a specific special area) is essential to ensure the quality of the term extraction.

**erminusus 2.0**

Home | User's guide | User details | Projects | Sources | Contact us | Exit

User: Natalia Seghezzi  
Group: Workshop1

**Conceptual structure**  
Concept tree

**Documents**  
Upload a document  
Upload a group of documents  
File manager

**Corpus**  
New corpus  
Corpus manager

**Analysis**  
New analysis  
Saved analysis manager  
Validated analysis manager

**Glossaries**  
Glossary manager  
Import a glossary

**Terms**  
New term  
Term search and export

**Corpus analysis**  
Step 2 of 2

Term extraction:

Choose analysis options

[Train the extractor](#)

[Apply the extractor](#)

Back

UNIVERSITAT POMPEU FABRA | iulaterm

In order to train the term extractor the following steps must be followed: first, select the training language (even though the extractor is language independent, the work language must be indicated to enhance performance); second, include a training dictionary: indicate a name, a brief description and the file containing the complete dictionary (please note that it must be a TXT file, in UTF8 encoding, with one term per line). Once the extractor has been trained, it can be applied to extract terms from a textual corpus.



In order to apply the term extractor please indicate first the analysis name (which later will allow you to recover the files related with the extraction), Then select the language of the corpus to analyze, the training dictionary matching the domain of the specialized corpus from the list of available dictionaries<sup>1</sup>, the maximum number of results, the minimum frequency, the termhood threshold (this is a statistical measure; please note that 0.5 is a default value), and the desire to ignore the numbers in the corpus if necessary.

---

<sup>1</sup> New dictionaries can be uploaded by using the *Train the term extractor* option (remember that it must be a TXT file, in UTF8 encoding, with one term per line).



The term extraction yields three different files, which will be automatically stored in *Saved analysis manager*. The first file (*Part of speech tagging*) contains all the tagged<sup>2</sup> units of the corpus and their lemma; the second file (*Term candidates sorted by weighting*) shows the term candidates (simple and complex terms) ordered by their probability to be considered as terms according to the statistical analysis of the extractor; the third file (*Term candidates sorted by syntactic pattern*) contains the list of term candidates ordered according to their syntactic pattern and weighting.

---

<sup>2</sup> Further information about part of speech tagging in different languages can be found on the following webs:

<http://courses.washington.edu/hypertext/csar-v02/penntable.html> (English)

<http://www.ims.uni-stuttgart.de/~schmid/french-tagset.html> (French)

<http://www.recercat.net/bitstream/2072/77/1/98inf018.pdf> (Spanish and Catalan)



In the four types of statistical analyses (i.e. n-gram extraction, association measures, keyword in context and term extraction) the result is a list of units meeting the analysis options required by the user. The lists are automatically stored and can be consulted in *Saved analysis manager*.



In the case of the n-gram extraction, the association measures and the term extraction, the resulting units can be validated as terms and incorporated into a glossary already created in Terminus (see the *Glossaries* module, Section 7). In order to validate the term candidates that will be incorporated into the glossary go to *Saved analysis manager*, click to open the desired analysis result list, select the units and click on *Validate selected candidates*. The validated units will become part of another list (from

which the terms can be included in the glossary), and it will be possible to delete the rest of the units.

The screenshot shows the erminus 2.0 web interface. On the left, there is a navigation menu with categories: Conceptual structure, Documents, Corpus, Analysis, Glossaries, and Terms. The main content area displays a table titled "Validate selected candidates" with 15 rows of term candidates. The table has columns for "No.", "Units", "Absolute frequency", and "Relative frequency".

No.	Units	Absolute frequency	Relative frequency
1) <input checked="" type="checkbox"/>	tension-type headache	62	0.00211193
2) <input type="checkbox"/>	mayo clinic	40	0.00136254
3) <input type="checkbox"/>	migraine headaches	37	0.00126035
4) <input checked="" type="checkbox"/>	without aura	37	0.00126035
5) <input checked="" type="checkbox"/>	migraine headache	34	0.00115816
6) <input type="checkbox"/>	headache subtypes	32	0.00109003
7) <input type="checkbox"/>	migraine without	30	0.00102190
8) <input checked="" type="checkbox"/>	tension headaches	24	0.00081752
9) <input checked="" type="checkbox"/>	cluster headaches	24	0.00081752
10) <input type="checkbox"/>	blood vessels	18	0.00061314
11) <input type="checkbox"/>	chronic daily	14	0.00047689
12) <input type="checkbox"/>	headache subtype	12	0.00040876
13) <input type="checkbox"/>	accessed feb	12	0.00040876
14) <input type="checkbox"/>	primary headaches	11	0.00037470
15) <input type="checkbox"/>	met criteria	11	0.00037470

The lists of validated term candidates can be consulted in *Validated analysis manager*. In order to see the validated term candidates, select the desired analysis result list and click on *Create entries*. Please notice that the lists of validated term candidates only include the term candidates which have been selected and validated as terms by the user.

The screenshot shows the erminus 2.0 web interface in the "Validated analysis manager" section. It displays a list of three validated analysis result lists, each with a checkbox and a name:

- association\_measure\_headaches
- n-grams headaches
- term extraction headache

Below the list, there are three buttons: "Create entries", "Delete selection", and "Delete all".

Click on the desired validated term candidate in order to create the terminological record (see Section 8 for a detailed description of the process for introducing terms into Terminus).

erminus 2.0

Home | User's guide | User details | Projects | Sources | Contact us | Exit

Gestió de corpus i terminologia en línia

User: Natalia Seghezzi  
Group: Workshop1

**Conceptual structure**  
[Concept tree](#)

**Documents**  
[Upload a document](#)  
[Upload a group of documents](#)  
[File manager](#)

**Corpus**  
[New corpus](#)  
[Corpus manager](#)

**Analysis**  
[New analysis](#)  
[Saved analysis manager](#)  
[Validated analysis manager](#)

**Glossaries**  
[Glossary manager](#)  
[Import a glossary](#)

**Terms**  
[New term](#)  
[Term search and export](#)

**Validated analysis manager**

Validated terms from analysis *undefined* (18567)

Id	Name
<input type="checkbox"/>	<a href="#">headache</a>
<input type="checkbox"/>	<a href="#">sinus</a>
<input type="checkbox"/>	<a href="#">diagnosis</a>
<input type="checkbox"/>	<a href="#">nausea</a>
<input type="checkbox"/>	<a href="#">pain</a>
<input type="checkbox"/>	<a href="#">syndrome</a>
<input type="checkbox"/>	<a href="#">aura</a>

Create entries | Select all

Delete selection | Delete all

UNIVERSITAT POMPEU FABRA iulTerm

## 7. Glossaries module

Terminus 2.0

Home User's guide User details Projects Sources Contact us Exit

User: Natalia Seghezzi  
Group: Workshop1

**Conceptual structure**  
Concept tree

**Documents**  
Upload a document  
Upload a group of documents  
File manager

**Corpus**  
New corpus  
Corpus manager

**Analysis**  
New analysis  
Saved analysis manager  
Validated analysis manager

**Glossaries**  
Glossary manager  
Import a glossary

**Terms**  
New term  
Term search and export

Terminus is a web application for corpus and terminology management. It allows creating and managing terminological projects and selecting data categories. It integrates the complete work process of individual and/or team terminographic work: textual corpus search, compilation and analysis, term extraction, glossary and project management, database creation and maintenance, and dictionary edition. Terminus contains different articulated modules:

- **Projects:** for creating a terminological project.
- **Sources:** for managing the sources used in a terminological project.
- **Conceptual structure:** for creating a concept tree in order to structure the terms of a given glossary.
- **Documents:** for uploading text files which will constitute the working corpus.
- **Corpus:** for selecting documents to build a corpus.
- **Corpus analysis:** for statistical corpus exploitation. Word frequency, concordances, n-grams and association measures can be extracted. It also for term extraction from LSP corpus.
- **Glossaries:** for creating and managing the glossaries of a project.
- **Terms:** for introducing terminological data into a glossary, retrieve them and export them.

UNIVERSITAT POMPEU FABRA iuliaTerm

This module allows managing the glossaries created by the user, i.e. edit them (change their name or description) or delete them, and creating new glossaries. Users may have an unrestricted number of glossaries, but must create at least one in order to be able to use the *Terms* module and introduce new terms into the glossary afterwards.

To create a new glossary click on *Glossary manager*. This option also allows to see the list of previously created glossaries. When a new glossary is created it must be given a name; besides, the main language, the domain field and the project it is ascribed to must be indicated.

Terminus 2.0

Home User's guide User details Projects Sources Contact us Exit

User: Natalia Seghezzi  
Group: Workshop1

**Conceptual structure**  
Concept tree

**Documents**  
Upload a document  
Upload a group of documents  
File manager

**Corpus**  
New corpus  
Corpus manager

**Analysis**  
New analysis  
Saved analysis manager  
Validated analysis manager

**Glossaries**  
Glossary manager  
Import a glossary

**Terms**  
New term  
Term search and export

**Glossary manager**

Options	Glossary name	Language	Domain field	Project	Terms
<input type="checkbox"/>	Headaches	English	Medicine	Headaches 0	

To create a new glossary, fill-in this form (fields marked with "\*" are mandatory).

Glossary name: \*

Language: \*

Domain field: \*

Project: \*

Description:

Ok Delete selection Delete all Reset form

UNIVERSITAT POMPEU FABRA iuliaTerm



From this module, glossaries complying with certain requirements (XML format, etc.) can also be imported<sup>3</sup>.

## 8. Terms module

The screenshot displays the Terminus 2.0 web application interface. At the top, the logo 'erminus 2.0' is visible, along with navigation links: Home, User's guide, User details, Projects, Sources, Contact us, and Exit. The user profile section on the left identifies the user as Natalia Seghezzi, Group: Workshop1. Below this, there are several menu categories: Conceptual structure (with a link to Concept tree), Documents (with links to Upload a document and Upload a group of documents), Corpus (with links to New corpus and Corpus manager), Analysis (with links to New analysis, Saved analysis manager, and Validated analysis manager), and Glossaries (with a link to Glossary manager). A green arrow points to the 'Terms' module in the navigation menu, which includes sub-options for 'New term' and 'Term search and export'. The main content area on the right provides a description of Terminus as a web application for corpus and terminology management, listing its various articulated modules: Projects, Sources, Conceptual structure, Documents, Corpus, Corpus analysis, Glossaries, and Terms. The footer of the page features the logos for Universitat Pompeu Fabra and iulaterm.

The *Terms* module has two main options: introducing new terms into a glossary, and searching for terms which have already been saved and exporting them to an already created glossary.

*New term*: In order to introduce new terms into a glossary, a series of steps have to be followed. The terminological record you will find is exactly the same as that for introducing terms from a previous validated analysis. (see Section 6).

The terminological record is divided into several parts to be completed with the term-related information: some basic information (the glossary to be incorporated in, lemma, grammatical category, source, etc.), and other optional data categories (contexts, definitions, equivalents, term variants, collocations and other fields). It is possible to introduce as many contexts, equivalents, term variants, collocations and linguistic notes as necessary or desired.

---

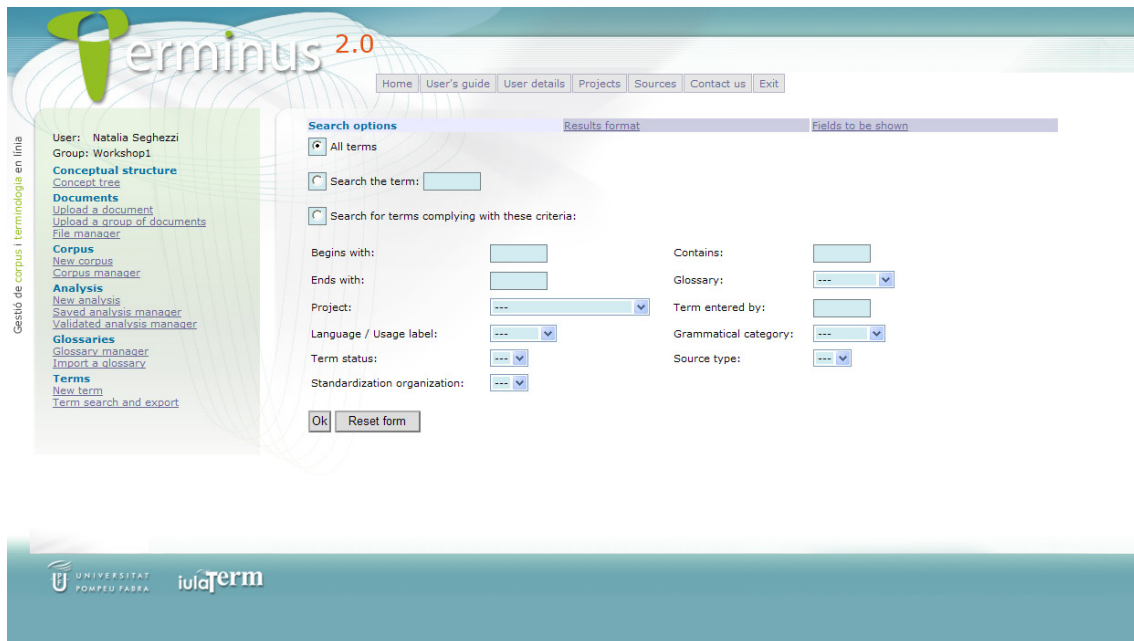
<sup>3</sup> For further information please contact [iulaterm@upf.edu](mailto:iulaterm@upf.edu).

The screenshot displays the 'erminius 2.0' web application interface. At the top, there is a navigation bar with links for Home, User's guide, User details, Projects, Sources, and Exit. Below this, a user profile section identifies the user as Natalia Seghezzi from the Workshop1 group. A sidebar on the left provides access to various tools such as Conceptual structure, Documents, Corpus, Analysis, Glossaries, and Terms. The main content area is titled 'Basic information' and contains a form with several fields: Lemma (text input), Grammatical category (dropdown), Term source (text input with a dropdown), Source type (dropdown), Subdomain (dropdown), Term status (dropdown), Standardization organization (dropdown), and Usage label (dropdown). Fields marked with an asterisk (\*) are compulsory. At the bottom of the form are two buttons: 'Create record' and 'Continue editing'. The footer of the page features the logos of the Universitat Pompeu Fabra and iulqterm.

Once the compulsory data categories (i.e. those marked with an asterisk) have been introduced (lemma, grammatical category, source and glossary), the terminological record can be created by clicking on *Create record*. Apart from the compulsory data categories, other term-related information can be introduced before creating the record, with the option *Continue editing* or by clicking on each of the screen tabs in the upper part of the screen (contexts, definitions, equivalents, term variants, collocations and other fields). You can click on *Create record* at any time. All created records are saved into the selected glossary and can be consulted and edited at any time.

*Term search and export*: Term searches can be made on different criteria and there are several data representation, visualization and exportation options.

First of all, the search criteria (*Search options*) must be indicated. Second, the representation format (*Results format*) must be chosen. Third, the data categories for each register have to be selected (*Fields to be shown*).



The term search is useful to edit the terminological records of the glossaries. As indicated, in order to be able to edit the term-related data, the *Search options* must be included in the fields to be shown.

As regard term exportation there are four main possibilities: to display the results on the browser, or to export the results in PDF, HTML, TXT or XML format. The last two options are useful to export the terminological data to other systems.