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 an 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. **Conceptual structure** module ..................................................5
4. **Documents** module .............................................................8
5. **Corpus** module ...................................................................12
6. **Analysis** module ...............................................................14
7. **Glossaries** module .............................................................24
8. **Terms** 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.

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

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.

**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 (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.

- **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.
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).

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

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.

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.

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 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).
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.

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

1 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 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.

Further information about part of speech tagging in different languages can be found on the following webs:
http://courses.washington.edu/hypertxt/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 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.
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).
7. Glossaries module

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.
From this module, glossaries complying with certain requirements (XML format, etc.) can also be imported.

8. Terms module

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.

---

3 For further information please contact iulaterm@upf.edu.
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.