



ALterm HANDBOOK

Miran Željko

Contents of this document

| | |
|---|----|
| ABOUT ALterm..... | 2 |
| HOW TO USE ALterm..... | 4 |
| UPDATING ALterm DATABASE..... | 7 |
| DESCRIPTION OF ALterm WEBSITE..... | 9 |
| FOLDER STRUCTURE..... | 9 |
| FILES IN THE HTML FOLDER..... | 10 |
| FILES IN THE CGI-BIN FOLDER..... | 13 |
| DESCRIPTION OF MySQL DATABASE STRUCTURE..... | 14 |
| SUMMARY OF SERVERS, LOCATIONS, USERNAMES AND PASSWORDS..... | 16 |
| TROUBLESHOOTING..... | 17 |



ABOUT ALterm

ALterm (<http://alterm.mie.gov.al> and <http://albania.anyterm.info> on the backup server) is a multilingual terminology database developed during the process of translating the EU acquis in the Albanian language serving as an important translation tool. The purpose of this program is to have a unified and consistent terminology database for a range of fields, serving as a reliable source of information and reference for translators, institutions working to harmonise the Albanian legislation with the acquis, universities, organizations, as well as the general public.

The process of translating the acquis consists not only of translation, but also expert, legal and linguistic review. In order to achieve as qualitative a translation as possible, each term goes through all the review and evaluation phases before it is added to the terminology database. Thus, the level of review determines the reliability of the term.

The acquis covers a wide range of areas, which means that terms are from many different contexts and fields of science and human activity. This terminology database provides basic words, legal terminology, terms from different fields and phrases/compound words used regularly.

The majority of terms are accompanied by information on the specific field it belongs to. Occasionally, the same term may have different meanings in two or more fields. Such terms are treated separately by referring to its new meaning. Thus, the same term may have two or more synonyms in a specific field.

Because EU legal acts are translated mainly from English, each new term consists of the term in English, its translation in Albanian and corresponding translations in the other languages where available (French, German, Italian or Spanish), for comparison and to understand them better. In addition to the term in two or more languages, an entry may also contain:

| | |
|-------------|--|
| Created on | Date the term was added to the termbase |
| Modified on | Date the term was last updated |
| Domain | Field the term belongs to |
| Acquis | Acquis chapter |
| CELEX | CELEX No. |
| Context | Original text extract with the term in it |
| Definition | Definition of the term in the given field |
| Grammar | The grammatical status of the term |
| Note | Notes, comments and additional information |



| | |
|-----------|---|
| Reference | The original term reference or source |
| Status | Term status (preferred, not preferred, to be reviewed, outdated, forbidden, etc.) |

As one may see, ALterm is a terminology database unlike internet dictionaries and it is linked strongly with the process of translating the acquis. As the process progresses, so shall the termbase be updated and developed.

In order to have a streamlined and consistent process of maintaining and updating the termbase, you are free to leave any suggestions, recommendations, comments and questions to Ms. Mikela Kakeli (mikela.kakeli@mie.gov.al).

The establishment of this database was commissioned in frame of the project "Support to Albanian Ministry of European Integration" (SMEI II; see: <http://smei.mie.gov.al/en/pages/Home>). This project is funded by the European Union (IPA 2008) and the Government of the Federal Republic of Germany (BMZ) and is implemented by Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH.



HOW TO USE ALterm

A search in the ALterm terminology database (<http://alterm.mie.gov.al> and <http://albania.anyterm.info> on the backup server) can be made by entering the search word(s) in the entry field on the left side of the ALterm's home page. When making a search, the following wildcards can be used:

- **_ . ?** (underline, dot or question mark) can substitute for any single character; e.g., if one wants to get all the hits containing the word *organisation* and *organization* with a single query, then the term **organi_ation** (or **organi.ation** or **organi?ation**) can be entered into the entry field.
- **% *** (percent sign or asterisk) can substitute for any number of characters; e.g., if one wants to find hits in which there are additional words between *illicit* and *drug* then the search query should be written as *illicit%drug* (or *illicit*drug*); in addition to usual word combinations with *illicit drug*, terms such as *illicit trafficking in narcotic drugs* are also obtained.

One can enter the search query (in any of the following languages: Albanian, English, French, German, Italian or Spanish) and the program will display the list of hits. Another possibility is to enter the "#" character followed by a number – in this case data about this entry number will be shown in the central part of the screen (e.g., by entering "#1" (without quotes) it is possible to see the first entry in the database.). This type of search is useful for a terminologist.

By clicking at the selected term, one can see a translation of this term into English, French, German, Italian and Spanish (where the terms in these languages are available). It is easier to compare similar terms by using this function. By clicking "Expand all" the software will display all translations of all terms on the list.

A click on or "Collapse all" has the opposite effect: by clicking at the selected term, the translations of this term will no longer be visible. By clicking "Collapse all", all translations of all terms on the list will be hidden.

A click on a particular hit will display all the data about this word/expression on the central and right side of the screen.

If the user wants to limit the search according to the position of the search word within possible terms, an appropriate wildcard can be used. Suppose you want to find word combinations that contain the word *company*:

- If only a translation of the word *company* is wanted, then enter **company\$** in the entry field (**\$** sign denotes the end of a word).



- If word combinations that start with the word *company* are wanted, enter this word only; in addition to *company* the hit list will contain *company accounts*, *company being acquired* etc.
- If all word combinations that end with the word *company* are wanted, enter ***company\$** in the entry field; the hit list will also contain *acquired company*, *acquiring company* etc. (* denotes any number of any character(s)).
- If all terms that contain the word *company* are wanted, enter ***company** in the entry field; in addition to the previously mentioned terms, the hit list will also contain *abusive company behaviour* etc.

It is possible to use a combination of wildcards, e.g., the search query can be:
***organi?ation\$**.

If the search word is not found, the program will switch to fuzzy search and will list the terms similar to the search term.

If you want to check which terms in a complete document exist in ALterm, you can use the **TermALog** program (terminology analyser: <http://alterm.mie.gov.al/t/> or <http://albania.anyterm.info/t/> on the backup server).

A word, a sequence of words or any text can be entered in the TermALog entry field. The text can be written in English or Albanian. A complete file can be transferred to the TermALog window by using the "copy and paste" commands — e. g., in Microsoft Word: first select the whole document (Ctrl A), then copy the selected text (Ctrl C), switch to the browser window, click the TermALog entry field and paste the text (Ctrl V).

When you select the language and press the Process key, the program will analyse the text and mark those parts of the text ("terms") that exist in the ALterm terminology database. These terms will be transformed to hypertext links and, by clicking a particular link, detailed information on this term can be obtained.

Fast search during translation

[Intelli Web Search](http://www.intelliwebsearch.com) (IWS: <http://www.intelliwebsearch.com>) is a free program that allows shortcuts to be made to frequently used search sites (e.g. Google, One-look dictionaries — as well as ALterm and TermALog). IWS also allows you to search local resources (dictionaries installed on your PC). After installing IWS (this is a very simple operation), it is necessary to enter the following strings in the program settings (IWS. ver. 3):



| | |
|---------------|---|
| Start: | http://alterm.mie.gov.al/index.php?beseda= |
| Finish: | |
| Shortcut key: | Any combination, e.g.: Ctrl + Shift + Alt + A (press and hold down the following keys: Ctrl Shift Alt A) |

ALterm can now be used much faster in any MS Windows software: just select the search string, press the appropriate combination of keys on the keyboard or select the desired website from the IWS list. If you select parallel ("Pluri") search, IWS launches any number of searches (almost) simultaneously, which considerably improves search speed.

Feel free to send any comments about the contents of the terminology database, translations, terminology, etc. to the terminologist [Imirjon Peza](mailto:imirjon.peza@mie.gov.al) (imirjon.peza@mie.gov.al), while comments about the program should be directed to the author of the program [Miran Zeljko](mailto:miran.zeljko@gmail.com) (miran.zeljko@gmail.com).



UPDATING ALterm DATABASE

Currently, MIE terminology data are stored in the Trados MultiTerm database. These data are accessible only to the people on the MIE local area network that have Trados MultiTerm software installed. In order to provide access to freelance translators and other users (government employees working in EU-related fields, journalists, students and general public interested in this subject) it is necessary to transfer the data to the web server.

Since July 2011 this terminology database (termbase) is available at <http://alterm.mie.gov.al> website. In order to update the terminology data on this website the terminologist must do the following:

1. export all terms from the MultiTerm database,
2. send this file to the alterm.mie.gov.al ftp site,
3. start the procedure for processing the data.

Detailed description of the procedure:

1. Open the termbase in MultiTerm and export all entries in XML format. The export data should be stored to the **enalterm.xml** file (file name is case sensitive!).
2. The simplest way to send this file to the server is to use Windows Explorer, but you can use any other ftp software (e.g. FireFTP in Mozilla, Filezilla or WinSCP to name just a few).
The address of the server is: <ftp://10.10.201.58>
There should be no tick mark in the "Anonymous login" box.
Username is **mie**
Password is **altermpass**
Username and password are case sensitive. You can tick the "Remember password" box so that you do not need to enter username and password every time you connect to the server.
After that you simply copy enalterm.xml file from your PC to the ftp server in the same way as you copy a file from one folder to another.
3. After that you should manually start the updating procedure: enter the following address in the web browser:
<http://alterm.mie.gov.al/cgi-bin/process.sh>
and press Enter. Terminology data processing is completed when the line "Processing completed" message is displayed.

Backup ALterm server is located at <http://albania.anyterm.info>. It will be fully functional in the future, but it can act as a backup server only if the data are updated on that server as well. The procedure for updating the termbase on this server is the same as the one described above, only the contact information is different:



REPUBLIC OF ALBANIA
MINISTRY OF EUROPEAN INTEGRATION

Support to the Albanian Ministry of
European Integration
(SMEI II)



The address of the server is: **ftp://alterm@anyterm.info**

Username is **alterm@anyterm.info**

Password is **t1e2r3m4**

After that you should manually start the updating procedure: enter the following address in the web browser:

<http://albania.anyterm.info/cgi-bin/process.sh>

and press Enter. Terminology data processing is completed when the line "Processing completed" message is displayed.



DESCRIPTION OF ALterm WEBSITE

FOLDER STRUCTURE

The complete folder structure of the website is:

```
/var
 |----- www
         |----- cgi-bin
         |             |----- ftp
         |----- html
                 |----- graphic
                 |----- t
```

The basic folder of the website is /var/www. Within this folder the following subfolders exist:

```
    cgi-bin
    html
```

The **cgi-bin folder** contains programs for conversion of data from MultiTerm XML format to the SQL format used in the ALterm website database.

Within this folder is **ftp** folder. When the terminologist connects to the ftp server, he is located at this folder and the enalterm.xml file that he copies from his PC to this server is stored at this location.

The **html folder** contains the ALterm website data (php, css, js and gif files).

Within this folder are 2 subfolders: the **graphic** folder contains the graphic files used in the terminology database, and the **t** folder contains the TermALog files.

It is important that the permissions of the folders remain the same as they were at the setup of the system: 2 folders have rwx permissions for everybody: ftp and graphic. This was the only way to allow the program that processes the input data to move data between the folders (the httpd process (which moves the data) is not the owner of the data).



FILES IN THE HTML FOLDER

The following files should be in the "html" folder (currently the full path of this folder is: /var/www/html):

| FILENAME | FUNCTION |
|------------------|--|
| index.php | The main file shown if website address is entered without a filename. It joins other files and defines their layout on the screen. |
| svez_slovar1.php | Upper part of the screen (above the horizontal line). It displays the logo, the title of the website and provides links to TermALog, description of the terminology database, instructions for use and allows switching between Albanian and English user interface. |
| svez_slovar2.php | Lower part of the screen (below the horizontal line). It creates 2 frames and initially displays svez_slovar3.php on the left side and evroterm.php on the right side. |
| svez_slovar3.php | Left part of the screen. It displays the basic data about the terminology database (number of entries and the date of the last update) and allows the user to enter a search query. After the search it displays the basic results (a list of clickable hits). |
| svez_slovar5.php | Right part of the screen. When the user clicks a link on the list on the left side of the screen, this file creates the resulting output. |
| evroterm.php | Description of terminology database project. |
| navodila.php | Instructions for using ALterm (search options). |
| entries | |



| | |
|-----------------|---|
| | Number of entries in the database. Updated after the termbase administrator sends a file to the server and starts the updating procedure. |
| timer_start.php | A script for starting the timer (at the beginning of a search query). |
| timer_stop.php | A script for stopping the timer (at the end of a search query). |

The address of ALterm on the web: <http://alterm.mie.gov.al>

The address of ALterm on the MIE intranet: <ftp://10.10.201.58>

The address of the backup ALterm server on the web: <http://albania.anyterm.info>

A source file for the terminology database is a file exported from the MultiTerm database in XML format. The name of this file should always be enalterm.xml and it can be zipped and stored as enalterm.zip (names of files are case sensitive and if they differ from these, they will not be recognised by the "process.sh" script).

After the termbase administrator sends enalterm.txt or enalterm.zip file to the server, (s)he should start the <http://alterm.mie.gov.al/cgi-bin/process.sh> script (on the backup server this script is located at: <http://albania.anyterm.info/cgi-bin/process.sh>). This script has many comments in it so that it is easier to check its operation.



TermALog has similar (but slightly simpler) structure. The following files should be in the "/var/www/html/t" folder:

| FILENAME | FUNCTION |
|--------------|--|
| index.php | The main file shown if website address is entered without a filename. It joins other files and defines their layout on the screen. |
| t1.php | Upper part of the screen (above the horizontal line). It displays the logo and the title of the website and provides links to ALterm, instructions for use and allows switching between Albanian and English user interface. |
| t2.php | The core of the terminology analyser program. |
| navodila.php | Instructions for using TermALog. |
| EN | A list of English terms sorted by the term length (used by t2.php). This file is updated after the termbase administrator sends a file to the server and starts the updating procedure. |
| SQ | A list of Albanian terms sorted by the term length (used by t2.php). This file is updated after the termbase administrator sends a file to the server and starts the updating procedure. |

The address of TermALog on the web: <http://alterm.mie.gov.al/t>

The address of TermALog on the MIE intranet: <ftp://10.10.201.58/t>

The address of TermALog on the backup web server: <http://albania.anyterm.info/t>



FILES IN THE CGI-BIN FOLDER

The following files should be in the "cgi-bin" folder (currently the full path of this folder is: /var/www/cgi-bin):

| FILENAME | FUNCTION |
|----------------|--|
| process.sh | <p>The main script for processing the data sent by the terminology database administrator. The script has the following main functions:</p> <ul style="list-style-type: none">• it transforms the MultiTerm export file to SQL format• it deletes the data from the database (as a matter of fact – it renames the tables by appending the string "old" to their names• it stores the data into the MySQL database• it creates EN and SQ files for terminology analyser. <p>At the beginning of the script the system-dependent variables are defined (folders, MySQL host name, database name, username, password). There are many comments in the script so that it is easier to understand the script logic.</p> |
| proces1.php | <p>This script transforms header data from MultiTerm export file to a series of "insert into" SQL statements.</p> |
| proces2.php | <p>This script transforms terms and term-related data in various languages from MultiTerm export file to a series of "insert into" SQL statements. The script also converts web addresses and Celex numbers to clickable links.</p> |
| sort_by_len.pl | <p>This Perl script sorts a list of terms by the length of terms: the longest term is in the first line, the shortest term is in the last line.</p> |



DESCRIPTION OF MySQL DATABASE STRUCTURE

Name of the database is mirze_alendb. The database must be created manually only once during setup, while the tables in the database are created automatically during each update.

Currently, this database contains the following tables:

- header data table (**enalhead**),
- 7 language tables (**alb, eng, fre, ger, ita, lat, spa**),
- the tables with the same name that have the string "old" added at the end of their names are the tables created during the previous update; they can be used for restoring tables to sensible values if something went wrong during update.

Structure of the header table (**enalhead**):

| Field | Type | Key | Default |
|--------------|--------------|-----|---------|
| ID | mediumint(9) | MUL | NULL |
| CreationDate | varchar(10) | | NULL |
| ChangeDate | varchar(10) | | NULL |
| Subject | varchar(255) | | NULL |
| Multimedia | text | | NULL |
| Note1 | text | | NULL |

Structure of language tables (**alb, eng, fre, ger, ita, lat, spa**):

| Field | Type | Key | Default |
|--------|--------------|-----|---------|
| ID | mediumint(9) | MUL | NULL |
| Jezik | varchar(255) | MUL | NULL |
| Sjezik | varchar(9) | MUL | NULL |



| | | | |
|------------|--------------|-----|------|
| Lema | varchar(255) | MUL | NULL |
| Grammar | varchar(10) | | NULL |
| Status | varchar(20) | | NULL |
| Reference | varchar(255) | | NULL |
| Note | text | | NULL |
| CELEX | varchar(255) | | NULL |
| Definition | text | | NULL |
| Acquis | varchar(255) | | NULL |
| Context | text | | NULL |
| Korpus | tinyint(4) | | 0 |

The data in the tables are those that were defined by the terminologist in the SDL Trados MultiTerm software. Additional fields are the following:

- Sjezik – fuzzy search value of the term; used by fuzzy search
- Lema – lemmatised form of the term (currently not used but may be useful in the future if terms in the termbase are lemmatised)
- Korpus – a flag with 2 possible values:
 - 1 – the term exists in the corpus,
 - 0 – the term does not exist in the corpus.

Initially all values are set to 0. Currently this field is not used – it will be used when translation memory is converted to a corpus database and stored on the web.

The logic of combining terms in various languages is simple: the terms that have the same ID number are corresponding translations in various languages (or synonyms or abbreviations within one language).



SUMMARY OF SERVERS, LOCATIONS, USERNAMES AND PASSWORDS

Name of the ALterm web server: `http://alterm.mie.gov.al`
IP address in the internal network: `10.10.201.58`.
Internet IP address: `95.107.174.158`.

Name of the ALterm backup web server: `http://albania.anyterm.info`

Root password on the operating system level: `a4l7t2`

It seems that root cannot connect to the server via ftp. For this reason another user was created: its username is `alterm` and its password is `t1e2r3m4`

Username used by the terminologist for updating the termbase: `mie`
Password of mie user: `altermpass`

Username for updating the termbase on the backup server: `alterm`
Password of alterm user (on the backup server): `t1e2r3m4`

Root password of the MySQL database: `e0r1m1`
Hostname used for accessing the database: `localhost`
Database name for ALterm: `mirze_alendb`

PHP user of the MySQL database: `mirze_alenweb`
Password of the PHP user of the MySQL database: `w3e2b1`

Location of the ALterm website on the server: `/var/www/html`
Location of the cgi-bin files on the server: `/var/www/cgi-bin`



TROUBLESHOOTING

I suggest that first a backup of the file system below `/var/www` is made. This backup can be used in case of a disk crash of the server or if the ALterm website files get somehow corrupted while the server itself still runs properly. A better solution is to make a current image of the disk to another medium. If a disk crash occurs, this image can be used to recover the system. This must be done on the NAIS location (National Agency for Information Society).

If the ALterm website (<http://alterm.mie.gov.al>) is not available, check the following:

- On the PC click the Start button => Run => enter `cmd` and press Enter. In the window that opens now, enter the following command:

```
nslookup alterm.mie.gov.al
```

If you get a valid public IP address in response, the DNS functions correctly, otherwise contact people from NAIS.

- If ALterm server responds to nslookup command, enter the following command using the same window as in the previous command:

```
ping alterm.mie.gov.al
```

If you get a response from the server, it means that it is running, but maybe the http server (web server) is not running.

- To test this, you must first check whether the server accessible via telnet or remote shell. Use putty.exe program to test this (connect as root). If not, check with the people from NAIS.
- If you can make telnet or remote shell connection to the server using the root account, check if httpd process is running; enter the following command:

```
service httpd status
```

The server should respond with something like this:

```
httpd (pid xxxx) is running
```

If the service is not running, you can start it with the following command:

```
service httpd start
```



If the process does not start, contact the people from NAIS.

- Even if httpd process is running, it is possible that it has somehow got stucked in a loop. In such a case it may help if the process is restarted by entering the following command:

```
service httpd restart
```

If this does not help, go to the next step.

- If the web server is running (you can see the home page), but the search produces no results, check if mysqld process is running; enter the following command:

```
service mysqld status
```

The server should respond with something like this:

```
mysqld (pid xxxx) is running
```

If the service is not running, you can start it with the following command:

```
service mysqld start
```

If the process does not start, contact the people from NAIS.

- Even if mysqld process is running, it is possible that it has somehow got stucked in a loop. In such a case it may help if the process is restarted by entering the following command:

```
service mysqld restart
```

If this does not help, go to the next step.

- If the mysqld process is running but you still cannot obtain any results when making a search, send enalterm.xml file from the PC to the server and start the process.sh procedure (for details see the chapter "Updating ALterm database" in this handbook (page 7)).

If the terminologist cannot connect to the server using the ftp protocol, check if the vsftpd process is running by entering the following command:

```
service vsftpd status
```

The server should respond with something like this:



`vsftpd (pid xxxx) is running`

If the service is not running, you can start it with the following command:

```
service vsftpd start
```

If the process does not start, contact the people from NAIS.

If everything that was described above fails, and NAIS people are not accessible, you can try to solve the problem by rebooting the server by entering one of the following commands after making telnet or secure shell connection:

```
shutdown -r 0  
or  
reboot
```