



<b>Project ref. no.</b>	<i>IST-1999-11748</i>
<b>Project acronym</b>	<b>LIMBER</b>
<b>Project full title</b>	<b>Language Independent Metadata Browsing of European Resources</b>

<b>Security (distribution level)</b>	<i>Public (report) – Internal (Demonstrator)</i>
<b>Contractual date of delivery</b>	<i>M19 July 2001</i>
<b>Actual date of delivery</b>	<i>3 December 2001</i>
<b>Deliverable number</b>	<i>D-7.2</i>
<b>Deliverable name</b>	<i>LIMBER Client Implementation Notes</i>
<b>Type</b>	<i>Report &amp; Demonstrator</i>
<b>Status &amp; version</b>	<i>Final</i>
<b>Number of pages</b>	<i>20</i>
<b>WP contributing to the deliverable</b>	<i>WP7</i>
<b>WP / Task responsible</b>	<i>Intrasoft International</i>
<b>Other contributors</b>	<i>NSD</i>
<b>Author(s)</b>	<i>Nikos Giannelos, Jostein Ryssevik</i>
<b>EC Project Officer</b>	<i>Mr. Kimmo Rossi</i>
<b>Keywords</b>	<i>Cross-language information retrieval, multilingual thesaurus, Thesaurus Management System, TMS, Text Classification</i>
<b>Abstract (for dissemination)</b>	The overall LIMBER system enables multilingual data retrieval through a data access system by providing a Thesaurus Management System (TMS) as a basis for translation and searching over a set of metadata. The clients for the end user and administrator to use the TMS, as well as for the end user to use the TMS with the NESSTAR data access system to retrieve data are described in this report.

## 1. Contents

<b>1. CONTENTS .....</b>	<b>2</b>
<b>2. EXECUTIVE SUMMARY .....</b>	<b>3</b>
<b>3. THE TMS END USER CLIENT (INTRASOFT INTERNATIONAL).....</b>	<b>4</b>
3.1 Introduction.....	4
3.2 Installation .....	4
3.3 Using the TMS User Client.....	4
3.3.1 Hierarchy browsing .....	5
3.3.2 Translation .....	6
3.3.3 Star view .....	7
3.3.4 Top Term .....	7
3.4 Using the TMS Administrator Client .....	8
3.4.1 Adding a node.....	8
3.4.2 Deleting a node .....	9
3.4.3 Renaming a node .....	10
3.4.4 Adding a new link.....	10
3.4.5 Deleting a link .....	11
<b>4. THE LOCALISED NESSTAR DATA ACCESS SYSTEM CLIENT (NSD)</b>	<b>12</b>
4.1 Introduction.....	12
4.2 Examples of the localisation of the NESSTAR Interface .....	12
4.3 Conclusion.....	16
<b>5. THE TMS AND THE DATA ACCESS SYSTEM TOGETHER (NSD).....</b>	<b>17</b>
5.1 Introduction.....	17
5.2 Using the complete LIMBER system.....	17
5.3 Conclusion.....	20

## **2. Executive Summary**

The overall LIMBER system enables multilingual data retrieval through a data access system by providing a Thesaurus Management System (TMS) as a basis for translation and searching over a set of metadata.

The end user is be the person retrieving the data who would use the LIMBER client interface described in the first part of the report to navigate the thesaurus and determine search terms.

The LIMBER TMS will require an administrator to be responsible for maintaining and updating the thesaurus who will require their own interface described in the second section.

The end user will use a data access system to retrieve the actual data, and the localised interface to the NESSTAR data access system for English, French, German and Spanish is described in the third section.

The fourth section describes the use of the whole system to access data using both the TMS and the NESSTAR data access system.

### 3. The TMS End User Client (Intrasoft International)

#### 3.1 Introduction

The TMS User client is an application build in JAVA that allows a user to achieve on-line, over-the-web access to the TMS server. In the next few chapters, a short description of the application will be presented.

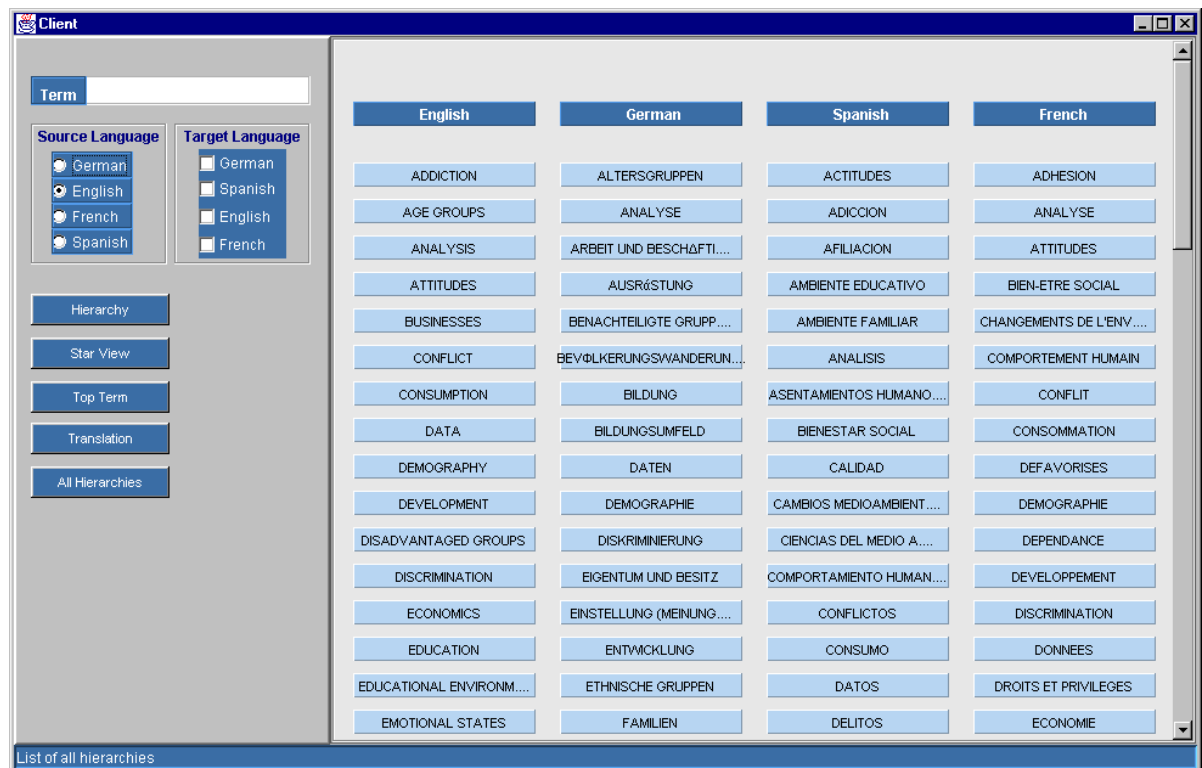
#### 3.2 Installation

The User client is distributed as a collection of files, either compressed in a ZIP archive, or gathered in a directory. In both cases, the installation is performed by simply copying the files to the PC's hard disk. The application may be launched by double clicking on the file named "runit.bat".

Since the application has been developed in JAVA, the user must have JDK 1.3 installed on his computer in order to execute the application. JDK 1.3 is include in the distribution CD, but it can also be downloaded from SUN's web page ([www.sun.com](http://www.sun.com)).

#### 3.3 Using the TMS User Client

As soon as the user launches the application, he is presented with the main screen, which can be seen in the following picture:



The program's screen is divided into two parts (also called panes):

- Control pane: The left part of the screen contains the necessary controls (edit boxes, buttons, lists, etc.) used to access and browse the TMS base

- Info pane: The right part of the screen is used for displaying the information that the user requests.

In general, the user clicks on the term he wishes to get info on and then clicks on the corresponding button. Alternatively, the user may type the term using the keyboard instead of selecting it from the Info pane.

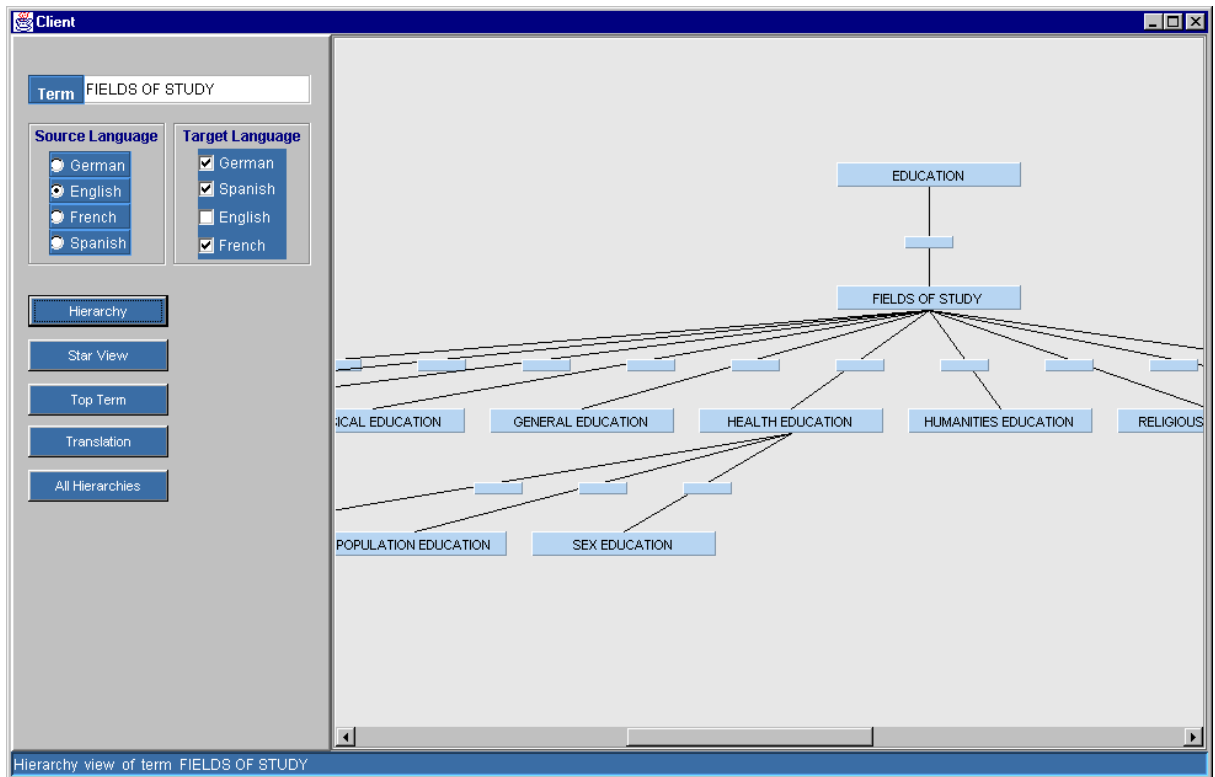
When the program starts, the Info pane presents all the hierarchies that reside in the TMS base, sorted alphabetically and by language. The user may return to this screen at any time by clicking on the button “All Hierarchies”, which is located at the bottom of the Control pane.

### 3.3.1 *Hierarchy browsing*

The user has the opportunity to browse and view the hierarchies that reside in the TMS base. In order to do so, the user has to specify the “depth” of the hierarchy. The corresponding drop-down list is located on the right of the “Hierarchy” button. This setting specifies the number of levels “up” and “down” the hierarchy, that the user wishes to see:



The user has also to specify the term and the source language. After specifying these settings, the user may view the hierarchy around the term he has selected:



### 3.3.2 Translation

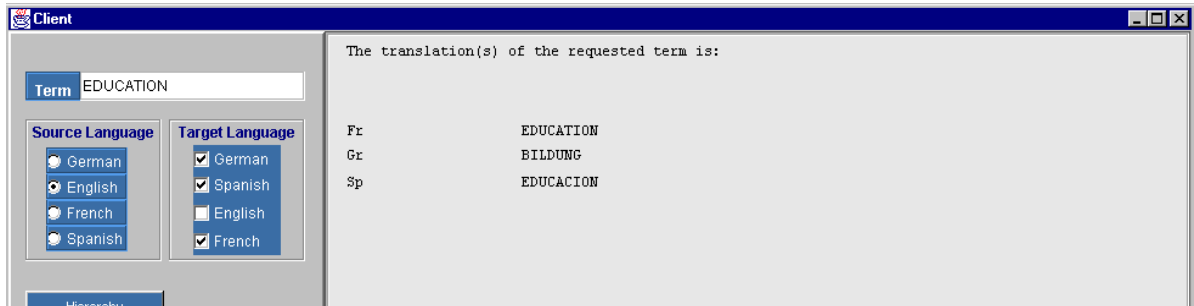
In order to get the translation of a term, the user has to:

- ❑ specify a term by typing it in the edit box, or by clicking on it in the Info pane
- ❑ specify the language the term belongs to, by checking the corresponding check-box as seen in the following picture:



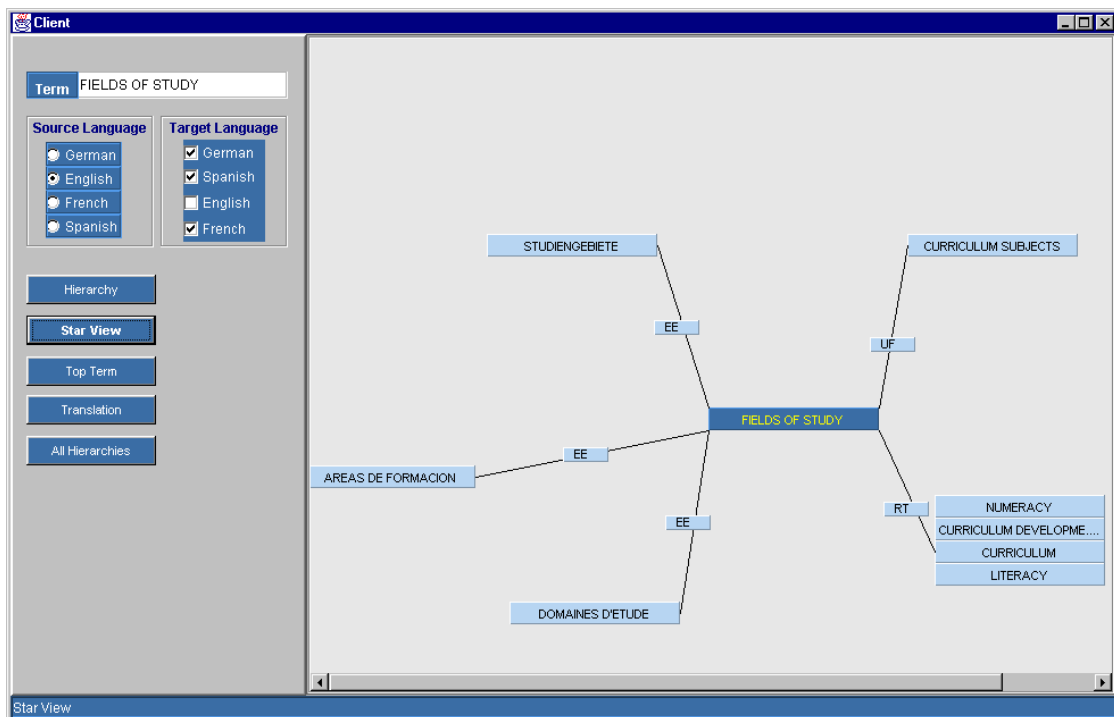
- ❑ request the translation by pressing the “Translation” button.

The translations of the term appear in the screen as text:



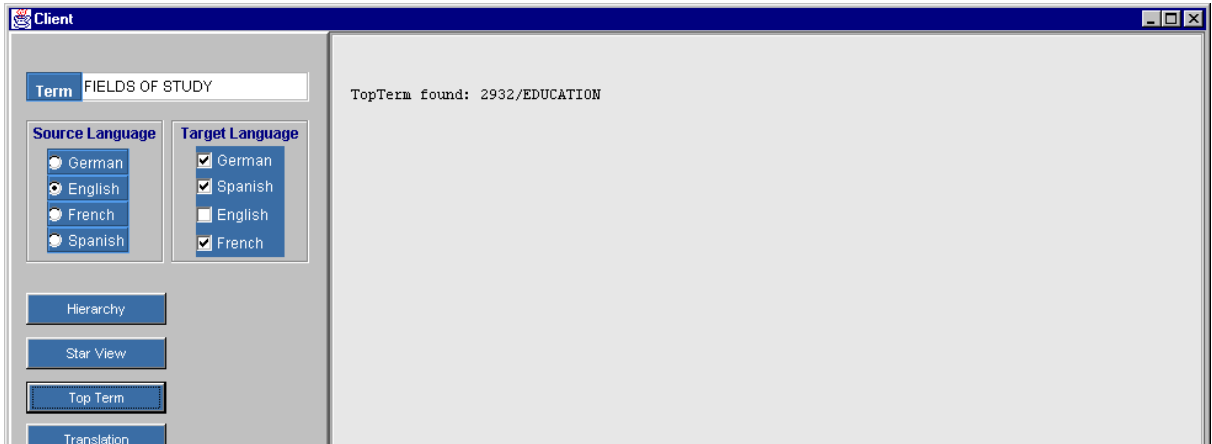
### 3.3.3 Star view

The star view is an option that allows a user to view all relations belonging to a term. After specifying the term and the source language, the user can click the “Star view” button. The following picture presents the star view of the English term “FIELDS OF STUDY”:



### 3.3.4 Top Term

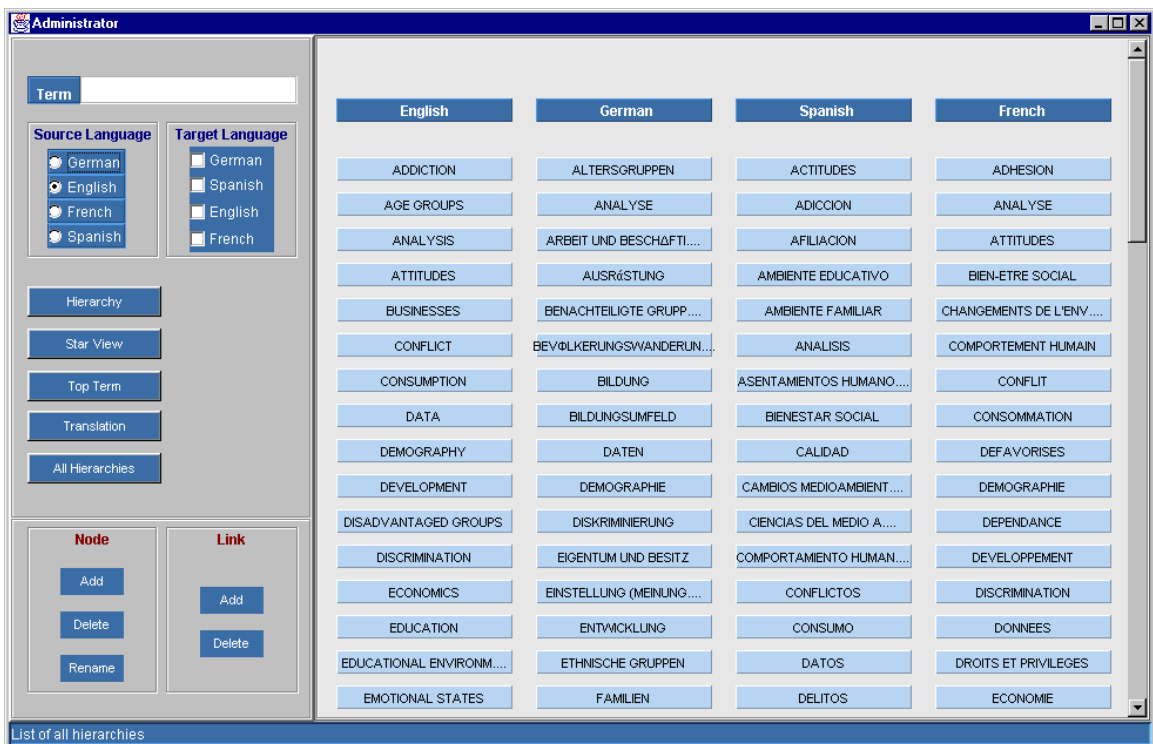
This function returns the term that is located on top of the hierarchy that the current term belongs to. For example, if the user presses the “Top Term” button in the above picture, the system returns the following message:



The number “2932” is the classification code of the term “EDUCATION”.

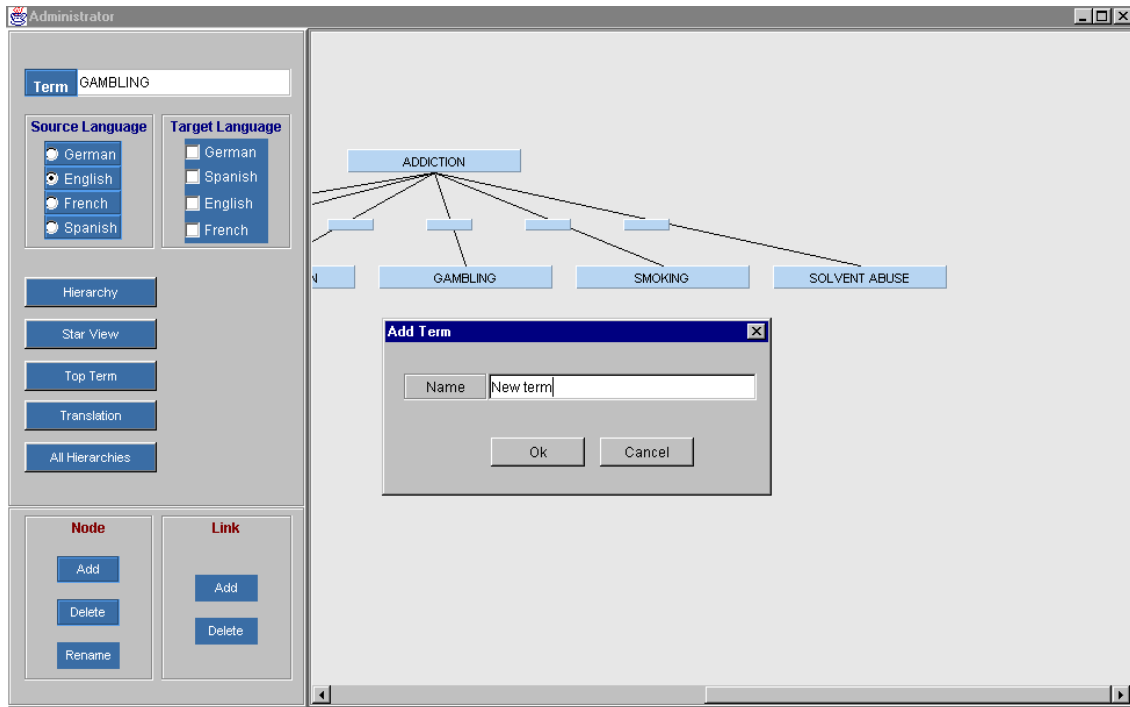
### 3.4 Using the TMS Administrator Client

The difference between the User and Administrator clients is that the Administrator client has the ability to edit the contents of the TMS base. This is performed through the use of the menu that is located on the bottom part of the Control pane:



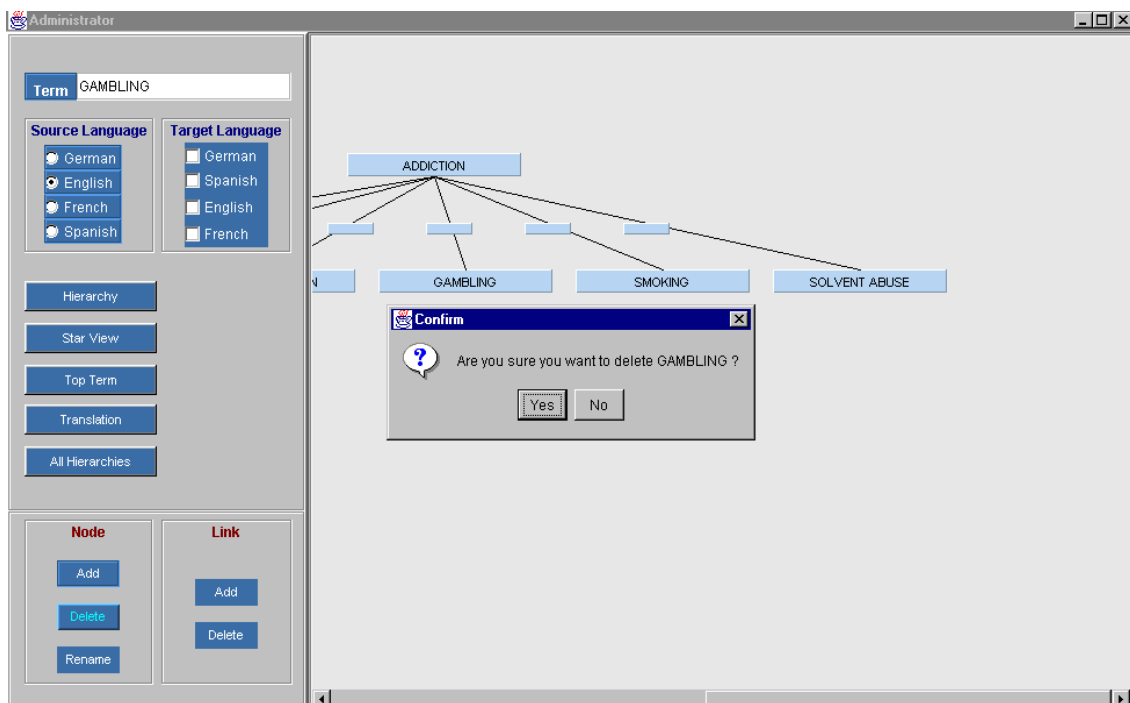
#### 3.4.1 Adding a node

The user may add a new node as a narrower term of an existing one. In order to do so, the user has to provide the broader term and its source language and then press the “Add” button which is located in the “Node” portion of the menu. An edit box appears and the user may specify the name of the new term:



### 3.4.2 Deleting a node

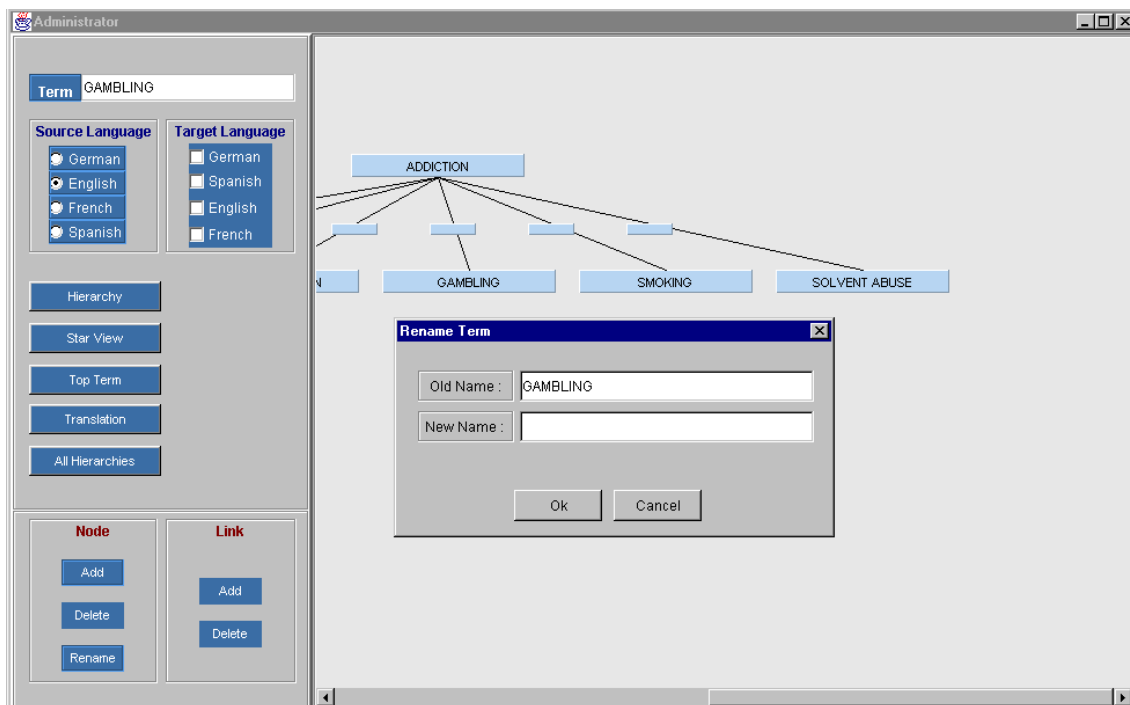
The user can delete a node by selecting it with the mouse or typing it and then pressing the “Delete” button in the “Node” menu. As soon as the user clicks on the button, a warning message appear:



If the user selects “Yes”, the node is deleted from the TMS base.

### 3.4.3 Renaming a node

The user can also rename a node, by clicking the “Rename” button. A dialog box appears and the user is prompted to type the new name of the node:



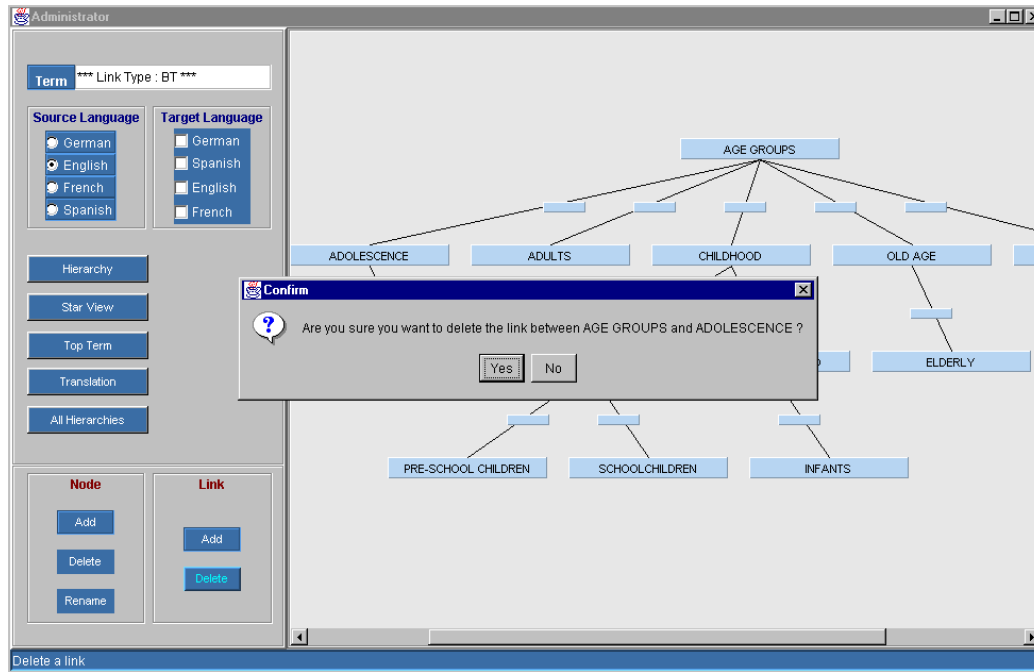
### 3.4.4 Adding a new link

The user can add a new relation between terms (link) by pressing the “Add” button in the “Link” part of the menu. A dialog box appears that requests from the user the node the link starts from, the node it points to and the type of link (UF, RT, etc.).



### 3.4.5 Deleting a link

The user may also delete existing links. In order to do so, the user must first select the link he wishes to delete. This is accomplished by clicking on the small rectangular that resides on the lines connecting the nodes. After the link is selected, the user may delete it by pressing the “Delete” button in the “Link” area of the Control pane:



## 4. The Localised NESSTAR Data Access System Client (NSD)

### 4.1 Introduction

The LIMBER project provided French, German and partially Spanish translations to the user interface of the Nesstar 1.0 client, which was released in July 2000. Nesstar use – and still uses – the localized terms in a system dependent way: it searches for each word in the language of the computer's operating system. If Nesstar cannot find it, it defaults to English, which is supposed to be available.

Since then, Nesstar 1.1 has been released with modifications in the user interface, but no corresponding updates of the French, German and Spanish translations. It was decided not to include the translations with the release, as an English-only program would be better than a partly translated one.

Nesstar 2.0, which is scheduled for 2002, will be released with updated translations. In addition, the user, not the operating system, will decide the language.

### 4.2 Examples of the localisation of the NESSTAR Interface

Just to give an idea, following are the first 20 expressions in the list that constitutes the user interface of the Nesstar client.

#### English:

- Variable
- : this field is mandatory
- Done
- Loading
- Location :
- Session: default
- This dataset contains no data
- Which format?
- About
- About...
- Abstract
- Access
- Activate (On/Off)
- Activation in progress
- Active Agent Slot Full (Remedy:Deactivate some of your Searches)
- Active searches
- Add
- Add all
- Add bookmark
- Add folder

French:

- Variable
- : champ obligatoire
- exécuté
- en cours de chargement
- adresse :
- Session: défaut
- Cette enquête ne contient pas de données
- Quel format voulez-vous?
- A propos de
- A propos de...
- Résumé
- Accès
- Activer (Oui/Non)
- Activation en cours
- Active Agent Slot saturé (Solution:Déactivez quelques-unes de vos recherches)
- Recherches actives
- Ajouter
- Ajouter tout
- Ajouter un signet
- Ajouter un dossier

German:

- Variable
- : Dieses Feld muß ausgefüllt werden
- Erledigt
- Laden
- Ort :
- Session: Vorgabe
- Diese Studie beinhaltet keine Daten.
- Welches Format ?
- Über
- Über...
- Abstrakt
- Zugang
- Aktivieren (An/Aus)
- Aktivierung findet statt.
- Active Agent Slot Full - (Lösung: Deaktivieren Sie einige Ihre Suchen)
- Aktive Suchen
- Hinzufügen
- Alles hinzufügen
- Lesezeichen hinzufügen
- Verzeichnis hinzufügen

Spanish:

- Variable

- : este campo es obligatorio
- Done
- Cargando
- Ubicación :
- Sesión : por defecto
- Este dataset no contiene datos
- ¿En qué formato?
- Acerca de
- Acerca de...
- Resumen/compendio/extracto
- Acceso
- Activar (On/Off)
- Activando
- Active Agent Slot Full (Solución: )
- Búsquedas en proceso
- Añadir
- Añadir todos
- Añadir señalador
- Añadir carpeta

In addition to the user interface of the Nesstar client, also the XML alias files of the codebook were translated. This allows the datasets themselves to be presented using a language of the user's choice. Following are the beginnings of the parallel Alias.xml files:

English:

```
<?xml version="1.0" ?>
- <!--
CodeBook.DTD Aliases Version 1.0 (FINAL)
-->
Error! Hyperlink reference not valid. <definition>
<codebook>Document</codebook>
<docdscr>Document Description</docdscr>
<guide>Guide To Codebook</guide>
<copyright>Copyright</copyright>
<docstatus>Documentation Status</docstatus>
<notes>Notes</notes>
<stdydscr>Study Description</stdydscr>
<citation>Bibliographic Citation</citation>
<stdyinfo>Study Scope</stdyinfo>
<subject>Subject Information</subject>
<topclas>Topic Classification</topclas>
<keyword>List of Keywords</keyword>
<vocab>Controlled Vocabulary</vocab>
<vocaburi>Controlled Vocabulary Location</vocaburi>
```

French:

```

<?xml version="1.0" ?>
- <!--
CodeBook.DTD Aliases Version 1.0 (FINAL)
-->
Error! Hyperlien non valide. <definition>
<codebook>Document</codebook>
<docdscr>Description du document</docdscr>
<guide>Guide pour la documentation</guide>
<copyright>Copyright</copyright>
<docstatus>Statut de la documentation</docstatus>
<notes>Notes</notes>
<stdydscr>Description du Projet</stdydscr>
<citation> Citation bibliographique</citation>
<stdyinfo>Portée du Projet </stdyinfo>
<subject>Contenu</subject>
<topclas>Thématique globale</topclas>
<keyword>Liste de mots-clés</keyword>
<vocab>Vocabulaire contrôlé</vocab>
<vocaburi> Adresse du vocabulaire contrôlé </vocaburi>

```

German:

```

<?xml version="1.0" ?>
- <!--
CodeBook.DTD Aliases Version 1.0 (FINAL)
-->
Fehler! Hyperlink nicht gültig. <definition>
<codebook>Dokument</codebook>
<docdscr>Dokumentenbeschreibung</docdscr>
<guide>Anleitung zum Codebuch</guide>
<copyright>Copyright</copyright>
<docstatus>Dokumentationsstatus</docstatus>
<notes>Anmerkungen</notes>
<stdydscr>Studienbeschreibung</stdydscr>
<citation>Bibliographisches Zitat </citation>
<stdyinfo>Studienumfang</stdyinfo>
<subject>Themeninformationen</subject>
<topclas>Themenklassifizierung</topclas>
<keyword>Liste der Schlüsselwörter </keyword>
<vocab>Kontrolliertes Vokabular </vocab>
<vocaburi>Ort des kontrollierten Vokabulars</vocaburi>

```

Spanish:

```

<?xml version="1.0" ?>
- <!--
CodeBook.DTD Aliases Version 1.0 (FINAL)
-->
Error! Hyperlink reference not valid. <definition>
<codebook>Documento con anotaciones</codebook>

```

<docdscr>Descripción general del documento</docdscr>  
<guide>Guía de términos y definiciones sobre el documento</guide>  
<copyright>Copyright</copyright>  
<docstatus>Informe de progreso sobre la documentación</docstatus>  
<notes>Notas</notes>  
<stdydscr>Ficha técnica del estudio</stdydscr>  
<citation>Cita bibliográfica</citation>  
<stdyinfo>Alcance del estudio</stdyinfo>  
<subject>Información del contenido</subject>  
<topclas>Clasificación temática</topclas>  
<keyword>Lista de palabras clave</keyword>  
<vocab>Terminología estructurada</vocab>  
<vocaburi>Ubicación de la terminología estructurada</vocaburi>

### **4.3 Conclusion**

The tools to translate the Nesstar client software have been readily provided by the LIMBER project. It needs to be updated since the 1.0 version of Nesstar, and maintained during further Nesstar development.

## 5. The TMS and the Data Access System Together (NSD)

### 5.1 Introduction

Support for the TMS multilingual thesaurus has been implemented in Nesstar 2.0, which is scheduled to be released in 2002. In the current implementation a Nesstar client communicates directly with the ELSST thesaurus server in Athens, Greece. The Nesstar server is not involved in the thesaurus support.

### 5.2 Using the complete LIMBER system

The thesaurus provided by the LIMBER project is a multilingual one. Each concept is seen as a single unit of meaning identified by a unique classification code and represented by a set of terms (or words and phrases). Each term has a language attribute (following ISO 639) and a value (literal string). In situations where there are more than one term representing a concept in a single language, one of these will always be classified as the preferred one. The concepts are arranged in hierarchies (trees) which makes it possible to identify broader and narrower concepts.

The thesaurus can support and enrich the searching in many different ways:

1. By allowing the user to browse the thesaurus (and traverse the hierarchies) it can help them to pick the most appropriate search terms (especially when performing keyword searching, but to some extent even when searching in free text).
2. Refining a search proposing the preferred terms
3. Extending a search by including parallel terms, related terms, narrower terms etc.
4. Create a multilingual search from mono-lingual input (by using the concept classification codes and/or the exact terms in different languages)

All search tools include an **open thesaurus** button. By pressing the button the user will open the thesaurus browser (in a separate window) where the concept hierarchies can be displayed and traversed and the relevant terms selected. In addition to traversing the hierarchies looking for the relevant term(s), the thesaurus can also be searched. If a term is highlighted in a search field of one of the search tabs when the thesaurus is opened, the term will be carried over to the thesaurus window and searched for in the hierarchies. The thesaurus can also be searched directly by typing a term into the search field of the thesaurus window.

The search tools also include a **thesaurus options** button that allows the user to set the parameters that drive the behavior of the thesaurus.

In a typical session with Nesstar the user first searches for a list of datasets that comply with a set of search criteria, like datasets that contain some specified terms. The inclusion of TMS in Nesstar allows the user to provide multilingual searches as well as standardizing the search terms and providing related terms.

For any term in the thesaurus TMS provides:

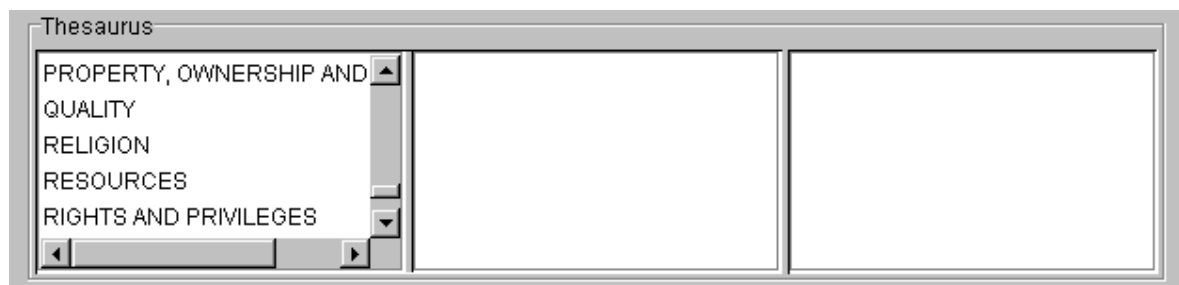
- a list of broader terms
- a list of narrower terms
- a list of related terms
- a list of synonyms that are unpreferred, in the sense that they cannot be used for further thesaurus lookup
- translation to the other languages. English, French, German and Spanish are supported.

In order to look up a term in the thesaurus the user needs to know it is there. Following this requirement, a complete thesaurus browser was needed in Nesstar. The user picks terms from the thesaurus, then performs the search. This **manual** method provides the best user control over the search expressions.

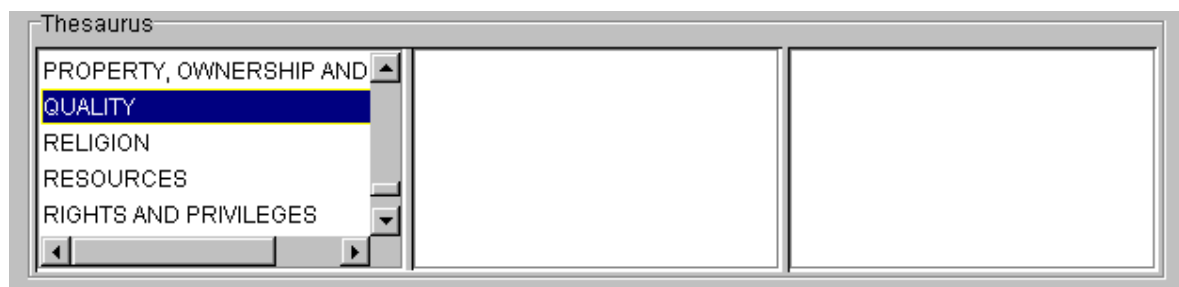
The thesaurus browser is implemented as a three column list of terms where the leftmost column contains the broader terms and the rightmost column contains the narrower terms. A click of the term will select it, shifting the columns if necessary so the selected term appears in the middle column, then show broader terms in the column to the left of the selected term, narrower terms in the column on the right side, and related terms and unpreferred synonyms below the selected term. The selected term can then be picked for insertion into a search expression or translation to other languages.

The thesaurus language can be chosen from any of the supported languages. Initially the left column is filled with all top terms, currently 49, in the chosen language.

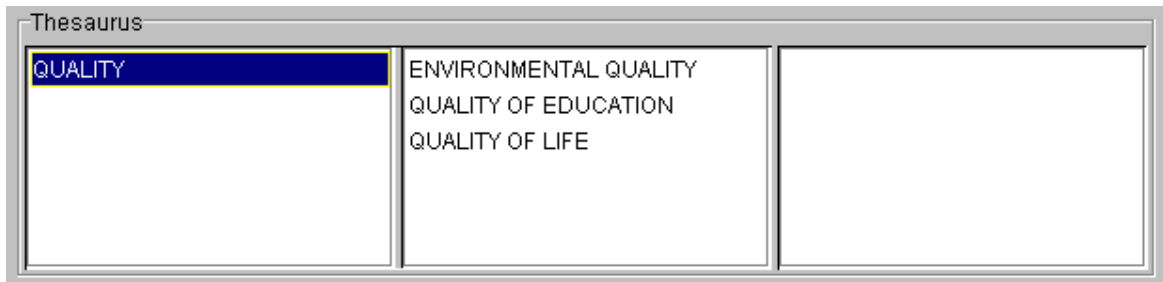
The following illustration gives an idea how the browser works:



*Initially, the 49 top terms are shown.*



*The term QUALITY is selected.*

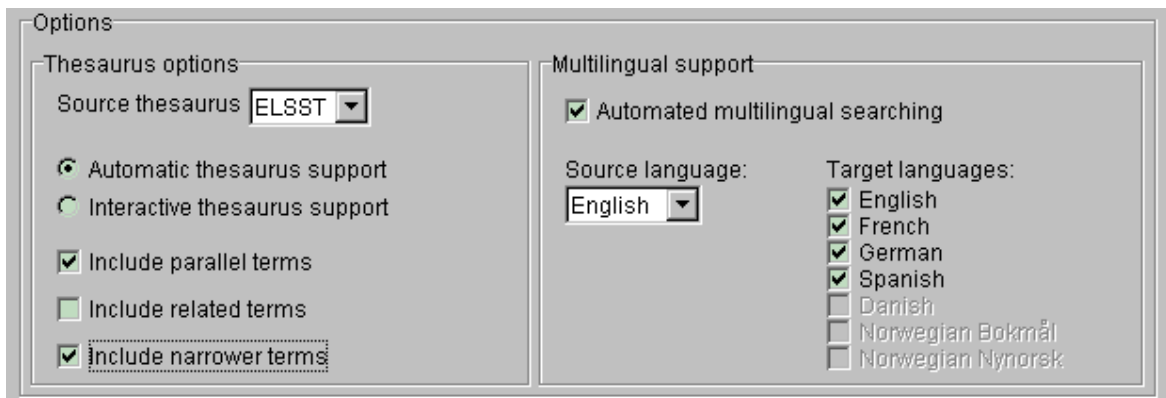


*Three narrower terms are shown in the column right to QUALITY. As a top term, QUALITY has no broader terms, so no right-shifting is necessary. It has no related terms or unpreferred synonyms, either. To continue browsing, select any of the narrower terms or reset the thesaurus to top terms.*

The user can also choose **automatic** expansion of the search expression. The provided search expression is expanded to also include any combination of broader terms, narrower terms, related terms, unpreferred synonyms and translated terms from any of the four thesaurus languages. This is the ideal search method for a user who wants a broad collection of datasets over a certain topic while relaxing a little the control over which search term that was used to provide each dataset.

Also bear in mind that Nesstar offers an extensive number of search options. Specific fields of the datasets can be searched for with different search expressions for different fields. It is especially for such advanced searches with multiple search expressions that automatic expansion can save work for the user.

Even for the automatic use of TMS, the user will probably start by consulting the thesaurus browser for search terms.



*Options for automatic expansion of search expressions.*

The **speed** of the TMS system has been a challenge. Each server request takes normally 5-20 seconds. For this reason all requests are cached locally by the Nesstar client. In addition Nesstar looks ahead and makes silent server requests in background for terms that the user may select. With these techniques the user experiences instantaneous response in most cases while browsing the thesaurus. In future Nesstar versions the cache may be stored on file, which will improve the response time even further.

An additional design challenge has been not to make the thesaurus support too obtrusive. A user who does not want to use the thesaurus tools, should not be bothered with their existence. The current design satisfies this challenge. The thesaurus doesn't come in the way or produce unexpected results for people who just want to make a normal search, nor does it steal an unpleasant amount of process time and memory.

### **5.3 Conclusion**

The LIMBER TMS has proven to be a valuable tool for the Nesstar search engine. TMS improves the accessibility of foreign language data sets and helps finding – and creating – proper search terms. Its full impact to access to statistical data remains to be seen, as Nesstar 2.0 isn't released yet.

It is clear, however, that TMS constitutes a valuable tool and its use will be refined and further developed along with the future development of the Nesstar software itself.