

How to add a thesaurus for search and data entering in an ABCD database

Piet de Keyser

March 2017

Introduction

In this short manual is explained how you can add a thesaurus to your ABCD database in order to use thesaurus terms as search terms or to add them to your record while entering data.

Do not use the ABCD thesaurus database to maintain the thesaurus, because at this moment there is no proper thesaurus maintenance module in ABCD; it would be as difficult as maintaining a thesaurus manually. It is better to use a dedicated thesaurus software (e.g. Thesaurus Builder), preferably with good export facilities, to do so.

You may find the technical details in Spanish on Guilda Ascencio's wiki:

<http://abcdwiki.net/wiki/es/index.php?title=Tesauros>

I wrote this manual because out of two considerations:

1. It is probably difficult to understand all details as explained on the Spanish wiki if you don't understand Spanish (my own knowledge of Spanish is very limited and I experienced this problem at first hand).
2. The field numbering on the Spanish wiki is different from the field numbering in the Agrovoc Thesaurus which comes with the ABCD download. In my experience it is better to stick to the Agrovoc structure, otherwise you will have to change many things all over.

1. Dr_path.def

Look for dr_path.def in main directory of your database directory (if it is not there, create it):

 ayudas	22/06/2016 22:56	Bestandsmap
 cnv	16/06/2016 23:01	Bestandsmap
 data	21/12/2016 22:31	Bestandsmap
 def	23/11/2016 22:06	Bestandsmap
 pfts	19/10/2016 23:12	Bestandsmap
 dr_path.def	2/11/2016 23:08	DEF-bestand

If the dr_path.def file exists it probably contains something like:

```
ROOT=/ABCD/www/bases/marc/dr/
```

Now add these lines to the file dr_path.def:

```
thesaurus=thes
prefix_search_thesaurus=DE_
```

Instead of “thes” you can enter any name of your thesaurus. DE_ is the prefix ABCD is going to use for searching the thesaurus. You can replace it by another prefix if you want to.

Now you should see the thesaurus icon in the toolbar of the data entry: .

2. Preparation of the thesaurus

2.1. Structure of the thesaurus database

The thesaurus is a normal ABCD database with all the normal parts: def, pft, fst etc. The latest download of ABCD comes with the Agrovoc Thesaurus: just re-use the all parts of this database (copy them into a new directory, e.g. Thes). This database has a different field numbering and uses different prefixes than the examples on the Spanish ABCD wiki. This is important for the conversion of your thesaurus data into this database.

Do not forget to copy the agrovoc.par file to thes.par in the par directory and make sure you have these lines in that thes.par:

```
thes.*=%path_database%thes/data/thes.*
thes.pft=%path_database%thes/pfts/en/thes.pft
prologoact.pft=%path_database%www/prologoact.pft
epilogoact.pft=%path_database%www/epilogoact.pft
```

2.1.1. The def

This is the def of the Agrovoc database:

F	094	ID	0 0 X ID_ 0 1 0
F	001	Descriptor (EN)	1 0 X DE_ 0 1 0
F	002	SN (EN)	0 0 X 0 1 0
F	003	USE (EN)	0 0 X 0 1 0
F	004	UF (EN)	0 0 X 0 1 0
F	005	BT (EN)	0 0 X 0 1 0
F	006	NT (EN)	0 0 X 0 1 0
F	007	RT (EN)	0 0 X 0 1 0
F	008	ND (EN)	0 0 X 0 1 0
F	101	DE (FR)	0 0 X 0 1 0

F	102	SN (FR)	0 0 X 0 1 0
F	103	USE (FR)	0 0 X 0 1 0
F	104	UF (FR)	0 0 X 0 1 0
F	105	BT (FR)	0 0 X 0 1 0
F	106	NT (FR)	0 0 X 0 1 0
F	107	RT (FR)	0 0 X 0 1 0
F	108	ND (FR)	0 0 X 0 1 0
F	201	DE (ES)	0 0 X 0 1 0
F	202	SN (ES)	0 0 X 0 1 0
F	203	USE (ES)	0 0 X 0 1 0
F	204	UF (ES)	0 0 X 0 1 0
F	205	BT (ES)	0 0 X 0 1 0
F	206	NT (ES)	0 0 X 0 1 0
F	207	RT (ES)	0 0 X 0 1 0
F	208	ND (ES)	0 0 X 0 1 0

This structure should be clear to you if you are familiar with the normal thesaurus terminology. Maybe ND is not very clear: it stands for “non descriptor”, i.e. the non preferred term, which is a rejected descriptor; in this case you only have one extra fields (in each used language), i.e. USE. In other words: it redirects the user from a “bad” descriptor to a “good” one.

Best thing is to use these fields numbering when you copy your thesaurus.

It is important to use v8 (ND) and not v1 (descriptor) for your non-preferred descriptor, otherwise the user can just select the wrong term. If your thesaurus contains this:

```
v1 Advocates
v3 Lawyers
```

You can convert this to:

```
V8 Advocates
v3 Lawyers
```

using this mx proc:

```
if p(v3) then
'd1', '<8>',v1,'</8>' fi
```

Then you need to replace those v1 into v8 in you entire database each time v8 is present, using this kind of mx command:

```
mx thes proc=@nd.prc create=thes2 now -all
```

2.1.2. *thes.dat*

In the same directory as the FDT of the thesaurus database you also need a file *thes.dat*. This file contains these lines:

```
alpha_prefix=DE_
perm_prefix=PER_
alpha_pft=v1
display=ficha
```

Also here you could use other prefixes (this would imply to use the same prefixes wherever necessary).

2.2. The pfts

You need a pft called *ficha.pft* in the *pfts* directory of *thes*. This pft contains these lines:

```
`<a href='javascript:Search(`v1`)'><img src=../dataentry/img/search.gif border=0></a>
  "<strong>"V1
  "</strong><br>"V45/"<br>"v32,/
  '<table>',
  if p(v3) then
    '<tr><td valign=top>Use</td><td>'
      (if p(v3) then
        `<a href='javascript:Show(`v3`)'>`v3'<br>' / fi),
    '</td>'
  fi,
  if p(v4) then
    '<tr><td valign=top>UF</td><td>'
      (if p(v4) then `<a href='javascript:Show(`v5`)'>`v4'<br>' /
fi),
      '</td>'/
  fi,
  "<tr><td valign=top>SN</td><td>"V2+|<br>|, "</td>"D2/
  if p(v5) then
    '<tr><td valign=top>BT</td><td>'
      (if p(v5) then
        `<a href='javascript:Show(`v5`)'>`v5'<br>' / fi),
      '</td>'/
  fi,
  if p(v6) then
    '<tr><td valign=top>NT</td><td>'
      (if p(v6) then `<a href='javascript:Show(`v6`)'>`v6'<br>' /
fi),
      '</td>'/
  fi,
  if p(v7) then
    '<tr><td valign=top>RT</td><td>'
      (if p(v7) then `<a href='javascript:Show(`v7`)'>`v7'<br>' / fi),
      '</td>'/
  fi,
  if p(v40) then
    '<tr><td valign=top>Véase además</td><td>'
      (if p(v40) then `<a href='javascript:Show(`v40`)'>`v40'<br>' /
fi),
```

```
'</td>' /  
fi,  
'</table>'  
'<p>' /
```

It is the same as on the Spanish wiki, but with a different field numbering. It is not in the pfts directory of the Agrovoc Thesaurus database; so create it.

2.3. Preparation of the thesaurus data

Your thesaurus must of course be an ISIS thesaurus with the same field numbering as the one in the def. If the numbering is different use MX retag to renumber your fields (see the CISIS Utilities manual). If your thesaurus is not an ISIS database you must convert it using e.g. the CISIS utilities. This may be a complex task depending on the export facilities of your thesaurus software.

When the thesaurus is in place index it using this MX comment:

```
mx thes fst=@thes.fst fullinv=thes now -all
```

3. Preparation of your (bibliographic) database worksheet

You must define the pick-list type of the field in which you want the thesaurus term to be copied as type “thesaurus”:

Type	Name
Pick list	

You do not have to name the thesaurus; the system will know which thesaurus comes with this database because this was already defined in dr_path.def.

The result can be like this:

(You can find more details about the structure of this field in 4.2.)

Now you will see the thesaurus icon next to your entry box in your worksheet:



4. Use of the thesaurus

What you want from your thesaurus module is this:

- You want to use it to search your database.
- While entering data,
 - You want to search for a proper thesaurus term to enter in a field for topical terms.
 - You want to copy this term automatically from your thesaurus into the worksheet of your (bibliographic) database.
 - You want to navigate through the thesaurus and explore the broader, narrower or related terms in order to find one that exactly expresses the content of a document.

Remember that there are three ways to see the thesaurus:

1. As an alphabetical list.
2. As a permuted list.
3. As a systematic display of one term (using BT, NT etc.)

The interface only uses the Spanish words “alfabético” and “permutado”. You can change these words in e.g. English by looking for them in the php scripts index.php, perm.php and show.php which you will find in the directory “thesaurus” of central and replacing them with the words you want to use.

At first the navigation through the thesaurus is tricky; it will be explained here.

4.1. Using your thesaurus for data search

When everything is installed properly you can click the thesaurus icon (the pile of books: ) in your tool bar in order to open the thesaurus. The result will be the alphabetical display (I am using the images of the Spanish wiki here):

The screenshot shows a web-based application interface for managing a thesaurus. On the left, there's a search bar with 'localhost:9090/central/dataentry/inicio_main.php' and a logo for 'ABCD'. Below the search bar is a toolbar with icons for navigation, search, and file operations. The main content area displays a list of terms under the heading 'Alfabético' (Alphabetical). The terms listed are:

- ANALISIS DE IMPACTO
- ANALISIS DE LA INFORMACION
- ANALISIS DE POLITICAS
- ANALISIS DE PROCESO
- ANALISIS DE REDES
- ANALISIS DE SISTEMAS
- ANALISIS ECONOMICO
- ANALISIS ESTADISTICO
- ANALISIS FINANCIERO
- ANALISIS HISTORICO
- ANALISIS INSTITUCIONAL
- ANALISIS JURIDICO
- ANALISIS NORMATIVO
- ANALISIS POLITICO
- ANALISIS SOCIAL
- ANTIGUEDAD
- ANTILLAS
- ANUARIO
- APOLITICISMO
- ARBITRAJE
- ARBOL DE DECISIONES
- ARCHIVO
- ARCHIVOS ADMINISTRATIVOS
- AREA
- AREA RURAL
- ARREGLO DE CONFLICTO
- ASCENSO
- ASESORES

At the bottom of the alphabetical list, there are links for 'Alfabético' and 'Permutado', and a search bar with a 'Continuar' button.

You can navigate the alphabetical display by:

- Clicking in the alphabet on the left-hand site.
- By clicking on “more terms” at the bottom of the box.
- By entering a term in the search box on the bottom and clicking “Continue”.

If you click on any term in the list it will start a search through your database. The result will be the first record of the set that is retrieved. If you are using v8 for the non-preferred term you will not see any non-preferred term in this list.

In order to start the systematic display (do not just click on a term...):

- First check the little check box on the upper right-hand site:
- Then click on a term.

Now the systematic view of this term will open:

MFN: 0018
08.01
Descriptor

BT [ESTRUCTURA DE LA ADMINISTRACION PUBLICA](#)

NT [GOBIERNO CENTRAL](#)
[MINISTERIOS](#)
[RELACIONES INTERGUBERNAMENTALES](#)
[SERVICIOS PUBLICOS](#)

RT [ADMINISTRACION INTERNA](#)
[ADMINISTRACION LOCAL](#)
[ADMINISTRACION SECTORIAL](#)
[CENTRALIZACION](#)
[EMPRESAS PUBLICAS](#)
[FINANZAS DE LA ADMINISTRACION CENTRAL](#)
[ORGANISMOS AUTONOMOS](#)
[REFORMA DE LA ADMINISTRACION CENTRAL](#)

[Alfabético](#) [Permutado](#)

Clicking the terms in this view will allow you to navigate from BT to NT to RT etc. etc. Only when you click on the icon of the looking glass next to the term () it will be used for searching the database.

You can also change to the permuted display by clicking on “Permutado”. The result will be something like this:

The screenshot shows a Mozilla Firefox window displaying the ABCD Thesaurus. The address bar shows the URL localhost:9090/central/thesaurus/perm.php?perm=Y&base=biblo. The main content area lists terms under category A, such as POLITICA DE ACCESO A LA INFORMACION, ACceso a la informacion, ACCESIBILIDAD, etc. A vertical sidebar on the left lists the alphabet from A to Z. At the bottom of the page, there is a search bar with the placeholder text "Ir a" and a button labeled "Continuar". Below the search bar, there are links for "Alfabético" and "Permutado".

You can walk through this list by using the alphabet or by entering a term in the search box on the bottom of the display. If you now click a term, it will be searched.

At this moment the permuted display only works correctly when the descriptors are in upper case in the thesaurus database. You could convert all descriptors to upper case, but then they also will appear in upper case in your worksheet when you use the thesaurus for data entry...

4.2. Using the thesaurus in data entry

Important notice: at this moment this only works correctly with fields *without subfields*. See form more details: paragraph 5. Known problems.

Click on the thesaurus icon (the pile of books) next to the field tag so that the thesaurus opens (in a different window):



The navigation and use is pretty much the same as in the one used for searching the database, but if you now click or select a term it will be entered in your worksheet field.

You can see that there is also a looking glass under the thesaurus icon. In fact you have the choice: selecting terms from the thesaurus. In that case you have to click the thesaurus icon. In this example you could also select terms that previously have been used in field 650 of your database. In that case you must click the looking glass. Now the dictionary of terms will open and you can choose terms by clicking on them:

Dictionary of terms: - Database: marc

Click on one or more terms to copy them to the data entry form. Use [more] to access more terms or enter the start of the required term in the box go to and click [Enter]. You can also select a letter in the left-hand index.

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

APPLICATION SOFTWARE
ARCHIVES
ART BRUT
BABIES
CATS
CELL MEMBRANE
CELL TRANSFORMATION, VIRAL
COLD (DISEASE)
COLLEGE GRADUATES
COMMUNISM AND PSYCHOANALYSIS.
DOCUMENTATION
DOGS
ELECTRONIC DATA PROCESSING.
ELECTRONIC DATA PROCESSING
EPIDEMIOLOGY
HEMAGGLUTININ
INFLUENZA EPIDEMIC, 1918-1919
INFLUENZA VACCINES
INFLUENZA VIRUSES
INFLUENZA.
INFLUENZA
INFORMATION SCIENCE
INSTITUTIONAL INVESTMENTS.
JAVA (COMPUTER PROGRAM LANGUAGE)

Preview Continue Go to:

This is the reason why we defined field 650 this way:

39 Field 650 Thesaurustermen Text/Textarea 1 Thesaura: MA_ [browse](#) v650 v650

- MA_ is the prefix to build the dictionary of terms.
- V650 is what should be displayed in that dictionary and also which field will be chosen when we click a term. This is not different from what you probably already defined for other fields too.

5. Known problems

- The **permuted display only works well if the descriptors are in upper case**. If they are in lower case you just see the words listed alphabetically without the context.
- In data entry **you cannot use the thesaurus to copy terms into a subfields**. It works if you add the thesaurus to the field itself, but not if you add it to a subfields. You still could copy from the thesaurus into the field as such, but after this you will have to add the subfield delimiter manually. An example:
 - You can copy this into the field:
650 : ABYSSOBENTHIC ENVIRONMENT
 - and then edit it manually into:
650 : ^aABYSSOBENTHIC ENVIRONMENT

This is tricky if you have to take care of indicators too (as is the case if the database has a MARC structure). An alternative could be that you copy the term from the “dictionary of terms” into the subfield entry box, but then you will only see the terms that have been used previously in data entry.