

Installing NCI Vocabularies in LexEVS 6.x

Contents of this Page

- [NCI Thesaurus](#)
- [NCI Metathesaurus](#)
 - [Loading the Metathesaurus](#)
 - [Resuming Loads](#)
- [NCI Thesaurus History](#)

NCI Thesaurus

This section describes the steps to download and install a full version of the NCI Thesaurus.

The NCI Thesaurus differs from other OWL formatted resources and, as a result administrators usually take advantage of some of the custom loading resources available in LexEVS. While users can download the Thesaurus.owl file from NCI and load it into LexEVS without it failing, many users will want to customize it with an XML formatted manifest and preferences files.



Note

The NCI Thesaurus has grown large enough that it can no longer be loaded on many typical desktop machines. We recommend a 64-bit operating system running on a multiprocessor computer with a minimum of 4g of memory. Server class Linux machines are the typical target for these loads. The time to load NCI Thesaurus will vary depending on machine, memory, and disk speed. Expect a couple of hours for a higher end machine.

Step	Action
------	--------

1

Using a web or ftp client go to the URL: <ftp://ftp1.nci.nih.gov/pub/cacore/EVS/>

Name	Size	Type	Modified
archive		File Folder	2/6/2006 1:24 PM
caBIG_lexGrid		File Folder	6/22/2005 12:00 AM
CDR		File Folder	6/14/2005 12:00 AM
fda		File Folder	10/18/2005 5:46 PM
protege		File Folder	10/19/2005 4:33 PM
ThesaurusSemantics		File Folder	2/6/2006 1:17 PM
ThesaurusTermsofUse_files		File Folder	9/23/2003 12:00 AM
Filtered_pipe_out.zip	0.97 MB	Compressed (zipped)...	8/26/2005 12:00 AM
Filtered_pipe_out_0601c.txt	64.0 KB	Text Document	2/6/2006 10:04 AM
Filtered_pipe_out_12f.txt	78.0 KB	Text Document	1/11/2006 12:18 PM
full_pipe_out.zip	2.25 MB	Compressed (zipped)...	7/1/2005 12:00 AM
full_pipe_out_0601c.txt	210 KB	Text Document	2/6/2006 10:04 AM
full_pipe_out_12f.txt	471 KB	Text Document	1/11/2006 12:18 PM
Metathesaurus_P051117.zip	625 MB	Compressed (zipped)...	1/27/2006 5:39 PM
mmsys.a.prop	19.0 KB	PROP File	5/25/2004 12:00 AM
MMSYS.jar	6.70 KB	Executable Jar File	5/25/2004 12:00 AM
mmsys.prop.sav	19.0 KB	SAV File	5/25/2004 12:00 AM
NCI_RRF_Addendum.pdf	130 KB	Adobe Acrobat Doc...	7/22/2004 12:00 AM
NCI_THESAURUS_license.txt	6.33 KB	Text Document	10/21/2003 12:00 AM
onlylog.dtd	7.18 KB	DTD File	1/4/2006 3:44 PM
ReadMe.txt	4.89 KB	Text Document	2/6/2006 10:38 AM
ReadMe_history.txt	3.70 KB	Text Document	2/6/2006 10:38 AM
Thesaurus_05.11f.FLAT.zip	3.04 MB	Compressed (zipped)...	1/4/2006 3:43 PM
Thesaurus_05.11f.OwL.zip	6.76 MB	Compressed (zipped)...	1/4/2006 3:43 PM
Thesaurus_05.11f.XML.zip	6.98 MB	Compressed (zipped)...	1/4/2006 3:43 PM
Thesaurus_05.12f.FLAT.zip	3.12 MB	Compressed (zipped)...	2/6/2006 10:18 AM
Thesaurus_05.12f.OwL.zip	6.86 MB	Compressed (zipped)...	2/6/2006 10:18 AM
Thesaurus_05.12f.XML.zip	7.10 MB	Compressed (zipped)...	2/6/2006 10:18 AM
ThesaurusTermsofUse.htm	10.8 KB	HTML Document	10/21/2003 12:00 AM
ThesaurusTermsofUse.pdf	82.2 KB	Adobe Acrobat Doc...	3/29/2005 12:00 AM

2

Select the version of NCI Thesaurus OWL file you wish to download. Save the file to a directory on your machine.

3

Extract the OWL file from the ZIP download and save in a directory on your machine. This directory will be referred to as NCI_THESAURUS_DIRECTORY in script examples.

4

Create Manifest and Preferences files. (optional)

Manifests update or fill empty terminology metadata and can make adjustments to names and alternate names, versions and many other things that the source terminology normally wants to say. This is very useful when the source terminology itself does not supply this information, but the user needs to record it in the terminology service representation. Some versions of the NCI Thesaurus may not load without a manifest. What follows is a sample manifest file used to update alternate names, language designations, versions and other metadata. Details of manifest elements are found in the Administration Guide to the [manifest file](#).

```
<?xml version="1.0" encoding="UTF-8"?>
<CodingSchemeManifest xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://LexGrid.org/schema/2010/01/LexOnt/CodingSchemeManifest"
  xmlns:owldef="http://LexGrid.org/schema/2010/01/LexOnt/CodingSchemeManifest"
  xmlns:lgNaming="http://LexGrid.org/schema/2010/01/LexGrid/naming"
  xmlns:lgRel="http://LexGrid.org/schema/2010/01/LexGrid/relations"
  xmlns="http://LexGrid.org/schema/2010/01/LexOnt/CodingSchemeManifest"
  id="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#">
  <codingScheme toOverride="true">NCI_Thesaurus</codingScheme>
  <entityDescription toOverride="true">NCI Thesaurus</entityDescription>
  <formalName toOverride="true">NCI Thesaurus</formalName>
  <codingSchemeURI>http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#</codingSchemeURI>
  <defaultLanguage toOverride="true">en</defaultLanguage>
  <representsVersion toOverride="true">11.10e</representsVersion>
  <localName toAdd="true">NCI Thesaurus</localName>
  <localName toAdd="true">NCI_Thesaurus</localName>
  <localName toAdd="true">NCI</localName>
  <source toUpdate="true" role="creator">NCI</source>
  <owldef:mappings xmlns="http://LexGrid.org/schema/2010/01/LexGrid/codingSchemes"
    <lgNaming:supportedCodingScheme localId="NCI_Thesaurus" uri="http://ncicb.
    <lgNaming:supportedRepresentationalForm localId="text_plain" uri="urn:oid:2
  </owldef:mappings>
  <owldef:associationDefinitions toUpdate="true">
    <assoc associationName="Has_Salt_Form" entityCode="A5" forwardName="Has_Sal
  </owldef:associationDefinitions>
</CodingSchemeManifest>
```

Preference files control how data is loaded and, while they do not add anything to the source, they can change the representation of a terminology by making choices as to what is loaded as a property, entity, or association from the source terminology. The following preference file sets root nodes for the terminology and processes a set of complex properties not handled by the OWL processor, among other things. The full preferences definitions are described in the Administration Guide to the [preferences file](#).

```
<?xml version="1.0" encoding="UTF-8"?>
<OWLLoaderPreferences xmlns="http://LexGrid.org/schema/LexBIG/2009/01/Preferences/load/OWLLoadPreferences"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://LexGrid.org/schema/LexBIG/2009/01/Preferences/load/OWLLoadPreferences
  http://lexgrid.org/schema/LexBIG/2009/01/Preferences/load/OWLLoadPreferences.xsd">
  <MatchConceptStatus>DONTMATCHANYTHING</MatchConceptStatus>
  <MatchRootName>(C12218|C12219|C12913|C14250|C16612|C17828|C1908|C97325|C20047|C20181|C20189|C20633
  |C22187|C22188|C26548|C28389|C28428|C3910|C43431|C7057)</MatchRootName>
  <ProcessComplexProps>true</ProcessComplexProps>
  <StrictOWLImplementation>false</StrictOWLImplementation>
</OWLLoaderPreferences>
```

Since the Thesaurus may be released as either a "by code" or "by name" formatted source, preference files can serve the purpose of adjusting the entity code to its intended place in LexEVS. Here are some differences in an OWL formatted Thesaurus source.

- "By Code" formatted Thesaurus OWL class

```
<owl:Class rdf:ID="C1324">
  <rdfs:label>Recombinant Interleukin-9</rdfs:label>
  <rdfs:subClassOf rdf:resource="#C593"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#R125"/>
      <owl:someValuesFrom rdf:resource="#C40739"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
```

which has its entity code formatted as an rdf:ID. This loads as the unique identifier without using a preferences file.

- "By name" version

```
<owl:Class rdf:about="#Therapeutic_Interleukin-9">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    >Recombinant Interleukin-9</rdfs:label>
  <rdfs:subClassOf rdf:resource="#Therapeutic_Interleukin"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#Chemical_Or_Drug_Has_Physiologic_Effect"/>
      <owl:someValuesFrom rdf:resource="#Hematopoiesis_Induction"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <DEFINITION rdf:parseType="Literal">
```

- With the code found as a "<code>" tagged attribute of the class

```
>Amino Acid, Peptide, or Protein</Semantic_Type>
<code rdf:datatype="http://www.w3.org/2001/XMLSchema#string">C1324</code>
<UMLS_CUI rdf:datatype="http://www.w3.org/2001/XMLSchema#string">C1522551</UMLS_CUI>
<FDA_UNII_Code rdf:datatype="http://www.w3.org/2001/XMLSchema#string">EY7FQ27ISK</FDA_UNII_Code>
<Contributing_Source rdf:datatype="http://www.w3.org/2001/XMLSchema#string">FDA</Contributing_Source>
```

In this case we can insure that the value "C1324" will be loaded into LexEVS as an entity code by supplying a preferences file with the following values. Loading this type of terminology without the preferences file may cause data truncation errors on the longer names and cause data loss on load.

```
<?xml version="1.0" encoding="UTF-8"?>
<OWLLoaderPreferences xmlns="http://LexGrid.org/schema/LexBIG/2009/01/Preferences/load/OWLLoadPreferences" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://LexGrid.org/schema/LexBIG/2009/01/Preferences/load/OWLLoadPreferences http://lexgrid.org/schema/LexBIG/2009/01/Preferences/load/OWLLoadPreferences.xsd">
  <MatchConceptCode>code</MatchConceptCode>
  <DatatypePropSwitch>association</DatatypePropSwitch>
</OWLLoaderPreferences>
```

Finally, post processing is also available for this and other sources, as explained in the Administrative guide [post processor](#) section.

- 5 Using the LexEVS command line, load the NCI Thesaurus with no options:
Change to the LexEVS home directory.

```
cd {LEXEVS_HOME}/admin
```

For Windows installation use the following command:

```
LoadOWL.bat -in "file:/// {NCI_THESAURUS_DIRECTORY} /Thesaurus_10.10d.owl
```

For Linux installation use the following command:

```
LoadOWL.sh -in "file:/// {NCI_THESAURUS_DIRECTORY} /Thesaurus_10.10d.owl
```

This should work best with a "by code" type Thesaurus source terminology.

6	<p>Using the LexEVS command line, load the NCI Thesaurus with options: Change to the LexEVS home directory.</p> <pre>cd {LEXEVS_HOME}/admin</pre> <p>For Windows installation use the following command:</p> <pre>LoadOWL.bat -in "file:/// {NCI_THESAURUS_DIRECTORY}/Thesaurus_10.10d.owl" -mf "file:/// {NCI_THESAURUS_DIRECTORY}/Thesaurus_MF.xml" -lp "file:/// {NCI_THESAURUS_DIRECTORY}/Thesaurus_prefs.xml"</pre> <p>For Linux installation use the following command:</p> <pre>LoadOWL.sh -in "file:/// {NCI_THESAURUS_DIRECTORY}/Thesaurus_10.10d.owl" -mf "file:/// {NCI_THESAURUS_DIRECTORY}/Thesaurus_MF.xml" -lp "file:/// {NCI_THESAURUS_DIRECTORY}/Thesaurus_prefs.xml"</pre>
---	--

Example output from load of NCI Thesaurus 05.12f

```
...
[LexBIG] Processing TOP Node... Retired_Kind
[LexBIG] Clearing target of NCI_Thesaurus...
[LexBIG] Writing NCI_Thesaurus to target...
[LexBIG] Finished loading DB - loading transitive expansion table
[LexBIG] ComputeTransitive - Processing Anatomic_Structure_Has_Location
[LexBIG] ComputeTransitive - Processing Anatomic_Structure_is_Physical_Part_of
[LexBIG] ComputeTransitive - Processing Biological_Process_Has_Initiator_Process
[LexBIG] ComputeTransitive - Processing Biological_Process_Has_Result_Biological_Process
[LexBIG] ComputeTransitive - Processing Biological_Process_Is_Part_of_Process
[LexBIG] ComputeTransitive - Processing Conceptual_Part_Of
[LexBIG] ComputeTransitive - Processing Disease_Excludes_Finding
[LexBIG] ComputeTransitive - Processing Disease_Has_Associated_Disease
[LexBIG] ComputeTransitive - Processing Disease_Has_Finding
[LexBIG] ComputeTransitive - Processing Disease_May_Have_Associated_Disease
[LexBIG] ComputeTransitive - Processing Disease_May_Have_Finding
[LexBIG] ComputeTransitive - Processing Gene_Product_Has_Biochemical_Function
[LexBIG] ComputeTransitive - Processing Gene_Product_Has_Chemical_Classification
[LexBIG] ComputeTransitive - Processing Gene_Product_is_Physical_Part_of
[LexBIG] ComputeTransitive - Processing hasSubtype
[LexBIG] Finished building transitive expansion - building index
[LexBIG] Getting a results from sql (a page if using mysql)
[LexBIG] Indexed 0 concepts.
[LexBIG] Indexed 5000 concepts.
[LexBIG] Indexed 10000 concepts.
[LexBIG] Indexed 15000 concepts.
[LexBIG] Indexed 20000 concepts.
[LexBIG] Indexed 25000 concepts.
[LexBIG] Indexed 30000 concepts.
[LexBIG] Indexed 35000 concepts.
[LexBIG] Indexed 40000 concepts.
[LexBIG] Indexed 45000 concepts.
[LexBIG] Indexed 46000 concepts.
[LexBIG] Getting a results from sql (a page if using mysql)
[LexBIG] Closing Indexes Mon, 27 Feb 2006 01:44:22
[LexBIG] Finished indexing
```

NCI Metathesaurus

Loading the Metathesaurus

This section describes the steps to download and install a full version of the NCI Metathesaurus for the LexEVS Service.

**Note**

NCI Metathesaurus contains many individual vocabularies some of which are large vocabularies in and of themselves. It requires many hours to load and index. It can require 36 hours on a multiprocessor machine with 6 GB or more of memory. The total time to load the NCI Metathesaurus will vary depending on machine, memory, and disk speed. Because this loader uses a batch loading strategy it is less dependent on memory, but some users will see 3 or 4 day load times with average multiprocessor processing power. While the batch loading process itself is not heavily memory dependent, the creating the index following the load will require at least 3 GB of memory.

Step	Action
1	Using a web client go to the URL: https://cbiit.nci.nih.gov/evs-download/metathesaurus-downloads Note that a valid UMLS license is required to download the NCI Metathesaurus due to the UMLS content inside. You will find out how to manage this on the site above if you do not have a license.
2	Download Metathesaurus.RRF.zip. Save the file to a directory on your machine. The compressed file is approximately 2GB in size and may take some time to download depending on your connection speed.
3	Extract the files from the ZIP download and save to a directory on your machine. This directory will be referred to as NCI_METATHESAURUS_DIRECTORY. RELEASE_INFO.RRF is required to be present for the load utility to work.
4	Check that you are able to open a large number of files before starting the load. <pre>ulimit -Hn</pre> Usually having around 10,000 available open files is sufficient. If your limit is set to low this will need to be raised.
5	Using the LexEVS utilities load the NCI Thesaurus: <pre>{LEXEVS_HOME}/admin</pre> For Windows installation use the following command: <pre>LoadMetaBatch.bat -in "file:/// {NCI_METATHESAURUS_DIRECTORY} /"</pre> For Linux installation use the following command: <pre>LoadMetaBatch.sh -in "file:/// {NCI_THESAURUS_DIRECTORY} /"</pre>

Resuming Loads

**Remember**

Killed processes cannot be restarted. The load can recover from an application error, but not from an outside activity that stops the process. This does not include out of memory errors, hung processes that have been killed, or system shutdown.

Since this loader is resource hungry we provide the option to restart should you find your resource settings to be inadequate. Resuming loads which have crashed is possible using the ResumeBatchLoad script set.

Step	Action
1	Open the lbGUI
2	Find in the table the terminology with the broken or stopped load. This should have a status of pending
3	Highlight and double click the terminology row in the table.
4	On the resulting window note for the following command line execution: The URI and the version.

5	<p>Using the LexEVS utilities restart the load of the NCI Metathesaurus:</p> <pre>{LEXEVS_HOME}/admin</pre> <p>For Windows installation use the following command:</p> <pre>ResumeMetaBatch.bat -in "file:/// {NCI_METATHESAUROS_DIRECTORY}/" -s "NCI Metathesaurus" -uri "urn:oid:2.16.840.1.113883.3.26.1.2" -version "200601"</pre> <p>For Linux installation use the following command:</p> <pre>ResumeMetaBatch.sh -in "file:/// {NCI_THESAURUS_DIRECTORY}/" -s "NCI Metathesaurus" -uri "urn:oid:2.16.840.1.113883.3.26.1.2" -version "200601"</pre>
---	--

NCI Thesaurus History

This section describes the steps to download and install a history file for NCI Thesaurus.

Step	Action
1	Using a web or ftp client go to the URL: ftp://ftp1.nci.nih.gov/pub/cacore/EVS/
2	Select the version of NCI History you wish to download. Save the file to a directory on your machine. Select the VersionFile download to the same directory as the history file.
3	Extract the History files from the zip download and save in a directory on your machine. This directory will be referred to as NCI_HISTORY_DIRECTORY
4	<p>Using the LexEVS utilities load the NCI Thesaurus:</p> <pre>cd {LEXEVS_HOME}/admin</pre> <p>For Windows installation use the following command:</p> <pre>LoadNCIHistory.bat -nf -in "file:/// {NCI_HISTORY_DIRECTORY}" -vf "file:/// {NCI_HISTORY_DIRECTORY} /VersionFile"</pre> <p>For Linux installation use the following command:</p> <pre>LoadNCIHistory.sh -nf -in "file:/// {NCI_HISTORY_DIRECTORY}" -vf "file:/// {NCI_HISTORY_DIRECTORY} /VersionFile"</pre>



Note

If a 'releaseld' occurs twice in the file, the last occurrence will be stored. If LexEVS already knows about a releaseld (from a previous history load), the information is updated to match what is provided in the file.

This file has to be provided to the load API on every load because you will need to maintain it in the future as each new release is made. We have created this file that should be valid as of today from the information that we found in the archive folder on your ftp server. You can find this file in the 'resources' directory of the LexEVS install.