## Instructions on creating new STF (NALT skill cartridge) for Web Studio

June 13, 2019 (SS): Started with instructions for Luxid 6 and changed for Luxid 7 (Web Studio)

## **Quick Checklist of procedure:**

- A. Export XML file from MultiTes.
- B. Convert XML to SKOS file
- C. Create new project in Web Studio using Thesaurus template
- D. Import Thesaurus data from SKOS file
- E. Configure Project
  - a. Languages
  - b. Static Mapping
  - c. STF Parameters
  - d. Other?
- F. Deploy to Annotation Server and Web Studio (last optional)
- G. Update Annotation Plan with new STF resource.
- H. Test your updated Annotation Plan with new STF/Thesaurus

# **General principles:**

The main working file in MultiTes is the only place where thesaurus additions, deletions and edits are performed. All files, including the required STF for Automated Indexing, are derived as a report from this main thesaurus file. Therefore, Luxid Knowledge Editor (Luxid 6) or Web Studio (Luxid 7) is NOT used for the manipulation of thesaurus data for Automated Indexing.

Thesaurus data is backed up at the conclusion of an editing session. Data is automatically saved on the "Theo" thesaurus computer at C:\mtm2007\data. (Note that 2007 is the year designation for the MultiTes software and is not associated with the annual update of the thesaurus). Data is manually backed up by thesaurus staff at the conclusion of an editing session at:

S:\NAL|DPD\Indexing&Informatics\mtm\20xx, where xx in the file name represents the working file data year. The naming convention for files is "NALT20xx.th2. In calendar year 2017, the working file is 2018. In calendar year 2018, the working file is 2019, and so on.

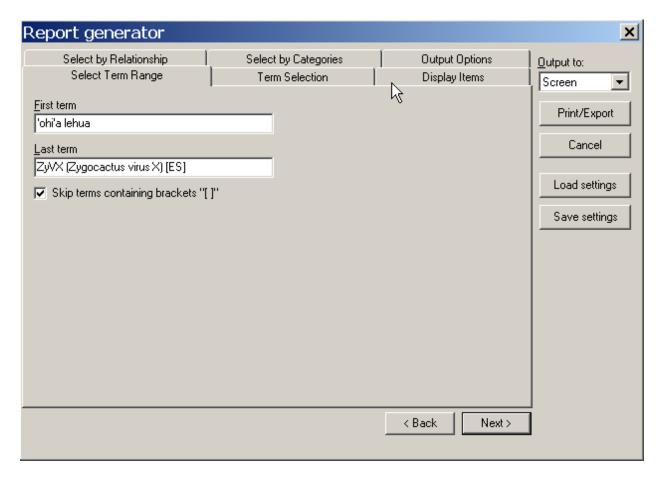
The procedure followed will require you to have:

- 1) Access to MultiTes. Access to latest main thesaurus working file on Theo or the latest backup on the G: drive. Get approval from thesaurus staff to create a new STF. Thesaurus staff will need to run ORPHAN, Pref have BT, Pref have SC reports and resolution must be completed before creating a new STF.
- Access to a Browser, such as Chrome (recommended over Internet Explorer for Luxid).
- 3) Access to Web Studio
- 4) Access to LuxidAdmin, password required.

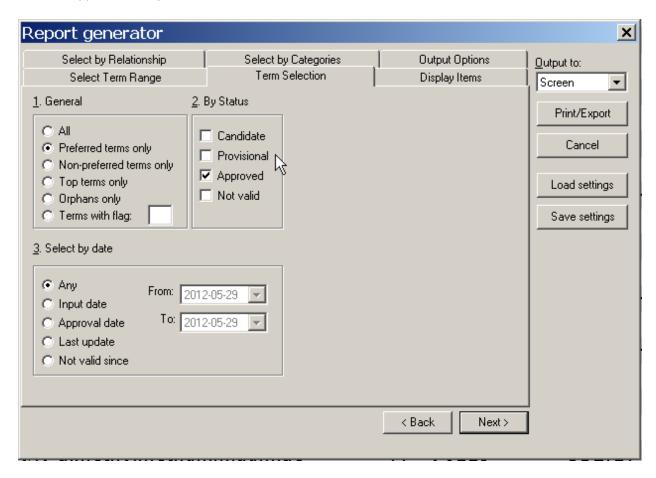
### Procedure:

- 1. Open MultiTes Pro and the main working file.
- 2. Ensure that you are in the main thesaurus working file before continuing (as MultiTes will open the last file opened, and someone may have been in another file for another purpose).
- 3. Check that you are in the ENGLISH "side of the database". The header on the first column should say "Term (English)".
- 4. At the top menu, select REPORTS and a pull down menu will appear. Select the first item, REPORT GENERATOR. A new window will appear.

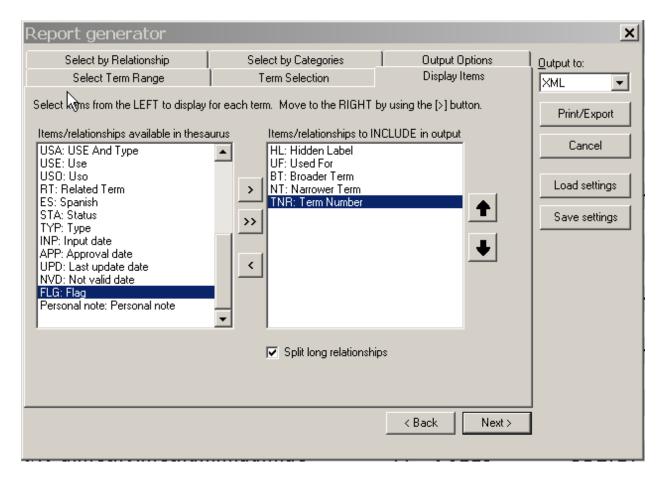
5. The first tab, is SELECT TERM RANGE. In this window, check the box "Skip term containing brackets []" as shown below.



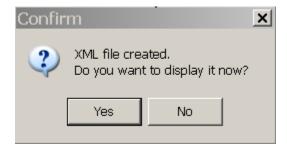
6. Click on the TERM SELECTION tab. Select radio button "Preferred terms only". Select on the "Approved" only, remove checks on other values as shown below.



- 7. Click on the DISPLAY ITEMS tab. Click on the items in the left column and then click on the > arrow to move to the right column, which are the items that wish to include in the output. For the STF, select as shown below.
  - a. HL: HIDDEN LABEL
  - b. UF: USED FOR
  - c. BT: BROADER TERM
  - d. NT: NARROWER TERM
  - e. TNR: TERM NUMBER
  - f. TY: TYPOGRAPHICAL ERROR
  - g. RT: RELATED TERM (added new in this STF; June, 2019)
  - h. INP: Input date (added new in this STF; June, 2019)
  - i. UPD: Last update date (added new in this STF; June, 2019)



- 8. On the right side of the window, pull down the menu beneath OUTPUT TO, Select XML as shown above.
- 9. Click on the button PRINT/EXPORT. A file progress window will open and close.
- 10. Window will open that prompts you to name your file. Default is to put under C:\mtm2007\xml. Give a descriptive name, such as: NALT2018\_May29\_2017.xml. Click on SAVE.
- 11. Window opens "XML file created. Do you want to display it now?", select NO.



- 12. Close REPORT GENERATOR window
- 13. Close MultiTes.
- 14. Go to explore and copy the XML file you just created and place in Q:\NALT\_XML\NEW

15. Open Browser, and go



16. Now that we have our XML file from MultiTes, we can use this program to convert the XML file into a SKOS file. Click on the third item, "NALT-to-SKOS Converter". BROWSE to your file (usually in the C:\mtm2007\xml folder) and select.



17. Click on RUN CONVERSION. You will get a message at the bottom of the screen:

Sending request to ai-dev.nal.usda.gov

The request can take several minutes to process. Look for the blue progress bar. If it is stuck on "sending request", close your browser and try again. If you opened in Fireflox, close and try again in IE. "Done" should be on the bottom status bar and a new window appears.

18. To speed up the retrieval of the file, do not open it. Go back to the converter page, programs.nal.usda.gov By clicking on the "return to upload form".

NALT-to-SKOS Converter	
Upload MultiTes XML file.	
File to Upload:	Browse
Run Conversion	
go to output folder	

19. Select "go to output folder"

# Index of /SKOS

<u>Name</u>	Last modified	Size Description
Parent Directory		-
April2013NALT-SKOS.xml	16-Apr-2012 15:00	19M
April2013_forSTF-SKOS.xml	05-Apr-2012 09:21	19M
BreedingGenetics-SKOS.xml	03-Mar-2012 16:37	150K
EconomicsMinusProducts-SKOS.xml	03-Mar-2012 14:14	311K
<u>GeogLocations-SKOS.xml</u>	03-Mar-2012 14:13	238K
HealthPathology-SKOS.xml	03-Mar-2012 16:37	1.0M
Methodology-SKOS.xml	03-Mar-2012 14:50	163K
NALT-SKOS.xml	29-Feb-2012 16:19	19M
NALT2013 May29 2012-SKOS.xml	29-May-2012 16:09	18M
NaturalResourcesEarthEnv-SKOS.xml	03-Mar-2012 14:49	775K
OneRecordNutritiveValue-SKOS.xml	23-Feb-2012 19:38	3.4K
OrganismsGroups-SKOS.xml	03-Mar-2012 14:50	274K
PlantSciPlantProd-SKOS.xml	03-Mar-2012 14:46	407K
Plantae-SKOS.xml	03-Mar-2012 14:47	3.2M
Q:\SKOS\OneRecordNutritiveValue-SKOS.xml	23-Feb-2012 19:30	3.4K
SeveralRecordsNutritiveValue2Nylon-SKOS.xml	29-Feb-2012 10:06	7.6K
cluster_10-SKOS.xml	01-Mar-2012 17:45	336K
continouscropping-SKOS.xml	03-Apr-2012 13:49	1.0K
get_hierarchy.html	16-Apr-2012 09:08	1.1K

Apache/2.2.3 (Red Hat) Server at ai-dev.nal.usda.gov Port 80

- 20. Right Click on your file and select "save link as".
- 21. A standard "save as" window appears. Save to Q:\SKOS. It will take a few minutes for it to copy it from the server to the Q: drive location.
- 22. Now that you have your SKOS file, do data integrity/quality control testing on SKOS file. Right click on SKOS file and "Edit with Vim". Look for &# OR

<mark>&</mark>#145

<mark>&</mark>#146

<mark>&</mark>#147

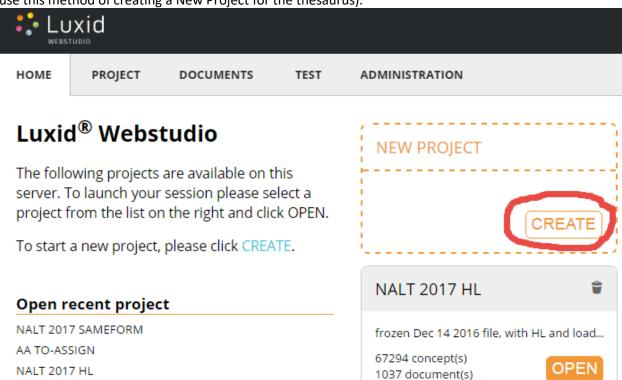
<mark>&#148</mark>

&#150 etc. in prefLabel/s or altLabel/s and if there is any of it, correct in NALT file and redo the whole process. (Note: These oblique quotes are Okay in hiddenLabel/s). This step is not needed after installing UNICODE version of MultiTes, more testing needed. Otherwise, proceed to create the STF!

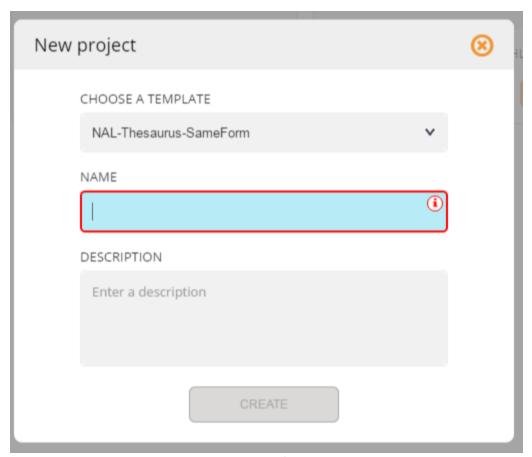
- 23. Open Web Studio in Chrome: <a href="http://ai-luxid.nal.usda.gov:8060/LuxidStudio">http://ai-luxid.nal.usda.gov:8060/LuxidStudio</a>
- 24. You will be prompted to login with username and password.

NALT 2017 HL FUZZY

25. Under HOME tab, the list of projects will appear. **Click on "Create" to create a New Project**. (Since you are not creating a new project from "project backup" but from a SKOS file, you will use this method of creating a New Project for the thesaurus).



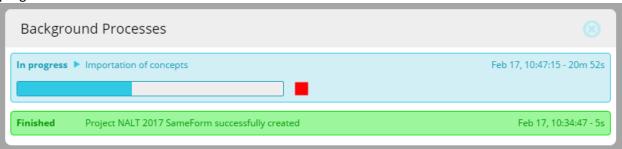
26. A new window will appear. Under "choose a template", select NAL-Thesaurus-SameForm. (Note: the "same form" template will mark all concepts in the thesaurus with Extraction Method "same form". If you use the "NAL-Thesaurus" template, then all concepts in the thesaurus will have the default value for the Extraction Method, which is "fuzzy matching".)



- 27. Enter a unique and descriptive Name for the thesaurus project, using NALT version and dates to identify the version of data, such as "NALT\_2018\_July\_18\_2017".
- 28. Enter a description, noting the data source, template used or other characteristics of the data.
- 29. Click on CREATE button.
- 30. Give it a few moments and you will see a blue message appear that shows that it is busy loading the data in the background. The "background processes" icon in the upper right will have a "1" in a blue circle to show that there is activity running in the background.



31. If you click on the "background processes" icon you will open a window that shows you the progress:



32. When the process is complete, you can go back to the HOME tab and see your project. The number of concepts should match the number of terms in the originating MultiTes file.

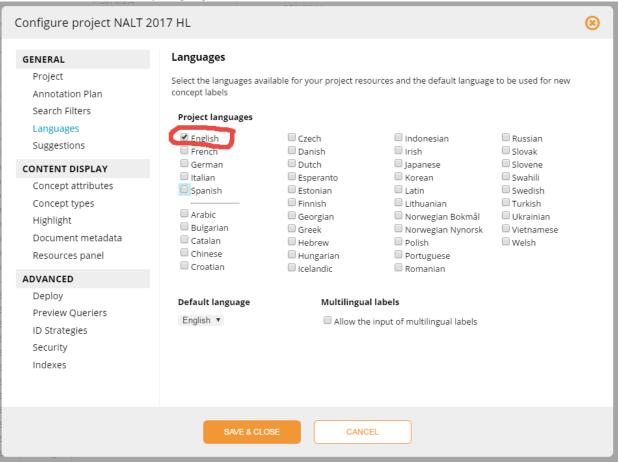


- 33. OPEN your new project:
- 34. Note: Murali suggests that you do the configurations of the project PRIOR to loading in data, as this will save time in the overall process.
- 35. **Next steps are to CONFIGURE your project**. Go to the upper right hand area and click on the name of your project to open the pull down menu.



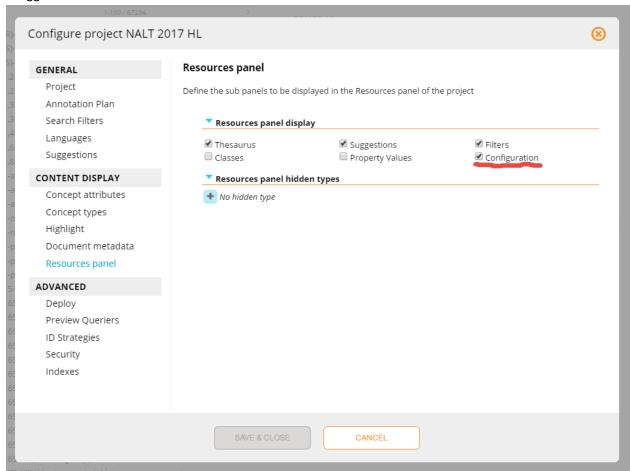
36. Select "CONFIGURE" from the pull down menu.

37. By default, the Project Languages are English, French and German. Uncheck French and German so that ENGLISH is the only language checked.

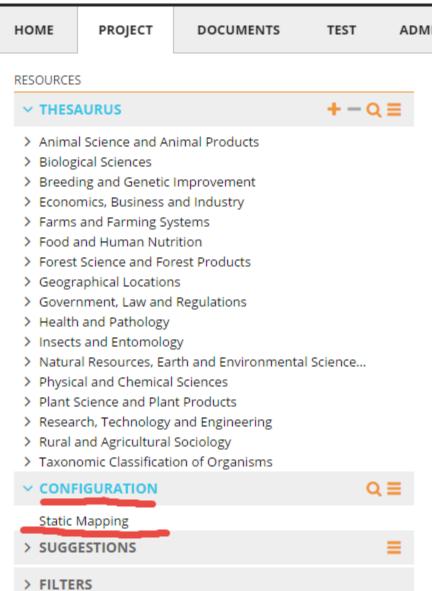


- 38. Click on SAVE & CLOSE.
- 39. **Next configuration is to set the STATIC MAPPING**. The reason we want to configure static mapping (setting the normalization value to \${prefLabel.english}), is that it will enable us to view the "prefLabel" of the concept in Annotation Workbench (instead of the URI number). This enables human readability when doing quality control of batches.

40. In the upper right, Use the pull down menu for the thesaurus project and select "Configure". Click on "Resources panel" under "Content Display" and check the box for "Configuration", "Suggestions" and "Filters". Click on "Save & Close"



41. Now, in the Project view, look for "Configuration" on the left side



42. Select "Static Mapping". On the panel called "Search in Static Mapping", click on the symbol to add a new concept, and input "/Thesaurus/Concept", then click ADD. (NOTE: If the concept already exists, please don't add another one. We may modify the existing concept).



43. Select this new concept. On the panel that would appear on the right, click on NORMALIZATION.



44. Add the value \${prefLabel.english} under the field NORMALIZATION. Then save the changes.

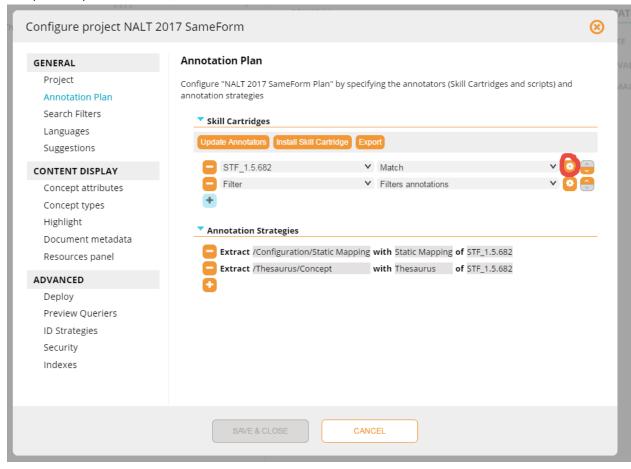


45. **Next, need to check the parameters that are set for STF/Thesaurus**. These should be static since they are inherited from the NAL-Thesaurus template, but check to make sure of the settings.

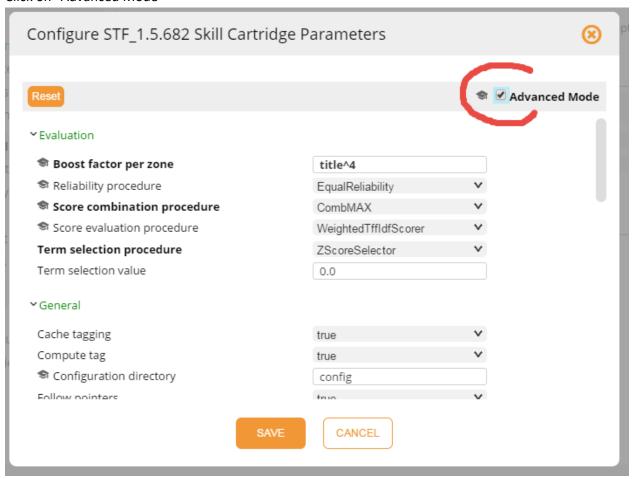
46. Go back to the upper right hand gray box with the name of the project and select CONFIGURE from the pull down menu:



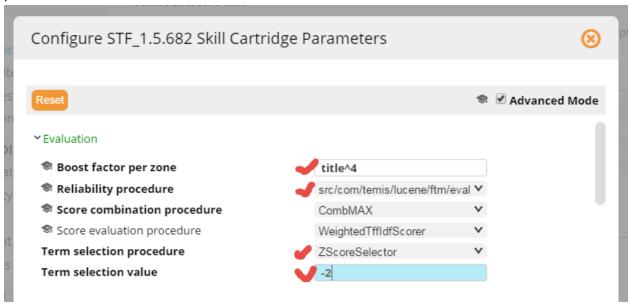
- 47. In the "Configure Project" window, select "Annotation Plan" from the left panel.
- 48. To open the parameters list, click on the flower icon:



# 49. Click on "Advanced Mode"



50. The parameters cover several screens, so you will need to scroll to see all parameters. Select the parameters shown in the screen shots:

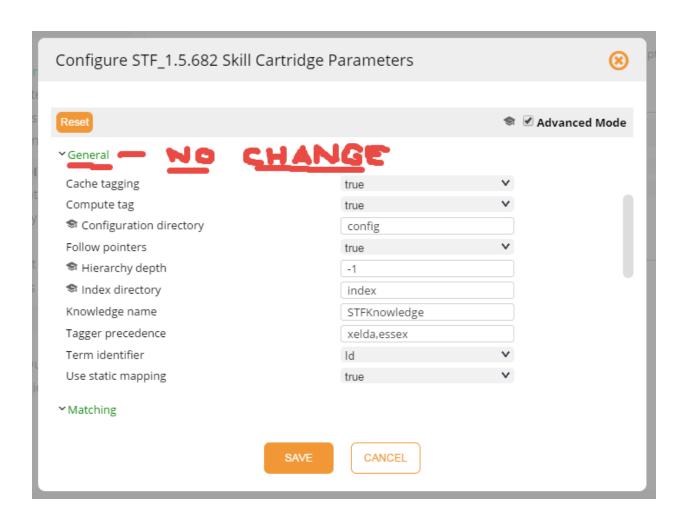


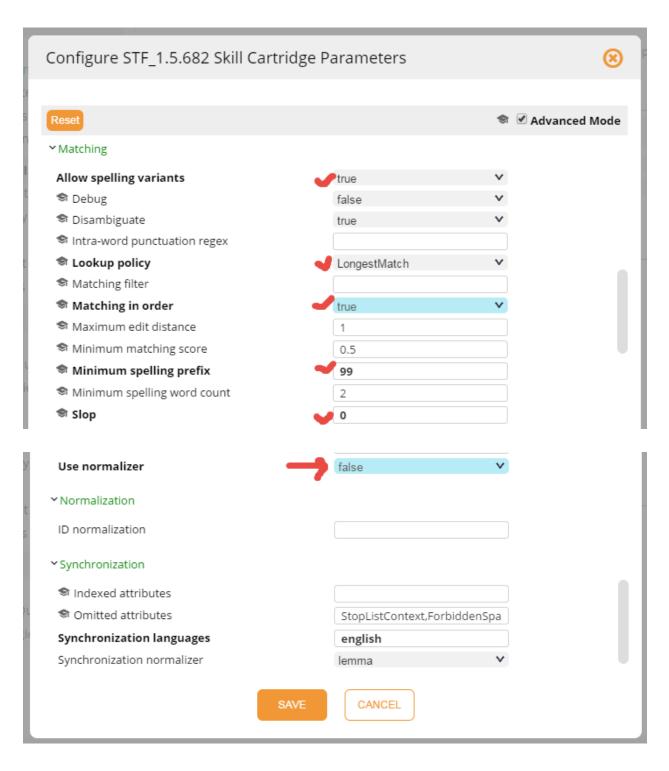
## Explanation of parameters:

Boost factor per zone: title^4 increases the ranking of terms that appear in the title by a factor of 4.

Reliability procedure: The selection above is using the special reliability procedure that was written for us that resembles the "NotNounsReliability" that we used in Luxid 6.

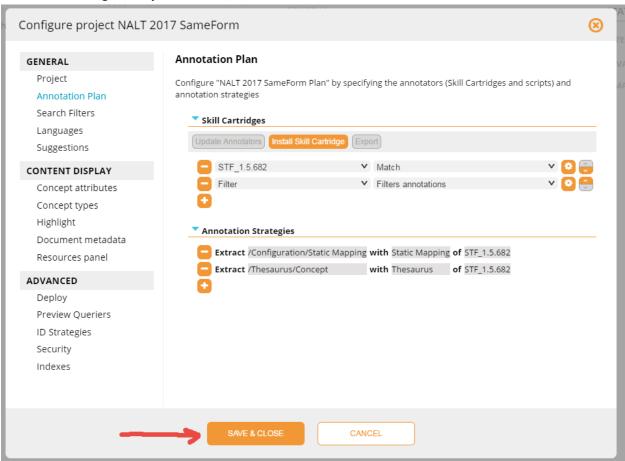
Term selection procedure: Z score selector setting at -2. See Z-score on Wikipedia for information on the population of terms that are extracted.





51. Don't forget to SAVE the parameters!

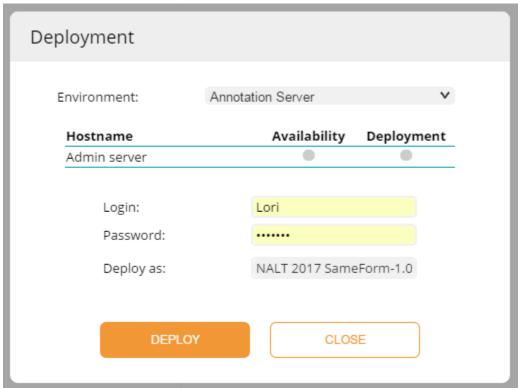
52. Back at the Configure Project window, click on SAVE & CLOSE.



- 53. Before continuing, make sure your configuration is completely saved. The "multiple spinning colored dots" should not be present.
- 54. Now, for this thesaurus to be active on the server, **this project must be deployed**, with a new version number/name, to environments in both the WebStudio as well as the Annotation Server.
- 55. Go to the gray project pull down menu:

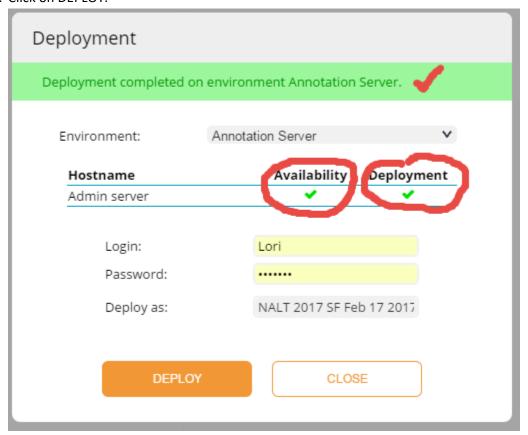


56. Use the pull down menu and select "Deploy". The "Deployment" window will open:



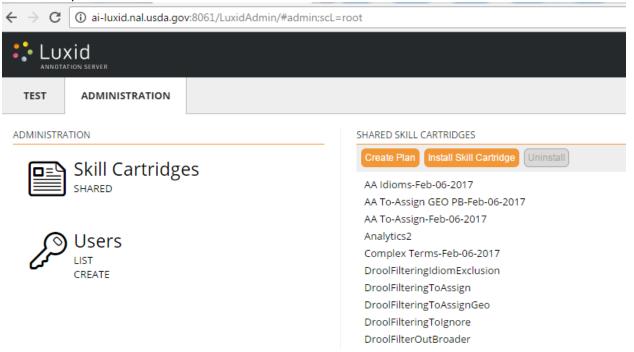
- 57. We have previously set up the deployment environments (see separate instructions for this procedure, but once setup you do not have to redo unless you change servers. Our current environments are "Annotation Server" and "My Web Studio". A new thesaurus will need to be at least deployed to the "Annotation Server" so that Luxid Annotation associated with our pipeline will use this new resource.
- 58. Select your "environment", login/password, and name the new thesaurus Skill Cartridge. Naming should use "NALT" and reflect date, e.g. NALT 2017 SF Feb 17 2017.

### 59. Click on DEPLOY.

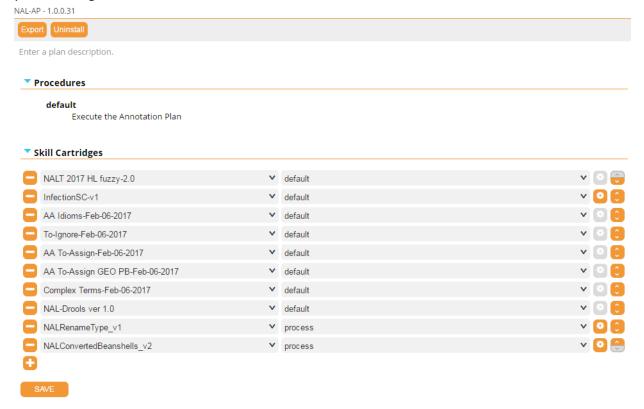


- 60. Availability is checked first, to see if the server is available and the Deployment is checked when the deployment is successful.
- 61. Deploy resource to "My Web Studio", repeating steps above except choosing "My Web Studio" for the environment. Close this window.
- 62. Next we need to **add this new thesaurus in the main Annotation Plan** in the Annotation Server (as well as in any annotation plan in the WebStudio if desired for testing!)
- 63. Go to LuxidAdmin at ai-luxid.nal.usda.gov:8061/LuxidAdmin.

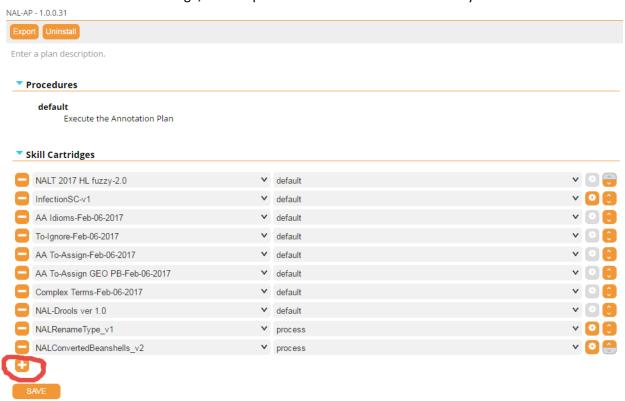
64. A list of SHARED SKILL CARTRIDGES will appear. This list contains skill cartridges, scripts, and annotation plans in one list:



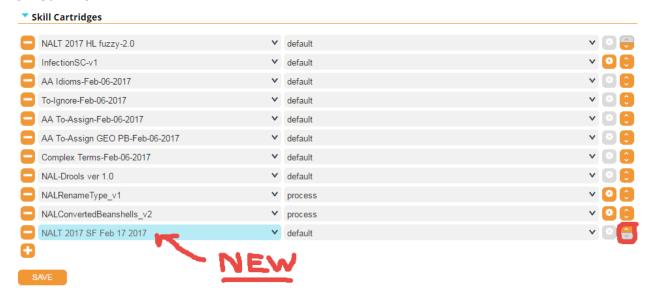
65. The annotation plan that we normally use for production will be NAL-AP. You just need to update the contents of NAL-AP with the new thesaurus. Click on NAL-AP and a new panel will open to the right:



66. To add the new skill cartridge, use the plus at the bottom of the list to add your new thesaurus:



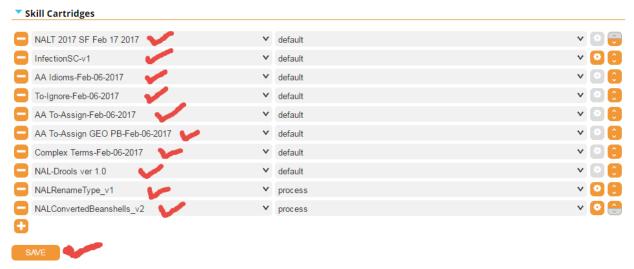
67. Use the pull down menu to select the name of the new Thesaurus skill cartridge, e.g. NALT 2017 SF Feb 17 2017.



68. Use the Orange "up arrow" on the right to move your Thesaurus to the top. Order of the skill cartridges in the plan DOES MATTER. The thesaurus must be the first item on the list.

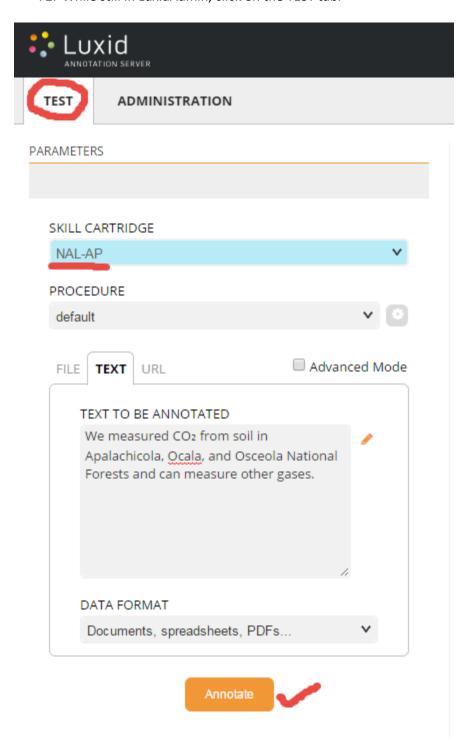


69. Use the Orange "- button" on the left to remove the old Thesaurus from the Annotation Plan.



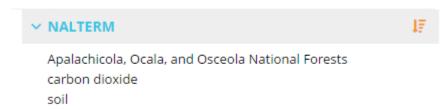
70. Check that all components are in the order as above. Then SAVE.

- 71. You are now ready to **test your new STF** using the TEST.
- 72. While still in LuxidAdmin, click on the TEST tab.



- 73. Select the NAL-AP for the "Skill Cartridge" from the pulldown menu.
- 74. Enter your Text that you want to test (use the "Good Sentences to test" under the Q:\Knowledge Files folder).

- 75. Click on ANNOTATE.
- 76. The NALTERMs that are annotated will appear in a right panel.



77. The center panel shows you the text you annotated with highlighting on extraction points.

A4BA9897CFEF5369EB9C042855B77BB6



78. Compare these results to the results expected in "Good sentences to test".