



Type here to search This Folder

Address Book Options Log Off

Reply Reply to All Forward Move Delete Junk Close

Re: Text Mining Evaluation Work

Dooley, Damion

You replied on 3/9/2018 8:04 AM.

Sent: Thursday, March 08, 2018 11:40 PM**To:** Gosal, Gurinder; ega12@sfu.ca; Duan, Jun; Fornika, Dan; Alghamdi, Dalia**Cc:** Hsiao, William; Macdonald, Kim - BCCDC

So the spreadsheet onto-lookup quickie app is finished :

<http://genepio.org/geem/validator/>

It enables us to see if selected FOODON item is the right one. Lookup ignores free text words. Its just focusing on the ontology identifiers.

ONE PROVISO: OLS may have a slightly different version of a term's label due to FoodOn updates I've made in last few months, e.g. The "different components" column might show "asiatic pennywort" whereas the OLS lookup shows "ASIATIC PENNYWORT PLANT AS FOOD SOURCE" - one can ignore the "as food source" part, the semantics are the same. When difference is a problem is e.g. when sample description "meat" term was matched to FoodOn FOODON_03317626: MEAT (CURED, COMMINUTED) - that's a problem since original sample said nothing about whether meat was cured or not.

Very brief instructions: Copy the row x column blocks you want to review into the spreadsheet app (in my case I did a few hundred rows x first 7 columns of the tricky component match entries, and pasted them into row 0 column A cell. It can take a few seconds for all the entries to paste in. You should see something like:

Not Secure genepio.org/geem/validator/

Apps Regex Tester mail Personal Ontology DNA EXTRACT OCR Bioinformatics Python NML EXTRACT Other Bookmarks

Demonstration spreadsheet with Excell cut&paste and ontology lookup on Ontology ID references

row	A	B	C	D	E	F	G	H	I
0	Encrypt_60	RTE deli salads containing meat - chicken salad	ready to eat deli salad containing meat chicken salad	{'chicken salad:FOODON_03306092', 'chicken meat product:FOODON_00001040', 'ready:Quality', 'deli:Structure-OR-Area]}	Component Match	{'Abbreviation-Acronym Treatment', 'Suffix (Product) Treatment to Input', 'Using Semantic Tagging Resources', 'Infection Treatment'}	Component1-> ready:Quality, Component2-> deli:Structure-OR-Area, Component3-> chicken salad:FOODON_03306092, Component4-> chicken meat		
1	Encrypt_61	rimco mix	rimco mixture	{'mixture:CHEBI_60004', 'snack:FOODON_03316370'}	Component Match	{'Abbreviation-Acronym Treatment', 'Synonym Usage'}	Component1-> snack:FOODON_03316370, Component2-> mixture:CHEBI_60004		
2	Encrypt_62	mamey, fcz	mamey frozen	{'mamey sapote:FOODON_03414238', 'frozen:FOODON_03470136'}	Component Match	{'Abbreviation-Acronym Treatment', 'Synonym Usage'}	Component1-> frozen:Process_FOODON_03470136, Component2-> mamey sapote:FOODON_03414238		
3	Encrypt_63	pennywort juice, fcz	pennywort juice frozen	{'juice:FOODON_03420300', 'asiatic pennywort:FOODON_03414238'}	Component Match	{'Abbreviation-Acronym Treatment', 'Synonym Usage'}	Component1-> juice:FOODON_03420300, Component2-> asiatic pennywort:FOODON_03414238		

Paste from Excel-compatible:

Copy a range of cells to clipboard in Excel
Select a cell on background
Use Ctrl-V keyboard shortcut to paste from the clipboard
Add rows to bottom of grid if paste overflows

Ontology Lookup:

Click on a cell to perform OLS ontology lookup of any [ONTLOGY_ID] e.g. FOODON_03414238 found in text. Each term lookup will be given its own cell to left.

Then navigate to the cell that has ontology ids to lookup – in this case, column G, and click on a cell. The lookup should take a second, and will be presented to right of cell you clicked on.

Demonstration spreadsheet with Excell cut&paste and ontology lookup on Ontology ID references

D	E	F	G	H	I	J	K
{'chicken salad:FOODON_03306092', 'chicken meat product:FOODON_00001040', 'ready:Quality', 'deli:Structure-OR-Area']}	Component Match	{'Abbreviation-Acronym Treatment', 'Suffix (Product) Treatment to Input', 'Using Semantic Tagging Resources', 'Infection Treatment'}	Component1-> ready:Quality, Component2-> deli:Structure-OR-Area, Component3-> chicken salad:FOODON_03306092, Component4-> chicken meat	FOODON_00001040: CHICKEN MEAT FOOD PRODUCT:	FOODON_00001040: CHICKEN SALAD:		
{'mixture:CHEBI_60004', 'snack:FOODON_03316370'}	Component Match	{'Abbreviation-Acronym Treatment', 'Synonym Usage'}	Component1-> snack:FOODON_03316370, Component2-> mixture:CHEBI_60004	Lookup unavailable.	CHEBI_60004: MIXTURE: A mixture is a chemical substance composed of multiple molecules, at least two of which are of a different kind		
{'mamey sapote:FOODON_03414238', 'frozen:FOODON_03470136'}	Component Match	{'Abbreviation-Acronym Treatment', 'Synonym Usage'}	Component1-> frozen:Process_FOODON_03470136, Component2-> mamey sapote:FOODON_03414238	FOODON_03414238: MAMEY SAPOTE PLANT AS FOOD SOURCE:	FOODON_03414238: PRESERVATION BY FREEZING: preservation by freezing a food product and keeping it at a temperature below the freezing point (20-30 degrees F) without		
{'juice:FOODON_03420300', 'asiatic pennywort:FOODON_03414238'}	Component Match	{'Abbreviation-Acronym Treatment', 'Synonym Usage'}	Component1-> juice:FOODON_03420300				

Hopefully this makes it a lot easier to review stuff...

Now I see this web spreadsheet software is powerful and I can immediately see a hundred things more we should add to the app to dive into the ontology validation challenge!

D.

From: Damion Dooley <damion.dooley@bccdc.ca>**Date:** Thursday, March 8, 2018 at 1:15 PM**To:** "Gosal, Gurinder" <Gurinder.Gosal@bccdc.ca>, Emma Griffiths <ega12@sfu.ca>, "Duan, Jun" <Jun.Duan@bccdc.ca>, "Fornika, Dan" <Dan.Fornika@bccdc.ca>, "Alghamdi, Dalia" <Dalia.Alghamdi@bccdc.ca>**Cc:** William Hsiao <William.Hsiao@bccdc.ca>, "Macdonald, Kim - BCCDC" <kim.macdonald@bccdc.ca>**Subject:** Re: Text Mining Evaluation Work

I'm going to do a little app that will take in selected spreadsheet columns and will for columns like RetainedSet_MappedTerms_with_Resource_IDs And

Different Components(In case of Component Match) , will go lookup all the ontology labels for them via OLS and will report them in a separate column to left.

I'll let you all know when its done - it should speed up review of the terms!

I'm doing this because I have to research browser based spreadsheet apps anyways today...

D.

On 2018-03-07, 3:48 PM, "Gosal, Gurinder" <Gurinder.Gosal@bccdc.ca> wrote:

Dear Evaluators,

I know you all are already doing a great job with the evaluation of the results 😊 and have also highlighted few points in between. It would be great if you can wind up this task in a week's time (say by next Friday). This would greatly help to have a comprehensive analysis of the evaluations results/ issues and speed up the subsequent work.

Thanks,
Gurinder.

