**Concept Frequency Program Documentation**

The Concept Frequency program is a GUI application that takes text as input and outputs an excel file. It takes as input a text file or text that's been copy-n-pasted into the input text area and outputs an excel spreadsheet that contains a list of concepts extracted and their associated frequency count. The program also outputs the total number of extracted concepts and the concept return ratio (total number of extracted concepts divided by that plus the number of poolparty terms found).

The program is written in java swing. The GUI can be edited using the google GWT software to accomplish faster programming. There are several external libraries it uses; notably the SPARQL and Apache http api. All libraries are included in TRACS.

The program does it's computation when the "Extract and import to Excel" button is pressed. First, it retrieves the text. If 'textArea', a JTextArea variable, is not empty it will get the text from there, otherwise it will get it from 'docString', the extracted string from the imported txt file. 'DocString' is extracted by using the java FileInputStream wrapper on the imported txt file. Next, a SupplierProfileCreator object is created. The SupplierProfileCreator object has the functions needed to retrieve concepts from the poolparty extractor (PPX). PPX is an api that works via HTTP requests. Documentation for the PPX api can be found on the poolparty's website. Next, the SupplierProfileCreator object creates a WebRetriever object that queries poolparty. Poolparty delivers back an xml string. The xml string is then parsed to get the skos:terms, skos:concepts, and skos:hasTopConcept fields; from that the list of concepts are saved. Next, the frequency of each concept is determined. A SparqlRetriever object is created that performs the following SPARQL query for every concept on the poolparty server:

PREFIX skos:<http://www.w3.org/2004/02/skos/core#>

SELECT distinct ?altLabelAndHiddenLabel

WHERE {

?x skos:prefLabel concept.

{ ?x skos:altLabel ?altLabelAndHiddenLabel. }

UNION { ?x skos:hiddenLabel ?altLabelAndHiddenLabel. }

}

Note: The poolparty SPARQL endpoint ( http://infoneer.poolparty.biz/PoolParty/sparql/Processes ) provides a GUI interface to test SPARQL queries. Sample queries are also given to help reduce the learning curve.

After the server is queried it returns a html table of alterate and hidden labels for every concept. So, given the prefLabel, alterate, and hidden labels the program scans through the original input text and finds the regular expression matching anyone of those for a given concept. The total number of matches equals the frequency count of that concept.

Finally, an xml file is generated using the JXL api. It creates two columns. The first column is the ppxFoundConcepts array and the second is the ppxConceptsFreq array.

The source code and libraries used in the program can be found on TRACS.