**Disease-Chemical/Drug association from CTD and NDF-RT**

**Comparative Toxicogenomics Database (CTD)**

Download the latest version from CTD <http://ctdbase.org/downloads/#cd>.

Delete the introduction lines.

Retrieve only chemical name (column 1), disease name (column 4) and disease ID (MeSH or OMIM identifier).  *$ cut -f1,4,5 INPUT\_FILE > INPUT\_FILE\_selectedColumns*

Identify unique concept identifier (CUI) for chemicals/drugs from the chemicals/drugs lexicon previously compiled from three resources namely UMLS Metathesaurus, DrugBank and PharmGKB (refer Project: DrugDict/DrugsDictionaryCompiler).

*$ javac DrugnamToLexicondrugidMapper.java*

*$ java DrugnamToLexicondrugidMapper INPUT\_FILE\_selectedColumns OUTPUT\_FILE1*

All chemicals/drugs may not have equivalent CUIs. Remove the records without CUI.

*$ awk -F"\t" '$4 != "null" { print $1 "\t" $2 "\t" $3 "\t" $4 }' OUTPUT\_FILE1 > OUTPUT\_FILE2*

Replace Mesh or OMIM ID with CUI for disease mentions.

*$ javac Conceptname2CUIMapper.java*

*$ java Conceptname2CUIMapper OUTPUT\_FILE2 OUTPUT\_FILE3*

MeSH or OMIM ID may not match with CUI for some disease synonyms. Remove the records without CUI.

*$ awk -F"\t" '$4 != "null" { print $1 "\t" $2 "\t" $3 "\t" $4 }' OUTPUT\_FILE3 >* ***OUTPUT\_FILE\_CTD***

**National Drug File – Reference Terminology (NDF-RT)**

We used MRREL.RRF file within UMLS Metathesaurus to retrieve NDF-RT records.

*$ grep 'NDFRT' MRREL.RRF > OUTPUT\_FILE*

Filter the disease-drug association records with annotation as ‘may treat’.

*$ grep 'may\_treat' OUTPUT\_FILE > OUTPUT\_FILE\_filtered*

Retrieve a unique list of disease-drug associations.

*$ cut -f1,5 -d'|' OUTPUT\_FILE\_filtered > OUTPUT\_FILE\_filtered\_uniquelist*

Retrieve a unique list of CUI and concepts from previously processed file from UMLS Metathesaurus (refer Project: ConceptMap/UMLSMetathesaurusCompiler).

*$ cut -f1,15 -d'|' mrconso.preferredVocabularies > mrconso.preferredVocabularies\_CUIAndConcept*

The associations from NDFRT are represented only with CUI. Map CUI to respective concept name in UMLS Metathesaurus.

*$ javac cui2ConceptnameMapper.java*

*$ java cui2ConceptnameMapper mrconso.preferredVocabularies\_CUIAndConcept OUTPUT\_FILE\_filtered\_uniquelist OUTPUT\_FILE1*

Map the chemicals / drugs to customized ID in the chemicals / drugs lexicon (refer Project: DrugDict/DrugsDictionaryCompiler).

*$ javac DrugnameToLexicondrugidMapper.java*

*$ java DrugnameToLexicondrugidMapper OUTPUT\_FILE1 OUTPUT\_FILE2*

Remove the records that do not map to customized ID in the chemicals / drugs lexicon.

*$ awk -F"\t" '$5 != "null" { print $1 "\t" $2 "\t" $3 "\t" $4 "\t" $5 }' OUTPUT\_FILE2 > OUTPUT\_FILE3*

Reorder the columns in the output file to match with the output file from CTD.

*$ awk -F"\t" ' { print $4 "\t" $2 "\t" $5 "\t" $1 }' OUTPUT\_FILE3 >* ***OUTPUT\_FILE\_NDFRT***

**Disease-chemical/drug associations from CTD and NDF-RT**

Combine the output from both resources.

*$ cat OUTPUT\_FILE\_CTD OUTPUT\_FILE\_NDFRT >* ***OUTPUT\_FILE\_CTD\_NDFRT***

Retrieve a unique list of disease-chemical/drug associations based on customized chemical/drug ID and disease CUI.

*$ javac DiseaseChemicalORDrugAssociationUniqueList.java*

*$ java DiseaseChemicalORDrugAssociationUniqueList OUTPUT\_FILE\_CTD\_NDFRT* ***OUTPUT\_FILE\_CTD\_NDFRT\_unique***

**Disease-drugs only associations from CTD and NDF-RT**

Filter a list of disease-drugs only associations by defining drugs as the concepts belonging to UMLS semantic types clinical drugs, antibiotics and Pharmacologically active substances.

*$ javac DrugsOnlyFilterer.java*

*$ java DrugsOnlyFilterer OUTPUT\_FILE\_CTD\_NDFRT\_unique* ***OUTPUT\_FILE\_CTD\_NDFRT\_unique\_drugsOnly***

**Summary (Sep 2017)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **# of associations** | **# of diseases** | **# of chemicals/drugs** | **# of drugs only** |
| Disease-Chemicals/drugs | 2,079,755 | 5,823 | 23,102 | - |
| Disease-Drugs only | 1,213,069 | 5,720 | - | 16,326 |