**Comparison between countA.txt and countB.txt**

**Test Value #1 #2 #1-#2**

**------------------------ ------------------------------------------------------ --------- --------- ---------**

codeCount 388944 345373 43571

\*\*\*\* While the count for Content codes matches, certain Anonymous codes and Metadata codes are missing from the DB.

**codeTally null,Y,N,Anonymous 299439 255909 43530**

**codeTally null,Y,N,Metadata 171 130 41**

**nameCount 447712 702940 -255228**

\*\*\*\* Counts for Content names match, but Metadata names are still not being loaded properly.

**nameTally null,N,en,PT,FULL\_SYN,NCI\_Thesaurus,Metadata 178 178**

**nameTally null,N,en,SY,FULL\_SYN,NCI\_Thesaurus,Metadata 10 10**

**nameTally null,N,en,null,Display\_Name,NCI\_Thesaurus,Metadata 156 156**

**nameTally null,N,en,null,label,NCI\_Thesaurus,Metadata 169 169**

**nameTally null,Y,en,null,Preferred\_Name,NCI\_Thesaurus,Metadata 168 168**

\*\*\*\* "Bottom Thing" + 255908 labels probably correspond to 255909 additional Anonymous codes in DB.

**nameTally null,Y,en,null,Bottom Thing,NCI\_Thesaurus,Anonymous 1 -1**

**nameTally null,Y,en,null,label,NCI\_Thesaurus,Anonymous 255908 -255908**

\*\*\*\* Comment counts do not match.

\*\*\*\* TODO: sample comments and compare.

**commentCount 4307 4286 21**

\*\*\*\* Definition counts match for Content but Metadata definitions are still not being loaded.

**defCount 72242 72074 168**

**defTally NCI\_Thesaurus,Metadata 168 168**

propCount 852811 716645 136166

\*\*\*\* All of the Content property counts match with the exception of “type.”

**propTally en,type,NCI\_Thesaurus,Content 92932 92932**

\*\*\*\* "type" is not being loaded into the DB. Is this something we added to file counting previously to match the DB?

**propTally en,Contributing\_Source,NCI\_Thesaurus,Metadata 11 11**

**propTally en,Semantic\_Type,NCI\_Thesaurus,Metadata 163 163**

**propTally en,Unit,NCI\_Thesaurus,Metadata 2 2**

**propTally en,P106,NCI\_Thesaurus,Metadata 106 -106**

**propTally en,P107,NCI\_Thesaurus,Metadata 101 -101**

**propTally en,P108,NCI\_Thesaurus,Metadata 108 -108**

**propTally en,P90,NCI\_Thesaurus,Metadata 108 -108**

**propTally en,P97,NCI\_Thesaurus,Metadata 109 -109**

**propTally en,P98,NCI\_Thesaurus,Metadata 18 -18**

**propTally en,label,NCI\_Thesaurus,Metadata 109 -109**

**propTally en,domain,NCI\_Thesaurus,Metadata 80 83 -3**

**propTally en,range,NCI\_Thesaurus,Metadata 80 83 -3**

**propTally en,subPropertyOf,NCI\_Thesaurus,Metadata 17 17 0**

**propTally en,type,NCI\_Thesaurus,Metadata 321 128 193**

\*\* 106,107,108 are display name, preferred name, sty -

\*\* definition, design note, full syn

\*\* BUG: so metadata properties are being loaded but incorrectly.

**propTally en,type,NCI\_Thesaurus,Anonymous 299439 255908 43531**

\*\*\*\* The Content propProp count matches except for “label”. Metadata propProps (attached to Metadata properties)

are not loaded either.

propPropCount 397060 767135 -370075

\*\*\*\* These labels are being added in the DB. This is currently done for browser functionality but it is still a bug.

**propPropTally Accepted\_Therapeutic\_Use\_For,qualifier,label,Content 556 -556**

**propPropTally BioCarta\_ID,qualifier,label,Content 335 -335**

**propPropTally CAS\_Registry,qualifier,label,Content 10889 -10889**

**propPropTally Chemical\_Formula,qualifier,label,Content 9023 -9023**

**propPropTally Concept\_Status,qualifier,label,Content 4069 -4069**

**propPropTally Contributing\_Source,qualifier,label,Content 23812 -23812**

**propPropTally EntrezGene\_ID,qualifier,label,Content 2595 -2595**

**propPropTally Essential\_Amino\_Acid,qualifier,label,Content 10 -10**

**propPropTally Essential\_Fatty\_Acid,qualifier,label,Content 6 -6**

**propPropTally Extensible\_List,qualifier,label,Content 130 -130**

**propPropTally FDA\_Table,qualifier,label,Content 1820 -1820**

**propPropTally FDA\_UNII\_Code,qualifier,label,Content 12023 -12023**

**propPropTally GO\_Annotation,qualifier,label,Content 6350 -6350**

**propPropTally GenBank\_Accession\_Number,qualifier,label,Content 2733 -2733**

**propPropTally Gene\_Encodes\_Product,qualifier,label,Content 1975 -1975**

**propPropTally Homologous\_Gene,qualifier,label,Content 205 -205**

**propPropTally ICD-O-3\_Code,qualifier,label,Content 1062 -1062**

**propPropTally INFOODS,qualifier,label,Content 203 -203**

**propPropTally Image\_Link,qualifier,label,Content 6 -6**

**propPropTally In\_Clinical\_Trial\_For,qualifier,label,Content 837 -837**

**propPropTally KEGG\_ID,qualifier,label,Content 237 -237**

**propPropTally Legacy\_Concept\_Name,qualifier,label,Content 72308 -72308**

**propPropTally MGI\_Accession\_ID,qualifier,label,Content 154 -154**

**propPropTally Macronutrient,qualifier,label,Content 23 -23**

**propPropTally Micronutrient,qualifier,label,Content 32 -32**

**propPropTally NCBI\_Taxon\_ID,qualifier,label,Content 562 -562**

**propPropTally NCI\_META\_CUI,qualifier,label,Content 11571 -11571**

**propPropTally NSC\_Code,qualifier,label,Content 2043 -2043**

**propPropTally Neoplastic\_Status,qualifier,label,Content 7614 -7614**

**propPropTally Nutrient,qualifier,label,Content 55 -55**

**propPropTally OID,qualifier,label,Content 7 -7**

**propPropTally OLD\_ASSOCIATION,qualifier,label,Content 8 -8**

**propPropTally OLD\_CHILD,qualifier,label,Content 2659 -2659**

**propPropTally OLD\_KIND,qualifier,label,Content 3024 -3024**

**propPropTally OLD\_PARENT,qualifier,label,Content 3533 -3533**

**propPropTally OLD\_ROLE,qualifier,label,Content 524 -524**

**propPropTally OLD\_STATE,qualifier,label,Content 7 -7**

**propPropTally OMIM\_Number,qualifier,label,Content 7619 -7619**

**propPropTally PDQ\_Closed\_Trial\_Search\_ID,qualifier,label,Content 3194 -3194**

**propPropTally PDQ\_Open\_Trial\_Search\_ID,qualifier,label,Content 3194 -3194**

**propPropTally PID\_ID,qualifier,label,Content 169 -169**

**propPropTally PubMedID\_Primary\_Reference,qualifier,label,Content 387 -387**

**propPropTally Relative\_Enzyme\_Activity,qualifier,label,Content 1 -1**

**propPropTally SNP\_ID,qualifier,label,Content 23 -23**

**propPropTally Semantic\_Type,qualifier,label,Content 95660 -95660**

**propPropTally Subsource,qualifier,label,Content 1261 -1261**

**propPropTally Swiss\_Prot,qualifier,label,Content 2340 -2340**

**propPropTally Tolerable\_Level,qualifier,label,Content 27 -27**

**propPropTally UMLS\_CUI,qualifier,label,Content 72561 -72561**

**propPropTally USDA\_ID,qualifier,label,Content 134 -134**

**propPropTally US\_Recommended\_Intake,qualifier,label,Content 43 -43**

**propPropTally Unit,qualifier,label,Content 330 -330**

**propPropTally Use\_For,qualifier,label,Content 341 -341**

**propPropTally miRBase\_ID,qualifier,label,Content 147 -147**

\*\*\*\* These source propProps are attached to presentations and definitions that are not being loaded.

**propPropTally definition,source,CRCH,Metadata 5 5**

**propPropTally presentation,source,NCI,Metadata 178 178**

**propPropTally definition,source,NCI,Metadata 163 163**

**propPropTally presentation,source,CRCH,Metadata 10 10**

propPropTypeCount 8 8 0

propTypeCount 74 163 -89

relCount 410708 992397 -581689

\*\*\*\* The Content relationships counts match. The following are the exceptions.

**relTally NCI\_Thesaurus,disjointWith,Content 196 -196**

**relTally NCI\_Thesaurus,subClassOf,Content 82326 99736 -17410**

\*\*\*\* disjointWith is not being counted in the file.

\*\*\*\* TODO (jfw): sample subClassOf rels to see what is in the DB.

\*\*\*\* The Anonymous and Metadata relationships are not being counted in the file, only in the DB.

We have not yet determined a good way to count Anonymous rels in the file. They are being generated by the loader.

**relTallyNCI\_Thesaurus,Allele\_Absent\_From\_Wild-type\_Chromosomal\_Location,Anonymous 1 -1**

**relTally NCI\_Thesaurus,Allele\_Associated\_With\_Disease,Anonymous 10 -10**

**relTally NCI\_Thesaurus,Allele\_Has\_Abnormality,Anonymous 182 -182**

**relTally NCI\_Thesaurus,Allele\_Has\_Activity,Anonymous 76 -76**

**relTally NCI\_Thesaurus,Allele\_In\_Chromosomal\_Location,Anonymous 106 -106**

**relTally NCI\_Thesaurus,Allele\_Not\_Associated\_With\_Disease,Anonymous 5 -5**

**relTallyNCI\_Thesaurus,Allele\_Plays\_Altered\_Role\_In\_Process,Anonymous 128 -128**

**relTallyNCI\_Thesaurus,Allele\_Plays\_Role\_In\_Metabolism\_Of\_Chemical\_Or\_Drug,Anonymous 3 -3**

**relTally NCI\_Thesaurus,Anatomic\_Structure\_Has\_Location,Anonymous 301 -301**

**relTallyNCI\_Thesaurus,Anatomic\_Structure\_Is\_Physical\_Part\_Of,Anonymous 7682 -7682**

**relTallyNCI\_Thesaurus,Biological\_Process\_Has\_Associated\_Location,Anonymous 512 -512**

**relTallyNCI\_Thesaurus,Biological\_Process\_Has\_Initiator\_Chemical\_Or\_Drug,Anonymous 109 -109**

**relTallyNCI\_Thesaurus,Biological\_Process\_Has\_Initiator\_Process,Anonymous 125 -125**

**relTallyNCI\_Thesaurus,Biological\_Process\_Has\_Result\_Anatomy,Anonymous 62 -62**

**relTallyNCI\_Thesaurus,Biological\_Process\_Has\_Result\_Biological\_Process,Anonymous 87 -87**

**relTallyNCI\_Thesaurus,Biological\_Process\_Has\_Result\_Chemical\_Or\_Drug,Anonymous 32 -32**

**relTallyNCI\_Thesaurus,Biological\_Process\_Is\_Part\_Of\_Process,Anonymous 466 -466**

**relTallyNCI\_Thesaurus,Chemical\_Or\_Drug\_Affects\_Abnormal\_Cell,Anonymous 28 -28**

**relTallyNCI\_Thesaurus,Chemical\_Or\_Drug\_Affects\_Cell\_Type\_Or\_Tissue,Anonymous 263 -263**

**relTallyNCI\_Thesaurus,Chemical\_Or\_Drug\_Affects\_Gene\_Product,Anonymous 328 -328**

**relTallyNCI\_Thesaurus,Chemical\_Or\_Drug\_Has\_Mechanism\_Of\_Action,Anonymous 1007 -1007**

**relTallyNCI\_Thesaurus,Chemical\_Or\_Drug\_Has\_Physiologic\_Effect,Anonymous 1294 -1294**

**relTallyNCI\_Thesaurus,Chemical\_Or\_Drug\_Is\_Metabolized\_By\_Enzyme,Anonymous 607 -607**

**relTallyNCI\_Thesaurus,Chemical\_Or\_Drug\_Plays\_Role\_In\_Biological\_Process,Anonymous 18 -18**

**relTally NCI\_Thesaurus,Chemotherapy\_Regimen\_Has\_Component,Anonymous 10798 -10798**

**relTally NCI\_Thesaurus,Conceptual\_Part\_Of,Anonymous 345 -345**

**relTallyNCI\_Thesaurus,Cytogenetic\_Abnormality\_Involves\_Chromosome,Anonymous 516 -516**

**relTally NCI\_Thesaurus,Disease\_Excludes\_Abnormal\_Cell,Anonymous 12228 -12228**

**relTallyNCI\_Thesaurus,Disease\_Excludes\_Cytogenetic\_Abnormality,Anonymous 402 -402**

**relTally NCI\_Thesaurus,Disease\_Excludes\_Finding,Anonymous 9179 -9179**

**relTallyNCI\_Thesaurus,Disease\_Excludes\_Molecular\_Abnormality,Anonymous 337 -337**

**relTally NCI\_Thesaurus,Disease\_Excludes\_Normal\_Cell\_Origin,Anonymous 3965 -3965**

**relTallyNCI\_Thesaurus,Disease\_Excludes\_Normal\_Tissue\_Origin,Anonymous 705 -705**

**relTallyNCI\_Thesaurus,Disease\_Excludes\_Primary\_Anatomic\_Site,Anonymous 2635 -2635**

**relTally NCI\_Thesaurus,Disease\_Has\_Abnormal\_Cell,Anonymous 33933 -33933**

**relTallyNCI\_Thesaurus,Disease\_Has\_Associated\_Anatomic\_Site,Anonymous 24407 -24407**

**relTally NCI\_Thesaurus,Disease\_Has\_Associated\_Disease,Anonymous 817 -817**

**relTally NCI\_Thesaurus,Disease\_Has\_Cytogenetic\_Abnormality,Anonymous 246 -246**

**relTally NCI\_Thesaurus,Disease\_Has\_Finding,Anonymous 19286 -19286**

**relTallyNCI\_Thesaurus,Disease\_Has\_Metastatic\_Anatomic\_Site,Anonymous 168 -168**

**relTally NCI\_Thesaurus,Disease\_Has\_Molecular\_Abnormality,Anonymous 1086 -1086**

**relTally NCI\_Thesaurus,Disease\_Has\_Normal\_Cell\_Origin,Anonymous 13513 -13513**

**relTally NCI\_Thesaurus,Disease\_Has\_Normal\_Tissue\_Origin,Anonymous 13943 -13943**

**relTally NCI\_Thesaurus,Disease\_Has\_Primary\_Anatomic\_Site,Anonymous 16066 -16066**

**relTally NCI\_Thesaurus,Disease\_Is\_Grade,Anonymous 892 -892**

**relTally NCI\_Thesaurus,Disease\_Is\_Stage,Anonymous 2671 -2671**

**relTally NCI\_Thesaurus,Disease\_Mapped\_To\_Chromosome,Anonymous 181 -181**

**relTally NCI\_Thesaurus,Disease\_Mapped\_To\_Gene,Anonymous 133 -133**

**relTally NCI\_Thesaurus,Disease\_May\_Have\_Abnormal\_Cell,Anonymous 1067 -1067**

**relTally NCI\_Thesaurus,Disease\_May\_Have\_Associated\_Disease,Anonymous 1479 -1479**

**relTallyNCI\_Thesaurus,Disease\_May\_Have\_Cytogenetic\_Abnormality,Anonymous 9063 -9063**

**relTally NCI\_Thesaurus,Disease\_May\_Have\_Finding,Anonymous 13072 -13072**

**relTallyNCI\_Thesaurus,Disease\_May\_Have\_Molecular\_Abnormality,Anonymous 5755 -5755**

**relTally NCI\_Thesaurus,Disease\_May\_Have\_Normal\_Cell\_Origin,Anonymous 125 -125**

**relTally NCI\_Thesaurus,EO\_Disease\_Has\_Associated\_Cell\_Type,Anonymous 13 -13**

**relTallyNCI\_Thesaurus,EO\_Disease\_Has\_Associated\_EO\_Anatomy,Anonymous 3104 -3104**

**relTallyNCI\_Thesaurus,EO\_Disease\_Has\_Property\_Or\_Attribute,Anonymous 273 -273**

**relTally NCI\_Thesaurus,EO\_Disease\_Maps\_To\_Human\_Disease,Anonymous 1811 -1811**

**relTally NCI\_Thesaurus,Gene\_Associated\_With\_Disease,Anonymous 2391 -2391**

**relTally NCI\_Thesaurus,Gene\_Found\_In\_Organism,Anonymous 2948 -2948**

**relTally NCI\_Thesaurus,Gene\_Has\_Abnormality,Anonymous 181 -181**

**relTally NCI\_Thesaurus,Gene\_Has\_Physical\_Location,Anonymous 989 -989**

**relTally NCI\_Thesaurus,Gene\_In\_Chromosomal\_Location,Anonymous 2395 -2395**

**relTallyNCI\_Thesaurus,Gene\_Involved\_In\_Pathogenesis\_Of\_Disease,Anonymous 126 -126**

**relTally NCI\_Thesaurus,Gene\_Is\_Biomarker\_Of,Anonymous 25 -25**

**relTally NCI\_Thesaurus,Gene\_Is\_Biomarker\_Type,Anonymous 39 -39**

**relTally NCI\_Thesaurus,Gene\_Is\_Element\_In\_Pathway,Anonymous 4662 -4662**

**relTally NCI\_Thesaurus,Gene\_Plays\_Role\_In\_Process,Anonymous 17765 -17765**

**relTally NCI\_Thesaurus,Gene\_Product\_Encoded\_By\_Gene,Anonymous 8919 -8919**

**relTally NCI\_Thesaurus,Gene\_Product\_Expressed\_In\_Tissue,Anonymous 989 -989**

**relTally NCI\_Thesaurus,Gene\_Product\_Has\_Abnormality,Anonymous 18 -18**

**relTally NCI\_Thesaurus,Gene\_Product\_Has\_Associated\_Anatomy,Anonymous 1357 -1357**

**relTallyNCI\_Thesaurus,Gene\_Product\_Has\_Biochemical\_Function,Anonymous 5115 -5115**

**relTallyNCI\_Thesaurus,Gene\_Product\_Has\_Chemical\_Classification,Anonymous 612 -612**

**relTally NCI\_Thesaurus,Gene\_Product\_Has\_Malfunction\_Type,Anonymous 8 -8**

**relTally NCI\_Thesaurus,Gene\_Product\_Has\_Organism\_Source,Anonymous 909 -909**

**relTallyNCI\_Thesaurus,Gene\_Product\_Has\_Structural\_Domain\_Or\_Motif,Anonymous 897 -897**

**relTally NCI\_Thesaurus,Gene\_Product\_Is\_Biomarker\_Of,Anonymous 29 -29**

**relTally NCI\_Thesaurus,Gene\_Product\_Is\_Biomarker\_Type,Anonymous 268 -268**

**relTally NCI\_Thesaurus,Gene\_Product\_Is\_Element\_In\_Pathway,Anonymous 4671 -4671**

**relTally NCI\_Thesaurus,Gene\_Product\_Is\_Physical\_Part\_Of,Anonymous 342 -342**

**relTallyNCI\_Thesaurus,Gene\_Product\_Malfunction\_Associated\_With\_Disease,Anonymous 506 -506**

**relTallyNCI\_Thesaurus,Gene\_Product\_Plays\_Role\_In\_Biological\_Process,Anonymous 10803 -10803**

**relTally NCI\_Thesaurus,Molecular\_Abnormality\_Involves\_Gene,Anonymous 373 -373**

**relTallyNCI\_Thesaurus,Procedure\_Has\_Completely\_Excised\_Anatomy,Anonymous 92 -92**

**relTally NCI\_Thesaurus,Procedure\_Has\_Excised\_Anatomy,Anonymous 404 -404**

**relTally NCI\_Thesaurus,Procedure\_Has\_Imaged\_Anatomy,Anonymous 1 -1**

**relTallyNCI\_Thesaurus,Procedure\_Has\_Partially\_Excised\_Anatomy,Anonymous 192 -192**

**relTally NCI\_Thesaurus,Procedure\_Has\_Target\_Anatomy,Anonymous 645 -645**

**relTallyNCI\_Thesaurus,Procedure\_May\_Have\_Completely\_Excised\_Anatomy,Anonymous 6 -6**

**relTally NCI\_Thesaurus,Procedure\_May\_Have\_Excised\_Anatomy,Anonymous 8 -8**

**relTallyNCI\_Thesaurus,Procedure\_May\_Have\_Partially\_Excised\_Anatomy,Anonymous 9 -9**

**relTallyNCI\_Thesaurus,Regimen\_Has\_Accepted\_Use\_For\_Disease,Anonymous 305 -305**

**relTally NCI\_Thesaurus,equivalentClass,Anonymous 11614 -1161**

**relTally NCI\_Thesaurus,subClassOf,Anonymous 261724 -261724**

**relTally NCI\_Thesaurus,domain,Metadata 80 -80**

**relTally NCI\_Thesaurus,range,Metadata 80 -80**

**relTally NCI\_Thesaurus,subPropertyOf,Metadata 17 -17**

**relTally NCI\_Thesaurus,type,Metadata 121 -121**

\*\*\*\* relProp is not being counted in the file the same way as it is in the DB.

\*\*\*\* These CANNOT be sampled due to the colon but it is not an issue of Annoymous vs. Content.

**relPropCount 282843 573290 -290447**

**relPropTally owl:someValuesFrom,Content 282843 573290 -290447**

relTypeCount 109 130 -21

relPropTypeCount 1 1 0

\*\*\*\* The extra entries are:

> relTypeCount||AllDifferent|

> relTypeCount||args|

> relTypeCount||complementOf|

> relTypeCount||datatype|

> relTypeCount||datatypeValue|

> relTypeCount||differentFrom|

> relTypeCount||disjointWith|

> relTypeCount||domain|

> relTypeCount||equivalentClass|

> relTypeCount||equivalentProperty|

> relTypeCount||inverseOf|

> relTypeCount||maxArgs|

> relTypeCount||minArgs|

> relTypeCount||OLD\_SOURCE\_ASSOCIATION|

> relTypeCount||PAL-DESCRIPTION|

> relTypeCount||PAL-NAME|

> relTypeCount||PAL-STATEMENT|

> relTypeCount||range|

> relTypeCount||sameAs|

> relTypeCount||subPropertyOf|

> relTypeCount||type|

**sourceCount 46 56 -10**

\*\*\*\* The extra entries are ncicp,owl,owl2xml,protege,rdf,rdfs,swrl,swrlb,xsd,xsp.

\*\*\*\* QUESTION: Are these read from the file (i.e., should be counted) or hard coded by the loader?

**ttyCount 16 17 -1**

\*\*\*\* The extra entry is "text\_plain"