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

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

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

**codeCount 408086 358862 49224**

**codeTally null,Y,N,Metadata 176 133 43**

**codeTally null,Y,N,Anonymous 313142 263961 49181**

\*\*\*\* Looks like a good amount of Anonymous and Metadata codes are still not being loaded into the DB.

**commentCount 4911 4890 21**

\*\*\*\* TODO: sample comments

**defCount 81951 81778 173**

**defTally NCI\_Thesaurus,Metadata 173 173**

\*\*\*\* Metadata defs are still not being loaded.

**nameCount 486457 749720 -263263**

**nameTally null,N,en,PT,FULL\_SYN,NCI\_Thesaurus,Metadata 183 183**

**nameTally null,N,en,SY,FULL\_SYN,NCI\_Thesaurus,Metadata 11 11**

**nameTally null,N,en,null,Display\_Name,NCI\_Thesaurus,Metadata 157 157**

**nameTally null,N,en,null,label,NCI\_Thesaurus,Metadata 174 174**

**nameTally null,Y,en,null,Preferred\_Name,NCI\_Thesaurus,Metadata 173 173**

\*\*\*\* Metadata names are still not being loaded.

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

**nameTally null,Y,en,null,label,NCI\_Thesaurus,Anonymous 263960 -263960**

\*\*\*\* "Bottom Thing" + 255908 labels probably correspond to 263961 additional Anonymous codes in DB.

**propCount 905893 758507 147386**

**propTally en,type,NCI\_Thesaurus,Content 98503 98503**

**propTally en,type,NCI\_Thesaurus,Anonymous 313142 263960 49182**

\*\*\*\* "type" is not being loaded into the DB, and incompletely for Anonymous entities

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

**propTally en,Semantic\_Type,NCI\_Thesaurus,Metadata 173 173**

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

**propTally en,domain,NCI\_Thesaurus,Metadata 82 85 -3**

**propTally en,label,NCI\_Thesaurus,Metadata 112 -112**

**propTally en,range,NCI\_Thesaurus,Metadata 82 85 -3**

**propTally en,type,NCI\_Thesaurus,Metadata 330 131 199**

**propTally null,ontologyFormat,NCI\_Thesaurus,Metadata 1 -1**

**propTally en,P106,NCI\_Thesaurus,Metadata 111 -111**

**propTally en,P107,NCI\_Thesaurus,Metadata 102 -102**

**propTally en,P108,NCI\_Thesaurus,Metadata 111 -111**

**propTally en,P90,NCI\_Thesaurus,Metadata 111 -111**

**propTally en,P97,NCI\_Thesaurus,Metadata 112 -112**

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

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

\*\* definition, design note, full syn

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

**propTypeCount 81 168 -87**

**propPropCount 438409 836824 -398415**

**propPropTally miRBase\_ID,qualifier,label,Content 152 -152**

**propPropTally Accepted\_Therapeutic\_Use\_For,qualifier,label,Content 572 -572**

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

**propPropTally CAS\_Registry,qualifier,label,Content 11931 -11931**

**propPropTally CHEBI\_ID,qualifier,label,Content 3107 -3107**

**propPropTally Chemical\_Formula,qualifier,label,Content 10813 -10813**

**propPropTally Concept\_Status,qualifier,label,Content 4277 -4277**

**propPropTally Contributing\_Source,qualifier,label,Content 27341 -27341**

**propPropTally EntrezGene\_ID,qualifier,label,Content 3009 -3009**

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

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

**propPropTally Extensible\_List,qualifier,label,Content 267 -267**

**propPropTally FDA\_Table,qualifier,label,Content 1824 -1824**

**propPropTally FDA\_UNII\_Code,qualifier,label,Content 12628 -12628**

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

**propPropTally GenBank\_Accession\_Number,qualifier,label,Content 3190 -3190**

**propPropTally Gene\_Encodes\_Product,qualifier,label,Content 1973 -1973**

**propPropTally HGNC\_ID,qualifier,label,Content 2983 -2983**

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

**propPropTally ICD-O-3\_Code,qualifier,label,Content 1220 -1220**

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

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

**propPropTally In\_Clinical\_Trial\_For,qualifier,label,Content 826 -826**

**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 566 -566**

**propPropTally NCI\_META\_CUI,qualifier,label,Content 12425 -12425**

**propPropTally NSC\_Code,qualifier,label,Content 2366 -2366**

**propPropTally Neoplastic\_Status,qualifier,label,Content 7944 -7944**

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

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

**propPropTally OLD\_ASSOCIATION,qualifier,label,Content 51 -51**

**propPropTally OLD\_CHILD,qualifier,label,Content 2663 -2663**

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

**propPropTally OLD\_PARENT,qualifier,label,Content 3669 -3669**

**propPropTally OLD\_ROLE,qualifier,label,Content 551 -551**

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

**propPropTally OMIM\_Number,qualifier,label,Content 8793 -8793**

**propPropTally PDQ\_Closed\_Trial\_Search\_ID,qualifier,label,Content 3455 -3455**

**propPropTally PDQ\_Open\_Trial\_Search\_ID,qualifier,label,Content 3455 -3455**

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

**propPropTally PubMedID\_Primary\_Reference,qualifier,label,Content 473 -473**

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

**propPropTally SNP\_ID,qualifier,label,Content 87 -87**

**propPropTally Semantic\_Type,qualifier,label,Content 100868 -100868**

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

**propPropTally Swiss\_Prot,qualifier,label,Content 2764 -2764**

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

**propPropTally UMLS\_CUI,qualifier,label,Content 77271 -77271**

**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**

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

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

**propPropTally definition,source,NCI,Metadata 168 168**

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

**propPropTally presentation,source,NCI,Metadata 184 184**

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

**relCount 443319 1043429 -600110**

**relTally NCI\_Thesaurus,Disease\_Has\_Finding,Content 19088 19091 -3**

**relTally NCI\_Thesaurus,Disease\_May\_Have\_Cytogenetic\_Abnormality,Content 9049 9050 -1**

**relTally NCI\_Thesaurus,Disease\_May\_Have\_Finding,Content 12892 12932 -40**

**relTally NCI\_Thesaurus,Disease\_May\_Have\_Molecular\_Abnormality,Content 5799 5807 -8**

**relTally NCI\_Thesaurus,Molecular\_Abnormality\_Involves\_Gene,Content 619 621 -2**

**relTally NCI\_Thesaurus,subClassOf,Content 87567 105443 -17876**

\*\*\*\* There are extra rel entries in the DB being created by the loader.

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

\*\*\*\* disjointWith is not being counted in the file. TODO: count in file

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

**relTally NCI\_Thesaurus,Allele\_Associated\_With\_Disease,Anonymous 12 -12**

**relTally NCI\_Thesaurus,Allele\_Has\_Abnormality,Anonymous 184 -184**

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

**relTally NCI\_Thesaurus,Allele\_In\_Chromosomal\_Location,Anonymous 107 -107**

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

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

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

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

**relTally NCI\_Thesaurus,Anatomic\_Structure\_Is\_Physical\_Part\_Of,Anonymous 7793 -7793**

**relTally NCI\_Thesaurus,Biological\_Process\_Has\_Associated\_Location,Anonymous 517 -517**

**relTally NCI\_Thesaurus,Biological\_Process\_Has\_Initiator\_Chemical\_Or\_Drug,Anonymous 112 -112**

**relTally NCI\_Thesaurus,Biological\_Process\_Has\_Initiator\_Process,Anonymous 125 -125**

**relTally NCI\_Thesaurus,Biological\_Process\_Has\_Result\_Anatomy,Anonymous 63 -63**

**relTally NCI\_Thesaurus,Biological\_Process\_Has\_Result\_Biological\_Process,Anonymous 86 -86**

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

**relTally NCI\_Thesaurus,Biological\_Process\_Is\_Part\_Of\_Process,Anonymous 468 -468**

re**lTally NCI\_Thesaurus,Chemical\_Or\_Drug\_Affects\_Abnormal\_Cell,Anonymous 28 -28**

**relTally NCI\_Thesaurus,Chemical\_Or\_Drug\_Affects\_Cell\_Type\_Or\_Tissue,Anonymous 267 -267**

**relTally NCI\_Thesaurus,Chemical\_Or\_Drug\_Affects\_Gene\_Product,Anonymous 339 -339**

**relTally NCI\_Thesaurus,Chemical\_Or\_Drug\_Has\_Mechanism\_Of\_Action,Anonymous 1032 -1032**

**relTally NCI\_Thesaurus,Chemical\_Or\_Drug\_Has\_Physiologic\_Effect,Anonymous 1328 -1328**

**relTally NCI\_Thesaurus,Chemical\_Or\_Drug\_Is\_Metabolized\_By\_Enzyme,Anonymous 614 -614**

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

**relTally NCI\_Thesaurus,Chemotherapy\_Regimen\_Has\_Component,Anonymous 10802 -10802**

**relTally NCI\_Thesaurus,Conceptual\_Part\_Of,Anonymous 346 -346**

**relTally NCI\_Thesaurus,Cytogenetic\_Abnormality\_Involves\_Chromosome,Anonymous 516 -516**

**relTally NCI\_Thesaurus,Disease\_Excludes\_Abnormal\_Cell,Anonymous 12270 -12270**

**relTally NCI\_Thesaurus,Disease\_Excludes\_Cytogenetic\_Abnormality,Anonymous 401 -401**

**relTally NCI\_Thesaurus,Disease\_Excludes\_Finding,Anonymous 9119 -9119**

**relTally NCI\_Thesaurus,Disease\_Excludes\_Molecular\_Abnormality,Anonymous 329 -329**

**relTally NCI\_Thesaurus,Disease\_Excludes\_Normal\_Cell\_Origin,Anonymous 3968 -3968**

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

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

**relTally NCI\_Thesaurus,Disease\_Has\_Abnormal\_Cell,Anonymous 33871 -33871**

**relTally NCI\_Thesaurus,Disease\_Has\_Associated\_Anatomic\_Site,Anonymous 24691 -24691**

**relTally NCI\_Thesaurus,Disease\_Has\_Associated\_Disease,Anonymous 810 -810**

**relTally NCI\_Thesaurus,Disease\_Has\_Cytogenetic\_Abnormality,Anonymous 234 -234**

**relTally NCI\_Thesaurus,Disease\_Has\_Finding,Anonymous 19091 -19091**

**relTally NCI\_Thesaurus,Disease\_Has\_Metastatic\_Anatomic\_Site,Anonymous 160 -160**

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

**relTally NCI\_Thesaurus,Disease\_Has\_Normal\_Cell\_Origin,Anonymous 13519 -13519**

**relTally NCI\_Thesaurus,Disease\_Has\_Normal\_Tissue\_Origin,Anonymous 13963 -13963**

**relTally NCI\_Thesaurus,Disease\_Has\_Primary\_Anatomic\_Site,Anonymous 16063 -16063**

**relTally NCI\_Thesaurus,Disease\_Is\_Grade,Anonymous 842 -842**

**relTally NCI\_Thesaurus,Disease\_Is\_Stage,Anonymous 2098 -2098**

**relTally NCI\_Thesaurus,Disease\_Mapped\_To\_Chromosome,Anonymous 206 -206**

**relTally NCI\_Thesaurus,Disease\_Mapped\_To\_Gene,Anonymous 3321 -3321**

**relTally NCI\_Thesaurus,Disease\_May\_Have\_Abnormal\_Cell,Anonymous 1055 -1055**

**relTally NCI\_Thesaurus,Disease\_May\_Have\_Associated\_Disease,Anonymous 1475 -1475**

**relTally NCI\_Thesaurus,Disease\_May\_Have\_Cytogenetic\_Abnormality,Anonymous 9050 -9050**

**relTally NCI\_Thesaurus,Disease\_May\_Have\_Finding,Anonymous 12932 -12932**

**relTally NCI\_Thesaurus,Disease\_May\_Have\_Molecular\_Abnormality,Anonymous 5807 -5807**

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

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

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

**relTally NCI\_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 2190 -2190**

**relTally NCI\_Thesaurus,Gene\_Found\_In\_Organism,Anonymous 2703 -2703**

**relTally NCI\_Thesaurus,Gene\_Has\_Abnormality,Anonymous 228 -228**

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

**relTally NCI\_Thesaurus,Gene\_In\_Chromosomal\_Location,Anonymous 2693 -2693**

**relTally NCI\_Thesaurus,Gene\_Involved\_In\_Pathogenesis\_Of\_Disease,Anonymous 780 -780**

**relTally NCI\_Thesaurus,Gene\_Is\_Biomarker\_Of,Anonymous 24 -24**

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

**relTally NCI\_Thesaurus,Gene\_Is\_Element\_In\_Pathway,Anonymous 6089 -6089**

**relTally NCI\_Thesaurus,Gene\_Mutant\_Encodes\_Gene\_Product\_Sequence\_Variation,Anonymous 172 -172**

**relTally NCI\_Thesaurus,Gene\_Plays\_Role\_In\_Process,Anonymous 20847 -20847**

**relTally NCI\_Thesaurus,Gene\_Product\_Encoded\_By\_Gene,Anonymous 10537 -10537**

**relTally NCI\_Thesaurus,Gene\_Product\_Expressed\_In\_Tissue,Anonymous 1088 -1088**

**relTally NCI\_Thesaurus,Gene\_Product\_Has\_Abnormality,Anonymous 22 -22**

**relTally NCI\_Thesaurus,Gene\_Product\_Has\_Associated\_Anatomy,Anonymous 1716 -1716**

**relTally NCI\_Thesaurus,Gene\_Product\_Has\_Biochemical\_Function,Anonymous 5590 -5590**

**relTally NCI\_Thesaurus,Gene\_Product\_Has\_Chemical\_Classification,Anonymous 723 -723**

**relTally NCI\_Thesaurus,Gene\_Product\_Has\_Malfunction\_Type,Anonymous 1 -1**

**relTally NCI\_Thesaurus,Gene\_Product\_Has\_Organism\_Source,Anonymous 912 -912**

**relTally NCI\_Thesaurus,Gene\_Product\_Has\_Structural\_Domain\_Or\_Motif,Anonymous 1199 -1199**

**relTally NCI\_Thesaurus,Gene\_Product\_Is\_Biomarker\_Of,Anonymous 30 -30**

**relTally NCI\_Thesaurus,Gene\_Product\_Is\_Biomarker\_Type,Anonymous 275 -275**

**relTally NCI\_Thesaurus,Gene\_Product\_Is\_Element\_In\_Pathway,Anonymous 6121 -6121**

**relTally NCI\_Thesaurus,Gene\_Product\_Is\_Physical\_Part\_Of,Anonymous 361 -361**

**relTally NCI\_Thesaurus,Gene\_Product\_Malfunction\_Associated\_With\_Disease,Anonymous 507 -507**

**relTally NCI\_Thesaurus,Gene\_Product\_Plays\_Role\_In\_Biological\_Process,Anonymous 10996 -10996**

**relTally NCI\_Thesaurus,Gene\_Product\_Sequence\_Variation\_Encoded\_By\_Gene\_Mutant,Anonymous 351 -351**

**relTally NCI\_Thesaurus,Molecular\_Abnormality\_Involves\_Gene,Anonymous 621 -621**

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

**relTally NCI\_Thesaurus,Procedure\_Has\_Excised\_Anatomy,Anonymous 405 -405**

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

**relTally NCI\_Thesaurus,Procedure\_Has\_Partially\_Excised\_Anatomy,Anonymous 193 -193**

**relTally NCI\_Thesaurus,Procedure\_Has\_Target\_Anatomy,Anonymous 668 -668**

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

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

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

**relTally NCI\_Thesaurus,Regimen\_Has\_Accepted\_Use\_For\_Disease,Anonymous 307 -307**

**relTally NCI\_Thesaurus,equivalentClass,Anonymous 12077 -12077**

**relTally NCI\_Thesaurus,subClassOf,Anonymous 269779 -269779**

**relTally NCI\_Thesaurus,domain,Metadata 82 -82**

**relTally NCI\_Thesaurus,range,Metadata 82 -82**

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

**relTally NCI\_Thesaurus,type,Metadata 124 -124**

\*\*\*\* 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.

**relTypeCount 112 133 -21**

\*\*\*\* 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|

**relPropCount 299769 599646 -299877**

**relPropTally owl:someValuesFrom,Content 299769 299823 -54**

\*\*\*\* 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.

**sourceCount 46 46 0**

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

\*\*\*\* !!! This has been fixed!

**ttyCount 16 17 -1**

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