**concepts; 2nd map the concepts outside MeSH to the MeSH concepts.**

**In the comment of Term Map, “->” means “maps to”.**

**The meaning of some term types (TTY) are related to the MeSH xml elements. Description of the xml elements help us understand the what the meaning of TTY is.**

**It is very important to understand the three level of xml structure in MeSH.**

**There are a lot of concepts in Metathesaurus and these concepts have various names.**

**Metathesaurus select one of these names for the preferred name which uniquely represents the concept.**

**“Term Map” make the other names map to the preferred name by mysql.**

**This document descrebes the method which is used to map the concepts to the one in Mesh.**

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2232139/?page=1>

**MetaMap program**

[**http://metamap.nlm.nih.gov/**](http://metamap.nlm.nih.gov/)

**$conn->do("INSERT INTO TERMMAP SELECT DISTINCT SUI, STR, NULL, NULL FROM MRCONSO");**

line : 156

comment : There are a lot of records that have the same **STRING(SUI)** in MRCONSO . These duplicate strings represent either the same or different concepts, they belong to either the same or different sources. This sql select the disinct **STRING(SUI)** from MRCONSO and put them into table TERMMAP.

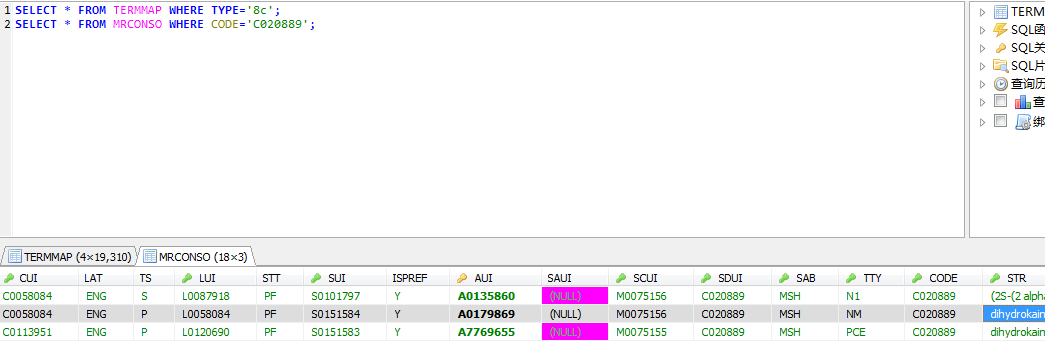
The sql “*select p1.\* from MRCONSO p1, MRCONSO p2 where p1.SUI=p2.SUI;*” help us find the records which have the same **STRING(SUI).**

**$conn->do("REPLACE INTO TERMMAP SELECT A.SUI, A.STR, B.STR MAPPING, IF(A.TTY='CE', '8b', '8c') FROM MRCONSO A, MRCONSO B WHERE A.TTY IN ('CE', 'N1') AND A.CODE=B.CODE AND B.TTY='NM'");**

line : 162

prior : we have to understand the meaning of *CODE NM CE* and *N1* with an aim to clear what the sql is dong.

**CODE:** unique Identifier or code for string in source. This identifier or code come from source.



*“dihydrokainic acid, (2S-(2 alpha,3 beta,4 beta))-2-carboxy-4-(1- methylethyl)-3-pyrrolidineacetic acid and dihydrokainate”* have the same CODE value and all of them come frome MeSH. In order to understand the relatationship between them, we have to know the structure of MeSH.

<https://www.nlm.nih.gov/mesh/>

<http://www.ncbi.nlm.nih.gov/mesh>

MeSH introduce:

If you’d like to be familiar with the structure of MEDLINE, please read this. It also helps you understand SA input files more clearly.

<https://www.nlm.nih.gov/mesh/introduction.html>

It describes the structure of MeSH xml files that are the input file of SA:

<https://www.nlm.nih.gov/mesh/concept_structure.html>

<https://www.nlm.nih.gov/mesh/xmlmesh.html>

the relation between MeSH and MEDLINE:

<http://www.nlm.nih.gov/pubs/factsheets/medline.html>

TR MeSH:

<https://images.webofknowledge.com/WOK46/help/MEDLINE/hsa_meshdetails.html>

It lists the mapping between the abbreviations of TTY values and MeSH xml elements

<http://www.nlm.nih.gov/mesh/xmlconvert.html>

comment : the sql *“ A.CODE=B.CODE”* means A and B come from the same record in MeSH. The relationship between them is synonyms, alternate forms or interchangeable.

The sql *“A.TTY IN ('CE', 'N1')”* means A.STR is either the name of an entry term or a Chemical Abstracts Service Type 1 name of a chemical .

The sql *“B.TTY='NM'”*  means B.STR is the preferred name of a supplementary record in MeSH.

This sql builds the mapping between A.STR and B.STR which is the preferred name of a supplementary record in MeSH.

*“8c”* means the string is a Chemical Abstracts Service Type 1 name and *“8b”* means the string is a usual name of an entry term.

Question: since the sql doesn’t contain the limitation *“SAB=MSH”* , does it imply that the CODE value doen’t appear in another soure?

**$conn->do("REPLACE INTO TERMMAP SELECT SUI, STR, STR MAPPING, '8a' FROM MRCONSO WHERE TTY='NM'");**

line: 171

comment: if the STR is the preferred name of a supplementary record in MeSH, set the TYPE as *“8a”*.

**$conn->do("REPLACE INTO TERMMAP SELECT A.SUI, A.STR, B.STR MAPPING, '8d' FROM MRCONSO A, MRCONSO B WHERE A.TTY='N1' AND A.CODE=B.CODE AND B.TTY='MH'");**

line: 177

comment: build the mapping between CAS Type 1 names and main headings which is also known as descriptors.

**$conn->do("REPLACE INTO TERMMAP SELECT A.SUI, A.STR, B.TOEXPR MAPPING, '6' FROM MRCONSO A, MRSMAP B WHERE A.CUI=B.FROMEXPR AND B.FROMTYPE='CUI' AND B.TOTYPE='BOOLEAN\_EXPRESSION\_STR' AND B.MAPSETSAB='MTH' AND (B.REL='RB' OR B.REL='SY')");**

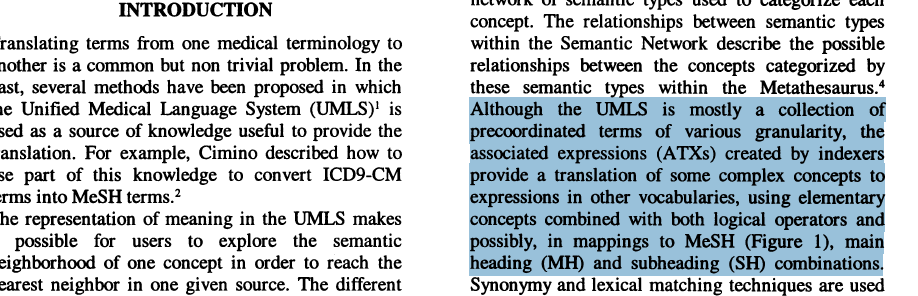
line: 192

prior: this document help you understand the representation of the mapping. Mapping is the relationship between concepts which are belong to different source vocabularies.

<http://www.ncbi.nlm.nih.gov/books/NBK9684/#ch02.I210_Mappings>

the mappings between complex concepts and associated expressions(ATXs).

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2232139/?page=1>

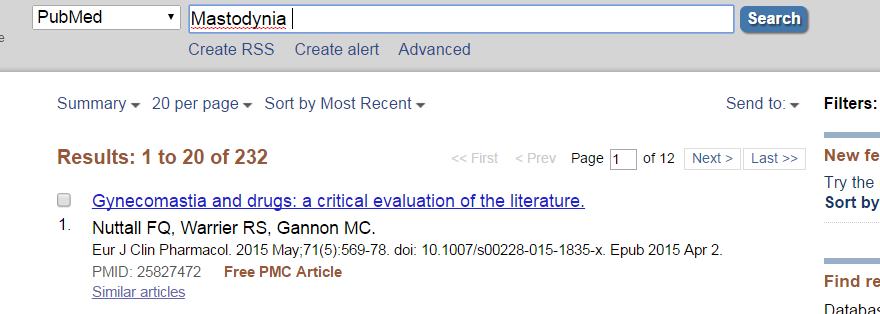


comment:

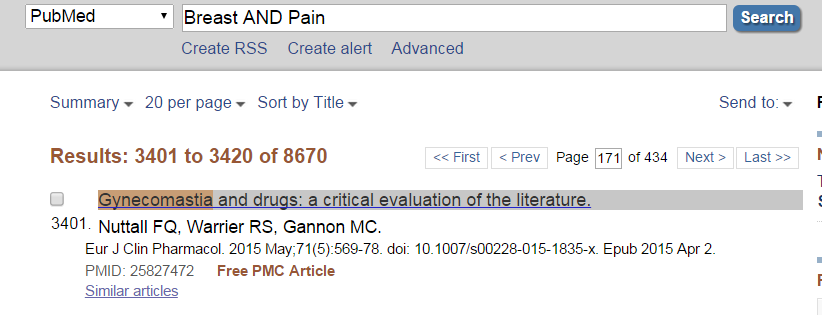


“Mastodynia”(乳腺痛), as well as “Breast” and “Pain”, is a descriptor in MeSH and represents a unique concept. The expression, “Bread and Pain” can indicate the same meaning of “Mastodynia”. There is a case which can prove the expression indicate the same meaning of the descriptor here.

When we search “Mastodynia” in [**PubMed**](http://www.ncbi.nlm.nih.gov/pubmed), we get an artical whose title is “**Gynecomastia and drugs: a critical evaluation of the literature.**”.



When we search “Breast AND Pain” in [**PubMed**](http://www.ncbi.nlm.nih.gov/pubmed), we get this artical again.



“Congenital cataract”(先天性白内障) is a descriptor in MeSH and represent a unique concept. “Cataract” is another descriptor in MeSH and “congenital” is a qualifier in MeSH. The expression “Cataract/congenital” indicate the same meaning of “Congenital cataract”.

This kind of expression is recorded as a unqiue concept in MRCONSO by MTH, but it doesn’t appear in MeSH. We need build the mapping between the preferred names between these expressions.

The sql *“AND B.MAPSETSAB='MTH'”* means this mapping was built by MTH .

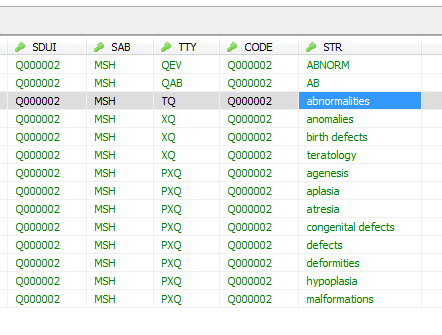
The sql *“B.REL='RB' OR B.REL='SY'”* means we only match the “SY” and “RB” mapping.

This is my own interpretation. It may be inaccurate.

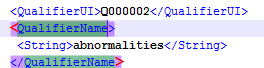
**$conn->do("CREATE TEMPORARY TABLE LIKE\_STR AS SELECT DISTINCT A.SUI SUI, A.STR STR, B.STR MAPPING, IF(B.TTY IN ('MH', 'TQ'), 1, 0) MAIN, B.CODE FROM MRCONSO A, MRCONSO B WHERE A.CUI=B.CUI AND B.SAB='MSH' AND B.TTY IN ('MH', 'TQ', 'CE', 'EN', 'EP', 'XQ')");**

line: 213

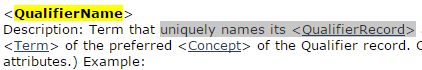
prior: “*TQ”* means “Topical qualifier”. We still can’t clear the meaning of “*TQ”*. The xml element can help us.



We find whcih element the value of is “abnormalities” in “Q000002” record in file, qual2014. The element is “<QualifierName>” .



The meaning of “<QualifierName>” is “ uniquely names its <[Q](https://www.nlm.nih.gov/mesh/xml_data_elements.html#QualifierRecord)ualiferRecord>”. So “TQ” represents the preferred name of the a qulifer record.



*“XQ”* is an alternate name for a qualifier. The relationship between *“XQ”* and *“TQ”* is *“SY”* or “BN/BO”.

*“CE” is supernumerary in the sql?*

*“EN/EP”* is related to the atrribute “PrintFlagYN” .

*“PM”* is related to the atrribute “IsPermutedTermYN” and exclude term map.

*“PEN”* means preferred non-print entry term. Should term map include this kind of term?

comment: create a table and build the mapping between a concept name and its MeSH synonym.

**$conn->do("REPLACE INTO TERMMAP SELECT L.SUI SUI, L.STR STR, B.STR MAPPING, '5' FROM LIKE\_STR L, MRCONSO B WHERE L.MAIN=0 AND L.CODE=B.CODE AND B.SAB='MSH' AND B.TTY IN ('MH', 'TQ')");**

line: 224

comment: build the mapping between a concept name which may be outside MeSH and its MeSH preferred name.

**$conn->do("REPLACE INTO TERMMAP SELECT SUI, STR, MAPPING, '4' FROM LIKE\_STR WHERE MAIN=1");**

line: 234

comment: it’s the same as line 224.

**$conn->do("REPLACE INTO TERMMAP SELECT A.SUI SUI, A.STR STR, B.STR MAPPING, '3' FROM MRCONSO A, MRCONSO B WHERE A.SAB='MSH' AND A.TTY IN ('CE', 'EN', 'EP', 'XQ') AND A.CODE=B.CODE AND B.SAB='MSH' AND B.TTY IN ('MH', 'TQ')");**

line: 243

**comment: this sql is similar to line 224. After this sql, the TYPE 4 indicates the mapping between a MeSH preferred name and a STR from another source and they are synonymy. The TYPE 5 indicates the mapping between a STR outside MeSH and a MeSH preferred name and they may not synonymy.The TYPE 3 indicates the mapping between a STR in MeSH and a preferred name in MeSH**.

**$conn->do("REPLACE INTO TERMMAP SELECT SUI, STR, STR MAPPING, '2' FROM MRCONSO WHERE SAB='MSH' AND TTY IN ('MH', 'TQ')");**

line: 254

comment: this sql set the TYPE of the preferred names of descripts and qualifiers as 2.

**generate\_termmap7();**

line 290

*“$maximum\_process”* : the thread number to generate type 7 file.

*“$each\_set\_records”* : the numbero of records that are processed by each thread.

line: 771

comment: SSH login without password!

**SubPremap.pl**

**my $sth = focus\_t($dbh, "SELECT SUI, STR FROM TERMMAP WHERE MAPPING IS NULL LIMIT $start\_limit,$end\_limit");**

line: 60

commnet: select the SUI AND STR which don’t build mapping.

**my $sth1 = focus\_t($dbh, "SELECT DISTINCT CUI FROM MRCONSO WHERE SUI='@$sui\_str[0]'");**

line: 63

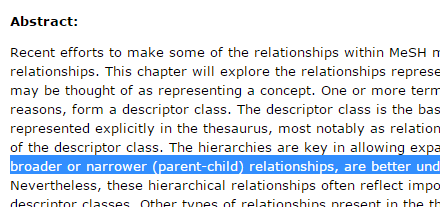
comment: find the cui in MRCONSO whose sui is identical with the one whose MAPPING is NULL in TERMMAP.

*TERMMAP-->CONSO-->CONCEPT*

This site decribes the relastionships in MeSH:

<https://www.nlm.nih.gov/mesh/meshrels.html>

*“BO/BN”* is synonymous with *“PAR/CHD”*.



I guess that “BO/BN” implies that the concepts are from the same record in MeSH, while “PAR/CHD” imples that the concepts are from different records in MeSH.

**sub children\_for\_seed**

line: 593 file: MeSHMAP.pm

prior: line 35 file “export\_mapping\_data.sql”  **create table RO as select cui1, cui2 from MRREL where rel='RO';**

***\*\*\*\*\*According to “Beyond Synonymy”, children and narrower concepts are all accepted to be the seed.***

In table “RO”, CUI1 and cui2 have relationship other than synonymous, narrower, or broader. They are not synonymous, narrower, or broader with each other.

line 45 file "export\_mapping\_data.sql"

**insert ignore into ROC select r.cui1, r.cui2 from RO r, MRSTY s1, MRSTY s2 where r.cui1 = s1.cui and r.cui2 = s2.cui and s1.tui = s2.tui;**

CUI1 and CUI2 have the same semantic type.

line: 11 file: "export\_mapping\_data.sql"

**create table T186 as select tui1, tui2 from SRSTRE1 where rel='T186';**

According to SRDEF, T186 represents “isa” semantic type.

line: 50 file : "export\_mapping\_data.sql"

**insert ignore into ROC select r.cui1, r.cui2 from RO r, MRSTY s1, MRSTY s2, T186 sn where r.cui1 = s1.cui and r.cui2 = s2.cui and s1.tui = sn.tui1 and s2.tui = sn.tui2;**

***\*\*\*\*According to “Beyond synonymy”, at least one of semantic types of the concept is identical to or has an “inverse\_isa” relationship to at least one of the sematic types of the argument concept.***

The semantic relationship between CUI1 and CUI2 is “isa” , that means CUI2 is a CUI1.

comment: the sql select the concepts whose semantic type are the same as the argument “concept”, or the ones that the relationship between the argument “concept” and is “isa”.

**sub siblings\_for\_seed**

line: 613 file: MeSHMap.pm

comment: select the concepts whose semantic types are the same as the argument “concept” . In addition, the reletionship between the concepts and the argument “concept” is “SIB” in MTH.

line: 3 file: "export\_mapping\_data.sql"

**create table PARRB as select cui1, cui2 from MRREL where rel in ('PAR','RB') AND cui1 != cui2;**

comment: the relationship from cui2 to cui1 is “PAR/RB”.

line: 19 file: "export\_mapping\_data.sql"

**insert ignore into PARRBC select r.cui1, r.cui2 from PARRB r, MRSTY s1, MRSTY s2 where r.cui1 = s1.cui and r.cui2 = s2.cui and s1.tui = s2.tui;**

comment: the relationship from cui2 to cui1 is “PAR/RB” and they have the same semantic type.

line: 24 file: "export\_mapping\_data.sql"

**insert ignore into PARRBC select r.cui1, r.cui2 from PARRB r, MRSTY s1, MRSTY s2, T186 sn where r.cui1 = s1.cui and r.cui2 = s2.cui and s1.tui != s2.tui and s1.tui = sn.tui1 and s2.tui = sn.tui2;**

comment: the relationship from cui2 to cui1 is “PAR/RB” and the semantic relationship from cui2 to cui1 is “isa”.

line: 469 file: "MeSHMap.pm"

**return if @concepts < 1 or @concepts > $MAX\_START\_CONCEPTS;**

comment: the max number of the children(CHD/RN) or siblings concepts is 1000.

line 483 file: "MeSHMap.pm"

**my $rows\_count = $conn->do\_count("CREATE TEMPORARY TABLE WORK$thread AS SELECT DISTINCT R.CUI1, R.CUI2, 0 AGE FROM $rel\_table R WHERE R.CUI1 IN ($start\_cuis)");**

comment: get the number of the argument concepts’ appearance in table *“PARRBC”*.

line: 491 file: "MeSHMap.pm"

comment: find the *“PAR/RB”* concepts of the argument concepts recursively and insert them into table *“WORK$thread”*. The smaller the *“AGE”* is, the older the *“CUI2”* is.

line: 553 file: "MeSHMap.pm"

comment:

transitive closure：

line 62~77 file Subpremap.pl

comment: select the the *sui* and *str* whose *map* is null in *TERMMAP* and find out the cui of the sui in *MRCONSO*. According the *cui*, find out the *map* in *CONCEPT*.

**map\_up\_from\_string**

comment: build the table *CONCEPT*

line: 469

return if @concepts < 1 or @concepts > $**MAX\_START\_CONCEPTS**

comment : **MAX\_START\_CONCEPTS=1000** for now.If the source concept has over 1000 chilldren or siblings , we just abandon this concept.

line: 482~503

comment: find out all of the compatible ancestors of the source concepts.

line : 537~548

comment: build the PAR/CHR relationship structure, *%is\_parent\_of*, is\_parent\_of{par}{chd} = 1. @pairs : [CHD][PAR], CHD is one of source concepts.

line: 553

comment: *@covers* contains all of compatible ancestors which have suitable coverage. The ancestors in @covers are parents of more than 70% source concepts.

line: 560~565

comment: select the *MH*, *XQ* or MeSH expressions of the *cuis* among the *@covers* from table *MRONSO* and *MRSMAP*.

line: 566~578

comment: exclude the candidates that are ancestors from each other.

**premap.pl**

line: 324

**$conn->do("LOAD DATA INFILE '" . File::Spec->catfile(File::Spec->rel2abs('.'), "termmap.type7") "' REPLACE INTO TABLE TERMMAP" . " FIELDS TERMINATED BY '|'");**

comment: import “termmap.type7” into TABLE *TERMMAP*.

line: 369~375

comment: create table *“TERMMAP\_MRCONSO”*

line: 377

**$conn->do("INSERT INTO TERMMAP\_MRCONSO SELECT T1.SUI, T1.STR, C1.CUI FROM MRCONSO C1, TERMMAP T1 WHERE T1.SUI=C1.SUI AND T1.MAPPING IS NULL");**

comment: table “TERMMAP\_MRCONSO” consitsts of *SUI*, *STR* which are from *TERMMAP*, and *CUI* which are from *MRCONSO*. **A *SUI* maybe have different *CUIs***.

line: 385

comment: create table “*MRCONSO\_TERMMAP*”

line: 391

**$conn->do("INSERT INTO MRCONSO\_TERMMAP SELECT C2.CUI, T2.MAPPING, T2.TYPE FROM MRCONSO C2, TERMMAP T2 WHERE C2.SUI=T2.SUI AND T2.MAPPING IS NOT NULL AND T2.TYPE!='7d'");**

commnet: the table “*MRCONSO\_TERMMAP*” records the mappings between the *CUI* from *MRCONSO* and *MAPPING* from *TERMMAP*.

*The “T2.TYPE!=7d” may be unnecessary.*

line 399

**$conn->do("INSERT INTO TERMMAP SELECT DISTINCT T1.SUI, T1.STR, CASE LEFT(T2.TYPE, 1) WHEN '7' THEN T2.MAPPING WHEN '8' THEN CONCAT('<<', T2.MAPPING, '>>') ELSE CONCAT('<', T2.MAPPING, '>') END, '7d' FROM TERMMAP\_MRCONSO T1, ROC R, MRCONSO\_TERMMAP T2 WHERE T1.CUI=R.CUI1 AND R.CUI2=T2.CUI GROUP BY T2.CUI");**

comment: if the relationship between the *CUI* of *SUI* from *TERMMAP* and the *CUI* of *MAPPING* from *TERMMAP* is *ROC*, make the TYPE be *‘7d’*. “*<<T2.MAPPING>>*” means the mapping is from *META SUPP*. *“T2.MAPPING*” means the mapping is from term\_type7. “*<T2.MAPPING>*” means the mapping is from META without SUPP. The records are grouped by the *CUI* of the *MAPPING* (instead of “distinct”?). *Maybe the ‘DISTINCT’ is unnessary because the table ‘MRCONSO\_TERMMAP’ and ‘TERMMAP\_MRCONSO’ has no primary key.*

line 411~412

comment: drop the tables MRCONSO\_TERMMAP and TERMMAP\_MRCONSO.

line 415~412

comment: backup the ‘7d’ mappings.

line 448~482

comment: refer to the comment of this code. Implement the “***group\_concat***” function. Because we have droped the prmiray key in *TERMMAP*, there are some duplicated *SUIs* which are have different *CUIs*. These SUIs have different *MAPPINGS*.

line 484

**$conn->do("DELETE FROM TERMMAP WHERE TYPE='7d'");**

comment: delete the type *‘7d’*, and we will insert the type *‘7e’* later.

line : 487~490

comment: import the file “*join\_ro.txt*”

line: 549

**$conn->do("CREATE TEMPORARY TABLE MRCONSO\_FILTER AS SELECT R.SUI SUI, R.STR STR, R.CUI CUI, R.TTY TTY, R.LUI LUI FROM MRCONSO AS R, TERMMAP AS HAS\_MAP WHERE HAS\_MAP.TYPE != '1' AND R.STR=HAS\_MAP.STR AND R.TS='P' AND R.STT='PF'");**

comment: *TS=’P’* means the *LUI* is preferred. *STT=’PF’* means that the *STR* is the preferred form of the term. *Is HAS\_MAP.TYPE != '1' unneccesary here?*

line: 558

**$conn->do("INSERT INTO TERMMAP SELECT DISTINCT L.SUI SUI, L.STR STR, R.STR MAPPING, '0' FROM MRCONSO AS L, AMBIG AS A, MRCONSO\_FILTER AS R where R.CUI=A.CUI AND L.TTY != 'MM' AND L.LUI=A.LUI");**

comment: *L.TTY!=’MM’, what is ‘MM’?* It seems that there is nothing with it in MRCONSO.

**$conn->do("INSERT INTO TERMMAP\_TZAB SELECT \* FROM TERMMAP WHERE TYPE='0'");**

comment: “*TERMMAP\_TZAB*” consists of ambiguous mappings.

line: 605

**$conn->do("CREATE TEMPORARY TABLE TERMMAP\_AB AS SELECT DISTINCT A.SUI A\_SUI, A.STR A\_STR, A.MAPPING A\_MAPPING, B.MAPPING B\_MAPPING FROM TERMMAP\_TZAB AS A, TERMMAP\_TZAB AS B WHERE B.STR=A.STR AND A.MAPPING!=B.MAPPING");**

comment: create a table which consists of *SUIs*, *STRs*, the *‘ambiguous’ MAPPINGs*.

line: 614

**$conn->do("INSERT INTO TERMMAP SELECT DISTINCT AB.A\_SUI, AB.A\_STR, AB.A\_MAPPING, '1' FROM TERMMAP\_AB AS AB, TERMMAP AS C, TERMMAP AS D**

**WHERE D.STR=AB.B\_MAPPING AND D.MAPPING!=C.MAPPING AND C.MAPPING!=concat('<',D.MAPPING,'>') AND D.MAPPING!=concat('<',C.MAPPING,'>') AND C.STR=AB.A\_MAPPING AND C.TYPE!='0' AND D.TYPE!='0' AND C.TYPE IS NOT NULL AND D.TYPE IS NOT NULL");**

comment: insert the records whose *TYPE* is ‘1’.

“C.MAPPING!=concat('<',D.MAPPING,'>')” and “D.MAPPING!=concat('<',C.MAPPING,'>')” mean