How To Use DO.db

Jiang Li

June 18, 2010

Contents

1 Overview

This vignette demonstrates how to easily use the DO.db package. The Disease Ontology (DO) is a manually inspected subset of Unified Medical Language System (UMLS) and includes concepts from outside the UMLS disease/disorder semantic network including various cancers, congenital abnormalities, deformities and metal disorders. Terms in DO are organized in Directed Acyclic Graph (DAG). DO.db is used to represent the content of Disease Ontology (for example, the child-parent relationship in DO) through Bimap object.

To start with DO.db package, type following code below:

```
> library(DO.db)
> help(DO.db)
```

2 Fetch whole DO terms

In DO.db, *DOTermsAnnDbBimap* object(a sub-class of Bimap) *DOTERM* represet the whole DO terms. Meanwhile a class named *DOTerms* represents Disease Ontology nodes

with four slots. Four S4 methods *DOID,Term,Synonym* and *Secondary* are defined in DO.db to fetch the corresponding attribute for a DO term node, they can work on *DOTermsAnnDbBimap,DOTerms* and *character* classes.

Here is an example of how to use these objects and functions, it fetchs the first 10 records and turn them to list object, then use the four S4 methods to fetch the corresponding attribute. Here is the code and result:

```
> FirstTenDOBimap <- DOTERM[1:10]
> class(FirstTenDOBimap)

[1] "DOTermsAnnDbBimap"
attr(,"package")
[1] "DO.db"

> xx <- as.list(FirstTenDOBimap)
> DOID(xx[[2]])

[1] "DOID:0014667"

> Term(xx[[2]])

[1] "disease of metabolism"

> Synonym(xx[[2]])

[1] "\"disorder of metabolism NOS\" EXACT [SNOMEDCT_2005_07_31:190961002]"
[2] "\"disorder of metabolism NOS (disorder)\" EXACT [SNOMEDCT_2005_07_31:267456000]"
[3] "\"Generalized metabolic disorder (disorder)\" EXACT [SNOMEDCT_2005_07_31:30390004]
[4] "\"Metabolic disease \" EXACT [SNOMEDCT_2005_07_31:75934005]"
[5] "\"Metabolic disorder\" EXACT [NCI2004_11_17:C3235]"
[6] "\"Metabolic disorder\" EXACT [SNOMEDCT_2005_07_31:154733004]"
[7] "\"metabolism disorder\" EXACT [CSP2005:1846-2030]"
[8] "\"Unspecified disorder of metabolism\" EXACT [ICD9CM_2006:277.9]"

> Secondary(xx[[2]])

character(0)
```

3 Fetch the relationship between DO terms

In this section, we will introduce four Bimap objects which represent relationship between DO terms. There are *DOANCESTOR,DOPARENTS,DOOFFSPRING* and *DOCHILDREN*. we will introduce them in the following sub-sections.

3.1 DOANCESTOR

This data set describes associations between DO terms and their ancestor terms, based on the directed acyclic graph (DAG) defined by the Disease Ontology Consortium. The format is an R object mapping the DO terms to all ancestor terms, where an ancestor term is a more general DO term that precedes the given DO term in the DAG (in other words, the parents, and all their parents, etc.).

For example, to fetch all the *DOANCESTOR* in a list object and display the first 4 DOID's ancestors, here is the code:

```
> xx <- as.list(DOANCESTOR)
> xx <- xx[!is.na(xx)]
> xx[1:4]

$`DOID:0000000`
[1] "DOID:4"  "DOID:13" "DOID:77" "DOID:7"

$`DOID:0014667`
[1] "DOID:4"   "DOID:344"

$`DOID:0050001`
[1] "DOID:104"     "DOID:4"       "DOID:1939"    "DOID:2313"    "DOID:8478"
[6] "DOID:0050117"

$`DOID:0050002`
[1] "DOID:104"     "DOID:4"       "DOID:1939"    "DOID:2313"    "DOID:8478"
[6] "DOID:0050117"
```

3.2 DOPARENTS

This data set describes associations between DO terms and their direct parent terms, based on the directed acyclic graph (DAG) defined by the Disease Ontology Consortium. The format is an R object mapping the DO terms to all direct parent terms, where a direct parent term is a more general DO term that immediately precedes the given DO term in the DAG.

For example, to fetch all the *DOPARENTS* in a list object and display the first 4 DOID's parents, here is the code:

```
> xx <- as.list(DOPARENTS)
> xx <- xx[!is.na(xx)]
> xx[1:4]

$`DOID:0000000`
[1] "DOID:77"
```

```
$`DOID:0014667`
[1] "DOID:344"

$`DOID:0050001`
[1] "DOID:8478"

$`DOID:0050002`
[1] "DOID:8478"
```

3.3 DOOFFSPRING

This data set describes associations between DO terms and their offspring terms, based on the directed acyclic graph (DAG) defined by the Disease Ontology Consortium. The format is an R object mapping the DO terms to all offspring terms, where an ancestor term is a more specific DO term that is preceded by the given DO term in the DAG (in other words, the children and all their children, etc.).

For example, to fetch all the *DOOFFSPRING* in a list object and display the first DOID's offsprings, here is the code:

```
> xx <- as.list(DOOFFSPRING)
> xx <- xx[!is.na(xx)]
> xx[1]

$`DOID:0000000`
 [1] "DOID:1568"  "DOID:10211" "DOID:9717"  "DOID:4058"  "DOID:8145"
 [6] "DOID:5275"  "DOID:9765"  "DOID:11755" "DOID:10210" "DOID:10254"
[11] "DOID:11377" "DOID:9766"  "DOID:4140"  "DOID:8157"  "DOID:1949"
[16] "DOID:11756" "DOID:6480"  "DOID:11753" "DOID:9764"  "DOID:11151"
[21] "DOID:3120"  "DOID:8090"  "DOID:2828"  "DOID:8135"  "DOID:9714"
[26] "DOID:12150" "DOID:4057"  "DOID:8167"  "DOID:4513"  "DOID:3121"
```

3.4 DOCHILDREN

This data set describes associations between DO terms and their direct children terms, based on the directed acyclic graph (DAG) defined by the Disease Ontology Consortium. The format is an R object mapping the DO terms to all direct children terms, where a direct child term is a more specific DO term that is immediately preceded by the given DO term in the DAG.

For example, to fetch all the *DOCHILDREN* in a list object and display the first DOID's children, here is the code:

```
> xx <- as.list(DOCHILDREN)
> xx <- xx[!is.na(xx)]
> xx[1]

$`DOID:0000000`
[1] "DOID:10211" "DOID:10254" "DOID:1949"  "DOID:3121"  "DOID:4140"
[6] "DOID:9714"  "DOID:9717"
```

4 Others

In this section, we will introduce the schema of DO.db and number of mapped keys for the maps in DO.db.

To get the schema of DO.db, type following:

```
> DO_dbschema()

  --
  -- DO_DB schema
  -- ====================
  --
  CREATE TABLE do_term (
  _id INTEGER PRIMARY KEY,
  do_id VARCHAR(12) NOT NULL UNIQUE, -- DI ID
  term VARCHAR(255) NOT NULL -- textual label for the DO term
  );

  CREATE TABLE do_synonym (
  _id INTEGER NOT NULL, -- REFERENCES do_term
  synonym VARCHAR(255) NOT NULL, -- label or DO ID
  secondary VARCHAR(12) NULL, -- DO ID
  like_do_id SMALLINT, -- boolean (1 or 0)
  FOREIGN KEY (_id) REFERENCES do_term (_id)
  );


  CREATE TABLE do_parents (
  _id INTEGER NOT NULL, -- REFERENCES do_term
  _parent_id INTEGER NOT NULL, -- REFERENCES do_term
  relationship_type VARCHAR(7) NOT NULL, -- type of DO child-parent
    relationship
  FOREIGN KEY (_id) REFERENCES do_term (_id),
  FOREIGN KEY (_parent_id) REFERENCES do_term (_id)
```

```
);

CREATE TABLE do_offspring (
_id INTEGER NOT NULL, -- REFERENCES do_term
_offspring_id INTEGER NOT NULL, -- REFERENCES do_term
FOREIGN KEY (_id) REFERENCES do_term (_id),
FOREIGN KEY (_offspring_id) REFERENCES do_term (_id)
);


CREATE TABLE do_obsolete (
do_id VARCHAR(12) PRIMARY KEY, -- DO ID
term VARCHAR(255) NOT NULL -- textual label for the DO term
)
;

CREATE TABLE map_counts (
map_name VARCHAR(80) PRIMARY KEY,
count INTEGER NOT NULL
);

CREATE TABLE map_metadata (
map_name VARCHAR(80) NOT NULL,
source_name VARCHAR(80) NOT NULL,
source_url VARCHAR(255) NOT NULL,
source_date VARCHAR(20) NOT NULL
);

CREATE TABLE metadata (
name VARCHAR(80) PRIMARY KEY,
value VARCHAR(255)
);

-- Indexes
```

And to get the number of mapped keys for the maps in DO.db, type following:

```
> DOMAPCOUNTS

 DOANCESTOR  DOCHILDREN  DOOBSOLETE DOOFFSPRING   DOPARENTS      DOTERM
      10262        3682        3778        3682       10262       10263
```