Patent application bibliographic data documentation

Bhaven N. Sampat
Assistant Professor
Mailman School of Public Health
Columbia University
bns3@columbia.edu

August 3, 2011

1 Summary

This file includes information on U.S. patent applications published between March 15, 2001 and December 31, 2010. The USPTO began publishing applications in 2001, under the American Inventors Protection Act.[1] The data were extracted from the USPTO's *Cassis* BIB Applications DVD.

The dataset includes the application number, publication number, application date, publication date, title, and patent classes at publication.

[1] http://www.uspto.gov/news/pr/2001/01-13.jsp

2 Codebook

```
. codebook
-------------------------------------------------------------------------------
class                                                   Class at publication
-------------------------------------------------------------------------------

                 type:  string (str244)
         unique values:  1135724                  missing "":  0/2625950
              examples:  "277/590"
                         "368/37"
                         "433/173"
                         "600/484 600/529"
               warning:  variable has embedded blanks
-------------------------------------------------------------------------------
appnum                                          Series code plus serial number
-------------------------------------------------------------------------------

                 type:  numeric (long)
                range:  [7153880,12923171]           units:  1
         unique values:  2623156                  missing .:  0/2625950
                 mean:   1.1e+07
              std. dev:    878271
           percentiles:        10%       25%       50%       75%       90%
                           1.0e+07   1.1e+07   1.1e+07   1.2e+07   1.2e+07
-------------------------------------------------------------------------------
pubnum                                                   Publication number
-------------------------------------------------------------------------------

                 type:  numeric (double)
                range:  [2.001e+10,2.010e+10]        units:  1
         unique values:  2625950                  missing .:  0/2625950
                 mean:   2.0e+10
              std. dev:   2.6e+07
           percentiles:        10%       25%       50%       75%       90%
                           2.0e+10   2.0e+10   2.0e+10   2.0e+10   2.0e+10
-------------------------------------------------------------------------------
appdate                                                          Filing date
-------------------------------------------------------------------------------

                 type:  numeric daily date (float)
                range:  [-18346,18525]               units:  1
       or equivalently:  [09oct1909,20sep2010]        units:  days
         unique values:  4939                     missing .:  0/2625950
                 mean:   16666.1 = 18aug2005 (+ 1 hour)
              std. dev:    948.32
           percentiles:        10%       25%       50%       75%       90%
                             15349     15914     16695     17444     17923
                         09jan2002 28jul2003 16sep2005 05oct2007 26jan2009
-------------------------------------------------------------------------------
pubdate                                                    Publication date
-------------------------------------------------------------------------------

                 type:  numeric daily date (float)
                range:  [15049,18626]                units:  1
       or equivalently:  [15mar2001,30dec2010]        units:  days
```

```
        unique values:  512                     missing .:  0/2625950
                 mean:   17078.5 = 04oct2006 (+ 11 hours)
              std. dev:   950.759
          percentiles:        10%       25%       50%       75%       90%
                            15714     16309     17128     17898     18339
                          09jan2003 26aug2004 23nov2006 01jan2009 18mar2010
```

```
title_trim                                          Title (Trimmed to 80 digits)
```

```
                type:  string (str80)
        unique values:  2157101                 missing "":  0/2625950
            examples:  "Derivatives of succinic and glutaric acids and
                       analogs thereof useful as inhibit"
                       "Iontophoresis method and apparatus for systemic
                       delivery of active agents"
                       "Methods of bonding metal surfaces, and components
                       therefor"
                       "SMART ARTEFACT AND USER TERMINAL HAVING A SHORT
                       RANGE INTERFACE AND LONG RANGE I"
             warning:  variable has embedded and trailing blanks
```

3 Data construction notes

Application numbers can be linked to inventor information in the `applications20012010_inventors` file, and to issued patents in the `bib19752010` file. Application numbers and publication numbers can also be used to merge to PAIR application data (coming soon).