Examiner citations data documentation

Bhaven N. Sampat
Associate Professor
Columbia University
bns3@columbia.edu

August 28, 2012

1 Summary

The file EXCITE includes information the 25,880,845 on "examiner" and other patent citations in U.S. utility and reissue patents issued between 2001 and 2010.

The data were extracted from the "Patent Office Bibliographic Data" XML files hosted by Google: [1]. A small number of patent numbers that were ultimately "withdrawn" were dropped from the set.[2]

The final dataset includes cited and citing U.S. patent numbers, the issue year of the citing patent, and an indication of whether the cited patent is listed as "cited by examiner'' or "cited by other." This file does not include citations to foreign patents or non-patent literature. These will be posted separately.

According to the Manual of Patent Examining Procedure[3]:

> Indication of whether a reference was listed by the examiner will be helpful in compiling statistical data related to prior art submissions so that the USPTO can better consider whether changes are required to the rules governing prior art statements.

[1] http://www.google.com/googlebooks/uspto-patents-grants-biblio.html
[2] http://www.uspto.gov/patents/process/search/withdrawn.jsp
[3] http://www.uspto.gov/web/offices/pac/mpep/documents/1300_1302_12.htm

> Indication of a reference with an asterisk should not be considered to reflect any significance other than that the reference was listed on a "Notice of References Cited," form PTO-892. When an examiner lists references on a form PTO-892, the examiner lists references that are relied upon in a prior art rejection or mentioned as pertinent. See MPEP § 707.05(c).
>
> The examiner does not list references which were previously cited by the applicant (and initialed by an examiner) on an Information Disclosure Statement ... See MPEP § 609 and § 707.05(b), (c) and (d).
>
> No distinction will be made in the "References Cited" section for other sources of references. Thus, references cited in a protest, by an attorney or agent not acting in a representative capacity but on behalf of a single inventor, and by the applicant will not be distinguished.

Citations other than those "cited by examiner" can thus include those from various sources, including applicant disclosures (based on prior art that applicants knew about when developing the invention, or learned of through searches conducted independently or by their attorneys), information from foreign search reports reported by the applicant to the USPTO, and others.

2 Codebook and descriptives

```
.
. codebook
-------------------------------------------------------------------------------
patent                                                          Citing Patent
-------------------------------------------------------------------------------

                 type:  string (str7)
         unique values:  1643596                  missing "":  0/25880845
              examples:  "6592240"
                         "6966374"
                         "7287765"
                         "7594001"
-------------------------------------------------------------------------------
cited                                                            Cited Patent
-------------------------------------------------------------------------------

                 type:  numeric (long)
                range:  [1,9989019]                  units:  1
         unique values:  3497511                  missing .:  0/25880845
```

```
                mean:   5.4e+06
             std. dev:   1.1e+06
         percentiles:        10%       25%       50%       75%       90%
                          4.0e+06   4.9e+06   5.6e+06   6.2e+06   6.6e+06
```

---

excite — Examiner Citation

```
                type:  numeric (byte)
               range:  [0,1]                        units:  1
      unique values:  2                         missing .:  0/25880845
         tabulation:  Freq.  Value
                     1.7e+07 0
                     8.7e+06 1
```

---

issyear — Citing Patent Issue Year

```
                type:  numeric (int)
               range:  [2001,2010]                  units:  1
      unique values:  10                        missing .:  0/25880845
                mean:   2006.09
             std. dev:   2.92139
         percentiles:        10%       25%       50%       75%       90%
                            2002      2004      2006      2009      2010
```

```
.
. tab issyear excite, row
```

Key
frequency *row percentage*

Citing Patent Issue Year	Examiner Citation 0	1	Total
2001	1,130,464 57.31	842,174 42.69	1,972,638 100.00
2002	1,197,997 58.41	852,925 41.59	2,050,922 100.00
2003	1,332,638 59.72	898,982 40.28	2,231,620 100.00
2004	1,368,863 61.00	875,241 39.00	2,244,104 100.00
2005	1,348,738 63.68	769,419 36.32	2,118,157 100.00
2006	1,861,847 66.25	948,650 33.75	2,810,497 100.00
2007	1,776,423 68.48	817,770 31.52	2,594,193 100.00

2008	1,836,024	794,864	2,630,888
	69.79	30.21	100.00
2009	2,166,031	829,544	2,995,575
	72.31	27.69	100.00
2010	3,160,936	1,071,315	4,232,251
	74.69	25.31	100.00
Total	17,179,961	8,700,884	25,880,845
	66.38	33.62	100.00

.

Note there has been an interesting decrease in the examiner share of all patent-to-patent citations over time.

3 Data construction notes

To verify completeness, we compared the data to citation information obtained from the Electronic Information Products Division of the USPTO, i.e. that compiled in the USCITES file. Overall, 99.96 percent of the patent-citation pairs from the USCITES file (for issue years 2001 to 2010) are also in EXCITE. The match rate is at least 99.94 percent every year.