USPTO Brief Data Evaluation

May 19, 2019

For this project I used the uspto.gov api to look at a handful of different topics. Out of a total 2,148,653 patents reviews, reviewed by me, there are 88% of all patents that are being published by men. It is grossly disproportional. The majority of the information is published after 2001; however, there is a separate endpoint that accesses data further back. The api contains 7 categories, each of which break down further into over 150 sub-categories. For example, one of the categories is patents. A few of the examples of the patents data include kind, date, category, lawyer, inventor, assignee, citations, non-patent citations, government organization id, government organization level, and much more.

One category of patents is botany. In review of 14,460 different patents ranging from 2001 to December of 2018, we can see that some plants are used more in patents.

Rosa hybrida	773
Chrysanthemum×morifolium	463
Calibrachoa sp.	265
Prunus persica	251
Petunia×hybrida	246
Rosa hybrid	209
Impatiens hawkeri	209
Pelargonium×hortorum	203
Osteospermum ecklonis	184
Hydrangea macrophylla	179
Chrysanthemum morifolium	157
Verbena hybrida	154
Alstroemeria hybrida	130
Phalaenopsis hybrid	122
Hibiscus rosa-sinensis	122
Fragaria×ananassa	112
Pelargonium peltatum	110
Vitis vinifera	108
Petunia hybrida	102
Kalanchoe blossfeldiana	101
Heuchera hybrid	97
Impatiens walleriana	92
Euphorbia pulcherrima	91
Phlox paniculata	85
Dahlia hybrida	83
Dianthus caryophyllus	81
Verbena×hybrida	76
Malus domestica	71
Pelargonium zonale	71
Phalaenopsis hybrida	71

We can look at things such as the number of pages within a patent document; all documents have at least 2 pages. 80% of documents are 10 pages or less, and the longest document is 4,011 pages long. You can look at the number of figures within a given document; a figure can be any thing from a page number to mathematical formula symbol. It may be possible to do an analysis on the amount made, or the usefulness of a patent, by the number of figures with a given document; probably won't find much, but it could happen.

We can see that by looking at the number of patents produced by the government, no surprise, the department of defense comes in at number one. The caveat here is that the government information here is limited. Out of 250 patents, 50 of those are from the Department of Defense. I'm not sure why there is so little information, it may be due to the information being dealt more as a trade secret. Darpa isn't too secretive though. Some other big organizations are National Institutes of Health, Army, Air Force, etc. You can also see the type of interest that the government has in a given patent:

'STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT The invention described herein was made in the performance of work under a NASA contract and by employees of the United States Government and is subject to the provisions of Public Law 96-517 (35 U.S.C. § 202) and may be manufactured and used by or for the Government for governmental purposes without the payment of any royalties thereon or therefore.'

There's much more to evaluate. And there are many company's that provide services for analysis of patents. Next steps would be to use nlp to better review what's in the actual text of the documents.