**Forecasting Promising Technology and Business Strategy through Corporate Patent Network**

**Jaekeun Lee**

***Keywords***

*Patent, Directed Network, Link Prediction, Text Mining, Corporate Strategy*

**Abstract**

This paper aims to generate directed patent asset network of individual companies from USPTO bulk data. The patent network can be seen as a historical evidence of evolution of company’s technological assets. Reference to theory of link prediction and comparison to other patent networks can enable descriptive analysis of company’s intellectual assets. Furthermore, company’s next technology of interest can be forecasted. The information can be utilized by various stakeholders: Government can prepare regulations for pertinent technology, media can set agendas that needs to be discussed and general public can use this for investment.

**Introduction**

In 2013, United States Patent & Trademark Office estimated that value of IP (intellectual property) to the U.S economy was more than $5 trillion and contributes to employment for about 18 million American people. Also, more developed nations have the tendency of presenting higher value in IP.[[1]](#footnote-0) Among the category of IPs, patent provides exclusive rights for successful applicants, endows strong market position and fosters development of innovative products. The importance of patent is in strategic and financial assets for business firms. The ongoing international patent war between Apple Inc and Samsung Electronics pertinent to their design rights is an exemplary case that highlights significance of managing patents.

Patent data is a self-explanatory information of technological breakthroughs or development. Also, because as it is well organized and structured set, facilitation of preprocessing led to various attempts of visualization, text-based, citation-based analysis of patents. However, analysis by “corporate-unit” seems to lack in its numbers. Patent portfolio analysis of companies can be utilized to diagnose strength of company and relative innovative power within a specific market segment. (Bernd et al. 2006)[[2]](#footnote-1) The results of diagnosis can be integrated with corporate data, providing insights about suitable cooperation or acquisition strategy. This may contribute to vitalization of business activity and encouragement of technological advancement.

In general, when corporate shifts focus outside of their scope of business area, acquisitions of startups with relevant technologies are adopted as prevalent strategy. This is for rapid transition by absorption of knowledge and technology in the new field. In this sense, analyzing patents may enable a peek into the next moves of business and corporates as well as suggest which corporate or patent owner they can negotiate with.

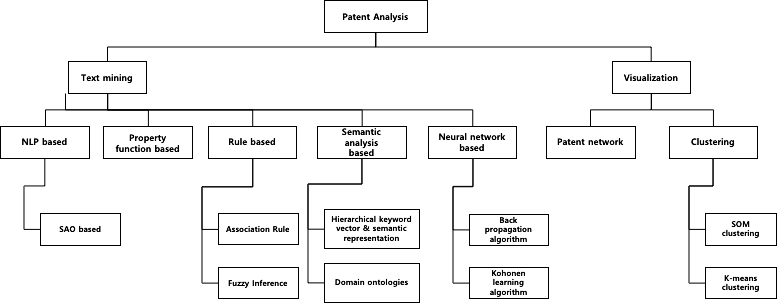
The acknowledgement of strategic movement of corporates indirectly disclose specific technologies receiving spotlights and lead to prediction of emerging technologies. Proper forecasts of promising technologies is elemental for companies where structural holes can be identified in patent network and recommendation of filling the gap with relevant patent may be suggested. Especially, establishment of efficient R&D strategy is possible as forecasting emerging technology is being recognized as a necessary phase of it. (Albright, 2002)[[3]](#footnote-2) Furthermore, Next, spillover effects can occur with the help of government officials or educators, who may create job opportunities or nurture experts of relevant technologies.

**Literature Review**

**Patent analysis techniques**

Patent is an open access data that contains information on both the developed technology and its usage rights. (Yoon et al. 2005)[[4]](#footnote-3) In recognition of disclosing technology, patents grant exclusive authority and rights to the inventors or assignees. (Trappey et al. 2012)[[5]](#footnote-4) With growing interests and movements on the importance of intellectual property rights and threat that reverse engineering can give, the contemporary society acknowledges the significance of patenting technology (Park et al. 2015)[[6]](#footnote-5). The role of patent is not confined to defensive purpose (Pénin et al. 2012) [[7]](#footnote-6); by practicing legal rights on patent infringement it can be used as a business strategy to generate profit.

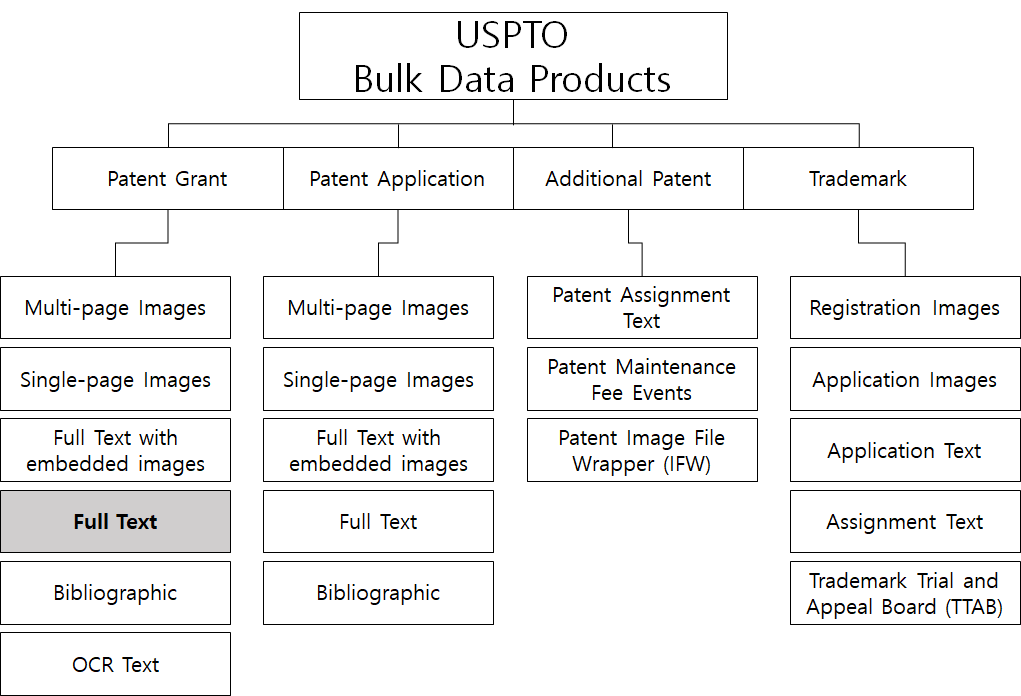
*Figure 1* displays the taxonomy of techniques for patent analysis are presented below (Assad et al. 2014)[[8]](#footnote-7). Preceding patent analysis techniques can be categorized into two main approaches: text mining approach and visualization approach. The text mining approach aims to derive significant pattern or meaningful information from textual data. Text mining techniques used in patent analysis are largely dependent on Natural Language Processing (NLP), property-function based analysis, rule-based approaches, semantic analysis approaches, and neural network approaches. Subject-action-object (SAO) structures, a type of NLP technique, were observed to be in a prominent use in text mining approach literatures. The literatures that adopted SAO structures include: identifying technological competition trends by analyzing patent vacuums and technological hot spots (Yoon et al. 2012)[[9]](#footnote-8), identification of promising patents for technology transfer (Park et al. 2013)[[10]](#footnote-9), identifying patent infringements using semantic technological similarity (Park et al. 2012)[[11]](#footnote-10).

**<Figure 1>**

Visualization approach implements text mining techniques to some degree for processing data, but its intention lies on representing and displaying patent information to facilitate understanding of technological trends or detect infringements. The visualization method includes clustering method and patent network method. KNN clustering method was used to determine trend shift for ubiquitous technologies (Kim et al., 2007)[[12]](#footnote-11) and patent network analysis was conducted with bibliometric analysis to monitor technological trends in an emerging field (Chang et al., 2009)[[13]](#footnote-12).

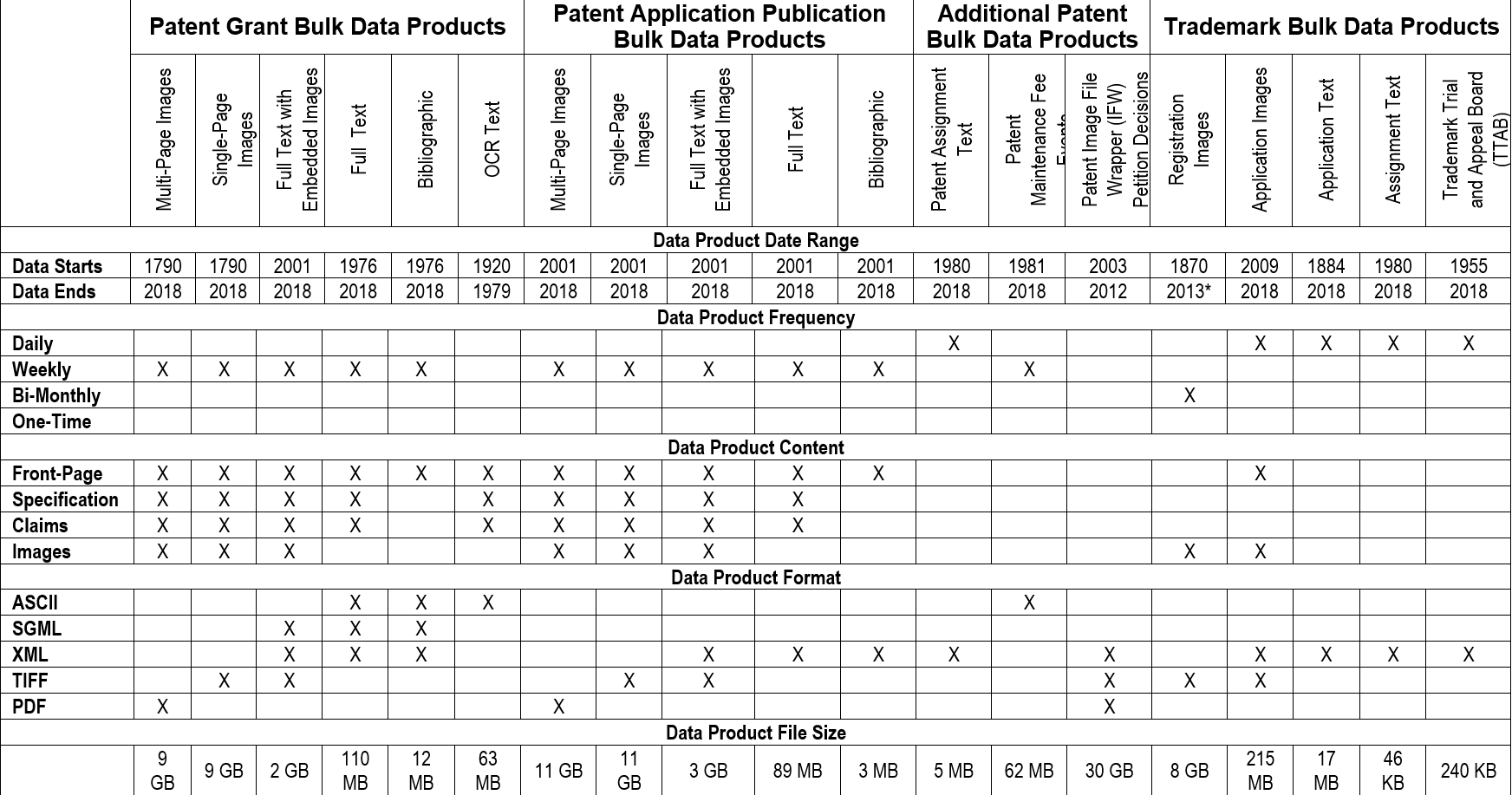
**Data and Variables**

The United States Patent and Trademark Office, an agency that issues patents to inventors and business for their inventions, provides bulk data of patents through USPTO Bulk Data Storage System (BDSS). USPTO BDSS provides wide range of patent data. The types of patent data products vary in accordance to their types, publication and purpose. There are four major patent product types: Patent Grant Bulk Data Products, Patent Application Publication Bulk Data Products, Additional Patent Bulk Data Products, Trademark Bulk Data Products (Figure 1). The products are also classified in accordance to its contents; whether the data file contains images, detailed information such as claims or specifications. The frequency of publication of product also differs.



**<Figure 1 - USPTO Bulk Data Product Types and Variations>**

Taking data size and contents into account, full-text grant patent data was chosen as main source of analysis (Figure 2). Full text data contains most information excluding images and as its format is in xml, parsing and preprocessing can be facilitated. Although the full text data dates back to 1976, grant full text data of most recent 10 year (2008-2018) has been downloaded and parsed to database file (56.54 GB). It was assumed that recent data was sufficient enough to perform contemporary analysis of corporate technology.

**<Figure 2 - USPTO Bulk Data Product Types and Specifications>**

Parsing was conducted using funginstitute’s patentprocessor library[[14]](#footnote-13), which provides xml parser for USPTO bulk data products and saves them into sqlite db file format. The xml data are parsed and organized into numerous tables: patent, claim, raw inventor, raw assignee, foreign citation, other references, ipcr, patent\_type, uspc, raw location, ipcr, etc… The data tables relevant to the research problem includes patent, claim, raw location, raw inventor, raw assignee, other reference. Detailed description and names of respective column variables of aforementioned data tables are listed below (Table 1)

**<Table 1 - Database file tables and included column name and descriptions>**[[15]](#footnote-14)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table** | **Column Name** | **Column Description** | **Table** | **Column Name** | **Column Description** |
| Patent | type | Category of patent, e.g. “design”, “reissue” | RawAssignee | uuid | Unique identifier for raw assignee record |
| id | Unique patent document number | patent\_id | Patent which contains this record |
| Country | Country of origin of patent | assignee\_id | Identifier of disambiguated assignee record |
| date | Date of grant of patent | rawlocation\_id | Location of raw assignee |
| abstract | Text of patent abstract | type | USPTO code for type of assignee |
| title | Text of patent title | name\_first | First name of assignee (if individual) |
| kind | USPTO code for type of patent | name\_last | Last name of assignee (if individual) |
| num\_claims | number of claims made by patent | organization | Name of assignee’s organization (if firm) |
| RawInventor | id | Unique raw inventor identifier | sequence | Order in which this assignee was listed on the patent |
| patent\_id | Patent which contains this record | Claim | uuid | Unique identifier for claim record |
| rawlocation\_id | Location of raw inventor | patent\_id | Corresponding patent document for this claim |
| name\_first | First name of inventor | text | Text of claim |
| name\_last | Last name of inventor | dependent | Sequence number of claim this record is dependent on |
| nationality | Nationality of inventor | sequence | Order in which this claim appears in its patent |
| sequence | Order that this inventor was listed on the patent |  | | |
| **Table** | **Column Name** | **Column Description** | **Table** | **Column Name** | **Column Description** |
| uspc | uuid | Unique  location identifier | ipcr | uuid | Unique identifier for citation relation |
| patent\_id | Raw city name | patent\_id | Patent making a citation |
| mainclass\_id | Main-class classification level | classification level | patent classification level |
| subclass\_id | Sub-class  classification level | sequence | Order in which the focal patent cited the document |
| ... | ... | ... | ... |

**Methodology**

Grouping the bulk data by organizations, dates and performing text analysis on patent contents can generate corpus and timeline of corporate’s interested technology. Comparison of corpus and measuring cosine similarity of vectorized words can provide technological similarity index of respective corporations. The propagation of technological interest of respective corporate can be drawn as a network, using patent issued dates.

**Key Attributes**

The USPTO data file contains numerous relations (tables) and attributes (columns). Among the relations, *“Patent”* and *“Raw Assignee”* are the constitutional subjects of patent analysis. The three relations include *“title”, “abstract”, “type”, “name\_first”, “name\_last”* which are general information of what the granted patent is about , who submitted the patent and who are authorized and assigned to the patents. The identification numbers such as “patent\_id” are used as foreign key and primary key for respective patents.

**Company Selection**

Based on the Fortune 500 rankings, top 8 companies with the intersection of highest market value, asset , revenue in the technology sector (as of March 29, 2018) and overall ranking above 100 were selected.[[16]](#footnote-15) Technology was chosen as a sector, assuming that patent is correlated to technological advantagesand best reflects its influence. The companies in list are: Apple, Alphabet, Microsoft, Facebook, Intel, Cisco Systems, Oracle and IBM.

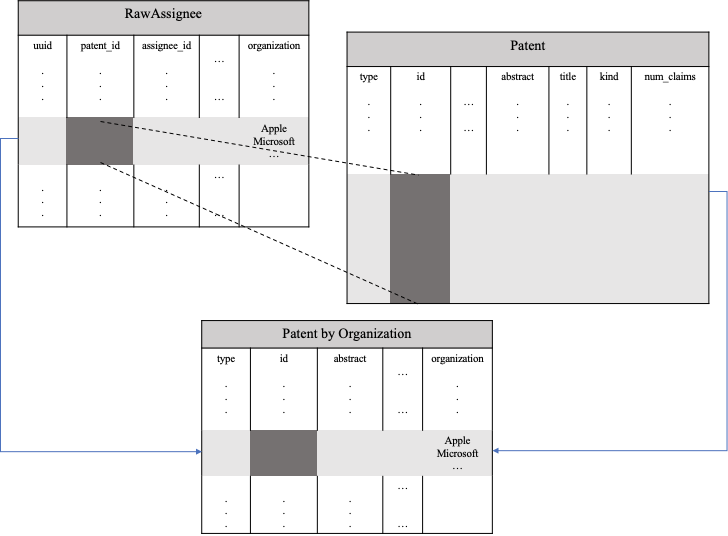
**<Table 2 - Top 8 Technology Sector Company >**

|  |  |  |
| --- | --- | --- |
| **Market Value** | **Asset** | **Revenue** |
| Apple (4)\* | Apple | Apple |
| Alphabet (22) | Microsoft | Alphabet |
| Microsoft (30) | Alphabet | Microsoft |
| Facebook (76) | Oracle | IBM |
| Intel (46) | Cisco Systems | Intel |
| Cisco Systems (62) | IBM | Cisco Systems |
| Oracle (82) | Intel | Facebook |
| IBM (34) | Facebook | Oracle |

\*number in the brackets are original Fortune 500 rankings (2018)

**Preprocessing : Company Name Recognition**

As *“Patent”* table did not contain information about the assignees, querying and joining data table was required to retrieve patent data of the selected 8 companies. Using python3 and sqlite3, tuples of *“patent\_id”* that included names of 8 organization were queried.

**<Table 3 - Patent by Organization>**

However, it was difficult to attain pertinent patents by organizations, because of inconsistency in corporate names. Perfunctory search was inadequate for retrieving all records relevant to the assignee names. For example, selecting records by organization name “Apple” presented following results:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Input:

> c.execute("SELECT DISTINCT organization   
 FROM rawassignee   
 WHERE organization   
 LIKE{} ".format('"apple%"'))

> apple\_names = [tuple[0] for tuple in c.fetchall()]  
> print(apple\_names)

-----------------------------------------------

Output:

[Apple Inc.', 'Apple Inc', 'Appleton Papers Inc.', 'Appleton Steel, Inc.', 'Applera Corporation',

'Apple, Inc.', 'Apple, Inc', 'Apple Computer, Inc.', 'Applegate Livestock Equipment, Inc.',  
 'Apple Ju Studio Inc.', 'APPLE Inc.', 'APPLE, INC.', 'APPLE INC.', 'Applexion', 'APPLE Inc', 'Apples-To-Go, Inc.', 'Apple inc.', 'Appleton Papers, Inc.', 'Appleton Papers Inc', 'Apple Biomedical, Inc.',  
 'Appleton Coated, LLC', 'Apple Biomedical Inc.', 'Appleton Grp LLC', 'APPLE INC',  
 'APPLEJACK 199, L.P.', 'APPLEJACK 199 L.P.', 'Appled Materials Israel Ltd.', 'APPLE']

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

The query result of selecting “Apple” includes irrelevant but similar corporate names such as “Apple Biomedical Inc, Appleton Grp LLC, APPLEJACK 199”. The case sensitivity, existence and location of punctuation marks makes it more difficult to retrieve desired outcomes. Hence, the name recognition of selected organizations performed by manual screening. After the corporate entity recognition process, querying was performed and relation table that represents patent by organization was created using the following code:

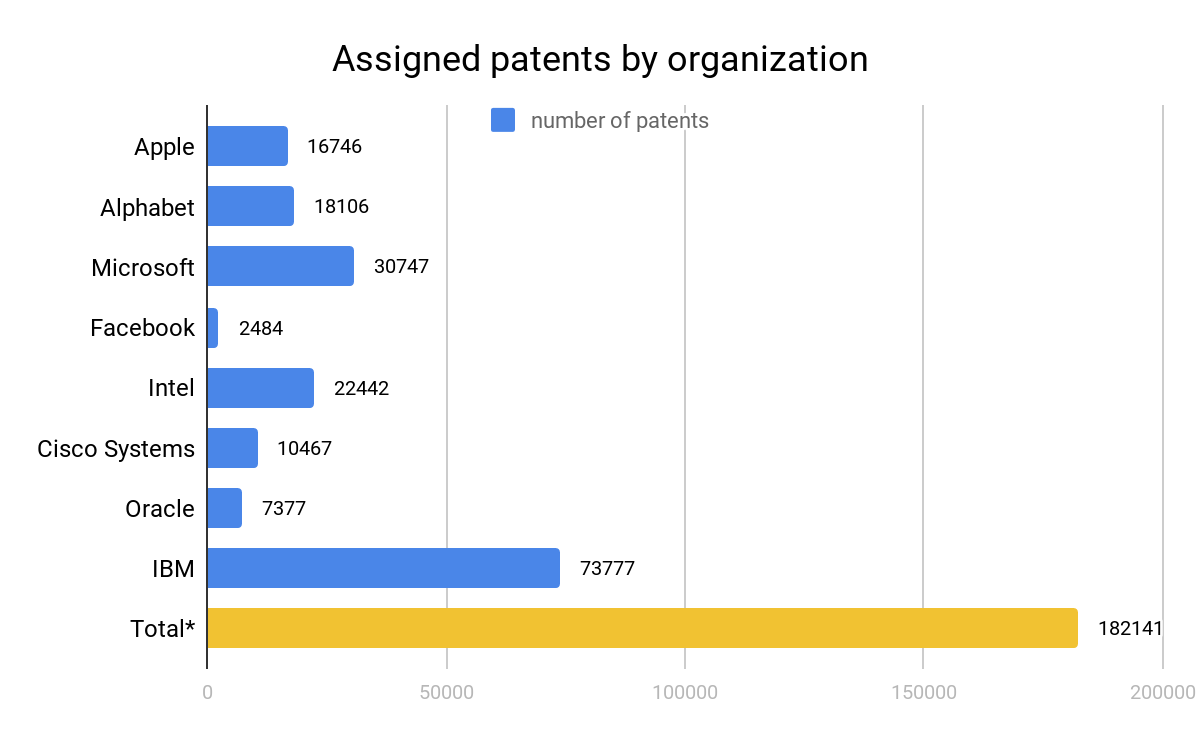
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

> query = c.execute("""SELECT \* FROM (SELECT patent\_id, name\_first, name\_last,   
 GROUP\_CONCAT(organization   
 FROM rawassignee   
 WHERE patent\_id IN {}  
 GROUP BY patent\_id) AS sample

INNER JOIN patent  
 ON patent.id = sample.patent\_id""".format(tuple(corp\_id)))

**Descriptive Analysis**

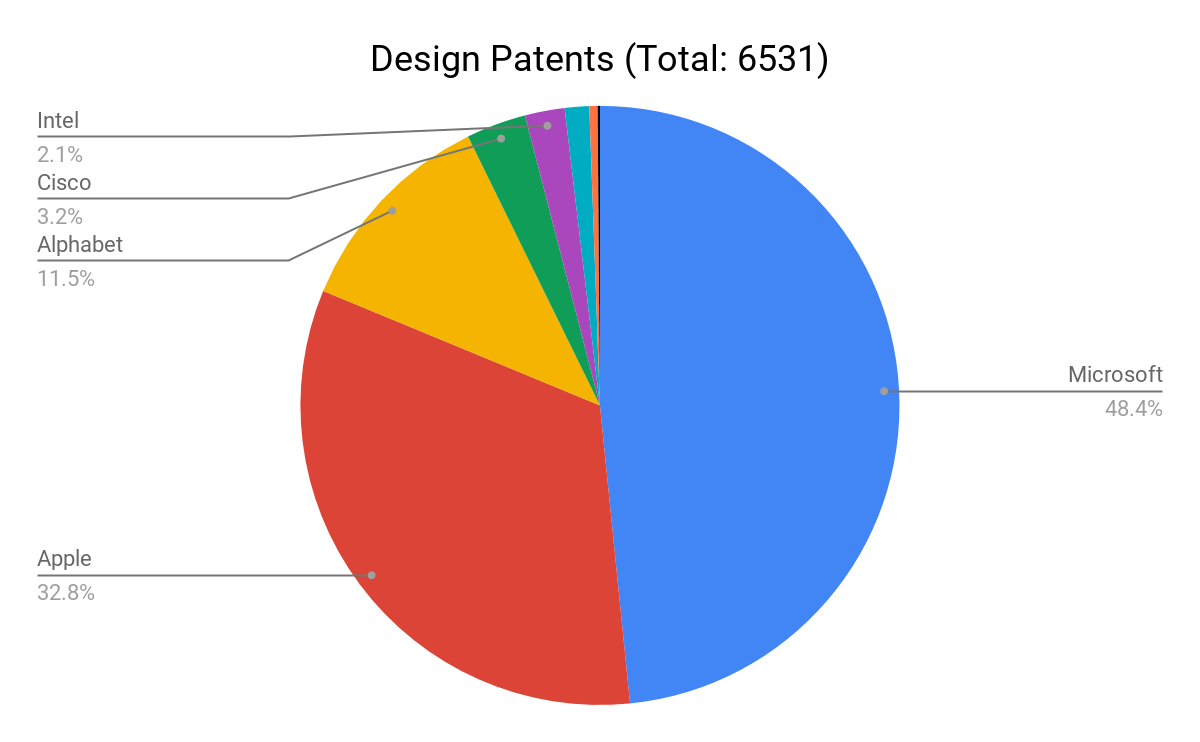
As illustrated in figure 3, IBM has predominant number of assigned patents. It accounts for 40% of the total number of assigned patents of all eight organizations. Although quantity does not necessarily promise quality and significance of technology, the number itself is quite remarkable.



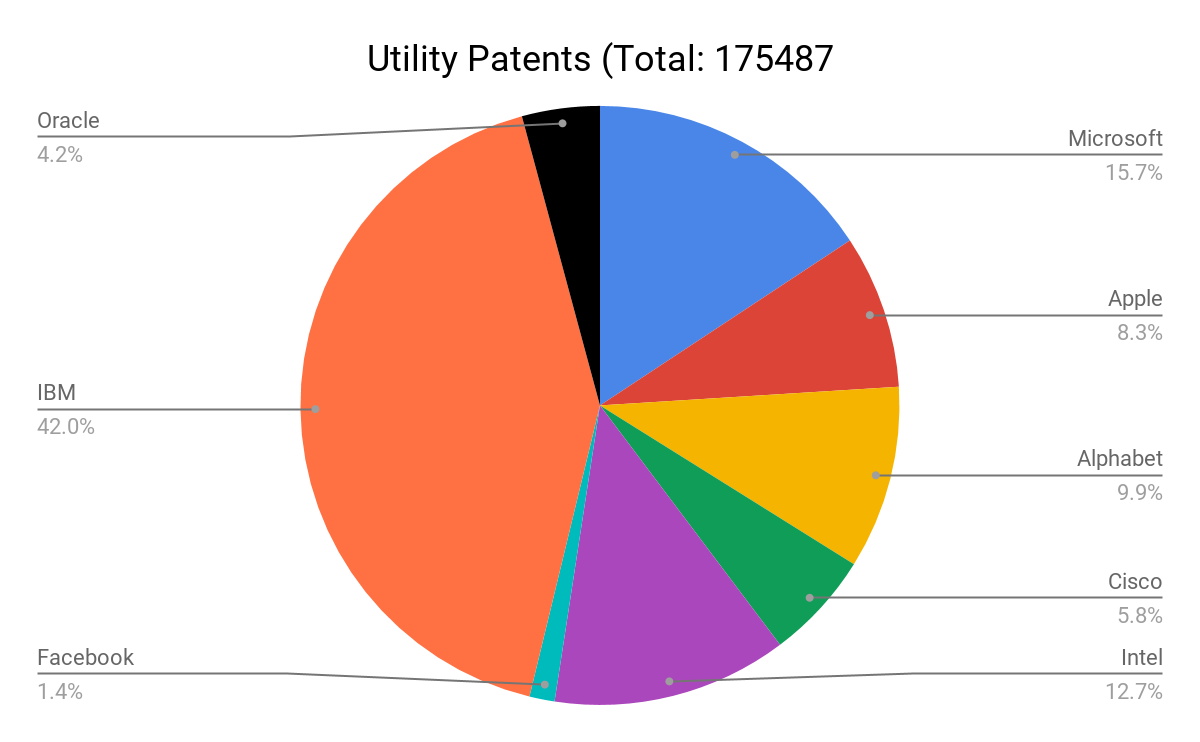
**<Figure 3 - Number of Assigned Patent by Selected Organizations>[[17]](#footnote-16)**

There are three types of patents: Utility patents, Design patents and Plant patents.[[18]](#footnote-17) Utility patents are granted to anyone who invents or discovers new, practical process or tangible machines or improvements. Design patents are granted for original and ornamental designs and plant patents are subject to distinct and new variety of plant discovery. Among the 182141 patents, 96% were utility patents, while 3.5% were design patents. The residual percentages were from re-issued patents and statutory invention registration (sir).

Figure 4 and 5, respectively shows the percentage of type of patents by the eight organizations. In design patents, Microsoft and Apple has more than 80% of share in granted design patents. Alphabet accounted for about 11.5 %. Meanwhile, IBM, Microsoft and Intel constitute 80% of utility patents. It was notable that Apple were leading in design patents with Microsoft, but falling behind in terms of the quantity of utility patents. IBM was invisible in design patents, but showed its significance in utility patents. Microsoft and Alphabet presented balanced shares in both design and utility patents.

****

**<Figure 4 - Number of Assigned Design Patent by Selected Organizations>**

****

**<Figure 4 - Number of Assigned Utility Patent by Selected Organizations>**

**Text Analysis:**

1. Thomas Bollyky (10 April 2013). ["Why Chemotherapy That Costs $70,000 in the U.S. Costs $2,500 in India"](https://www.theatlantic.com/health/archive/2013/04/why-chemotherapy-that-costs-70-000-in-the-us-costs-2-500-in-india/274847/). *The Atlantic*. The Atlantic Monthly Group. Retrieved 18 April 2013 [↑](#footnote-ref-0)
2. Fabry, Bernd & Ernst, Holger & Langholz, Jens & Köster, Martin, 2006. "[Patent portfolio analysis as a useful tool for identifying R&D and business opportunities--an empirical application in the nutrition and health industry](https://ideas.repec.org/a/eee/worpat/v28y2006i3p215-225.html)," [World Patent Information](https://ideas.repec.org/s/eee/worpat.html), Elsevier, vol. 28(3), pages 215-225, September [↑](#footnote-ref-1)
3. Albright, R.E., 2002. What can past technology forecasts tell us about the future? Technol.

   Forecast. Soc. Chang. 69 (5), 443–464 [↑](#footnote-ref-2)
4. Park, Y., Yoon, B., Lee, S., 2005. The idiosyncrasy and dynamism of technological innovation

   across industries: patent citation analysis. Technol. Soc. 27 (4), 471–485. [↑](#footnote-ref-3)
5. Trappey, A.J.C., Trappey, C.V., Wu, C.-Y., Lin, C.-W., 2012. A patent quality analysis for innovative

   technology and product development. Adv. Eng. Inform. 26 (1), 26–34. [↑](#footnote-ref-4)
6. Park, S., Lee, S., Jun, S., 2015. A network analysis model for selecting sustainable technology.

   Sustainability 7, 13126–13141 [↑](#footnote-ref-5)
7. Pénin, J., 2012. Strategic uses of patents in markets for technology: a story of fabless firms,

   brokers and trolls. J. Econ. Behav. Organ. 84 (2), 633–641. [↑](#footnote-ref-6)
8. Assad A., Limin Z., Samee U. K., 2014, A literature review on the state-of-the-art in patent analysis,

   World Patent Information,Volume 37, 3-13. [↑](#footnote-ref-7)
9. Yoon J, Park H, Kim K. Identifying technological competition trends for R&D planning using dynamic patent maps: SAO-based content analysis. Scientometrics 2013:1e19 [↑](#footnote-ref-8)
10. Park H, Ree JJ, Kim K. Identification of promising patents for technology transfers using TRIZ evolution trends. Expert Syst Appl 2013:736e43 [↑](#footnote-ref-9)
11. Park H, Yoon J, Kim K. Identifying patent infringement using SAO based semantic technological similarities. Scientometrics 2012;90(2):515e29. [↑](#footnote-ref-10)
12. Kim YG, Suh JH, Park SC. Visualization of patent analysis for emerging technology. Expert Syst Appl 2008;34(3):1804e12. [↑](#footnote-ref-11)
13. Chang PL, Wu CC, Leu HJ. Using patent analyses to monitor the technological trends in an emerging field of technology: a case of carbon nanotube field emission display. Scientometrics 2010;82(1):5e19. [↑](#footnote-ref-12)
14. https://github.com/funginstitute/patentprocessor [↑](#footnote-ref-13)
15. https://funginstitute.berkeley.edu/wp-content/uploads/2014/06/patentprocessor.pdf [↑](#footnote-ref-14)
16. http://fortune.com/fortune500/list/filtered?sortBy=mktval&sector=Technology [↑](#footnote-ref-15)
17. \*Co-assigned patents are counted as 1 in total number of patents [↑](#footnote-ref-16)
18. https://www.uspto.gov/patents-getting-started/general-information-concerning-patents [↑](#footnote-ref-17)