Paper 101-2013

Data Mining of U.S. Patents: Research Trends of Major Technology Companies

Kenneth M. Potter, Science Applications International Corporation, Huntsville, Ala.
Robert N. Hatton, Science Applications International Corporation, Huntsville, Ala.

ABSTRACT

Research initiatives are normally closely held corporate secrets. Insights to research trends are difficult to extract from public information, but data mining of the U.S. Patent and Trademark Office (USPTO) patent grants provides an opportunity to expose interesting trends and areas of interest as indicated by activity in related patent areas. This paper covers assessing the vast USPTO information repository and the analytical methodology that extracts patent grant information from multiple formats and produces interesting insights into research trends for several major technology companies.

KEYWORDS Data & Text Mining; Advanced Analytics; Research, Development, Test and Engineering, U.S. Patent and Trademark Office, Technology Trends, SAS®, SAS® Enterprise Guide, SAS® Enterprise Miner™ and Text Miner, cURL

INTRODUCTION

The U.S. Patent and Trademark Office (USPTO) is the official grantor of patents in the United States. The USPTO provides multiple online search mechanisms for obtaining patent information from their electronic patent products site,[1] including both images and full text. Recently, Google has started to provide full text patent information[2] in various bulk download formats at no cost.

The USPTO states their average file size is 54 MB (weekly compressed patents). For each year the estimated size would be 2,808 MB (compressed format). The total size for 1976-2011 is estimated to be over 98 GB of data (compressed). Extracting patent information from each compressed file resulted in approximately 400 MB (weekly uncompressed patents), 20.8 GB (yearly uncompressed patents) for a total of 728 GB (uncompressed patents) for 1976-2011.

Given the vast amount of data that can be extracted from the USPTO site, it would be time-consuming and require a large amount of storage space to assemble a single data repository and to assess this data using SAS Institute Inc. software. The authors investigated multiple research paths to determine a reasonable approach to deal with the large quantity of available data. Ultimately, the USPTO compressed files were downloaded and then extracted to form the corpus of data that was then manipulated via both base SAS® code and SAS® Enterprise Miner™ and Text Miner software.

The research purpose was crystallized to identify significant associated terms with various contemporary technology initiatives and to expose technology trends by major technology companies since 1990. Ultimately, the research would provide insights to what areas major technology companies are investigating as indicated by granted patents.

DATA PREPARATION

New U.S. patent grants are published every Tuesday by the USPTO. Through an agreement reached in June 2010, The USPTO makes all patent grant and application data from 1976 to present available for bulk download through Google® Books. The data are available as one .ZIP file for each week. For the purposes of this paper, we analyzed patent grants but did not account for applications. In addition, we limited our search to only those patent grants published between January 1, 1990 and September 25, 2012.

The Data Preparation Phase consisted of three steps:

1. Download target data (compressed)
2. Decompress target data
3. Extract assignee information across entire corpus to create a lookup table for mapping assignees to patent numbers

The first step in acquiring the desired data, given that we were intending to download 1,187 separate files, was to write a script to automate the download and organization of the bulk data provided through Google in ZIP format. Fortunately the file naming convention used by the USPTO and Google is nearly consistent throughout our target

[1] http://www.uspto.gov/products/catalog/patent_grants.jsp
[2] www.google.com/googlebooks/uspto-patents-grants-text.html

timeframe, which allowed us to easily generate a list of target URLs for our script. A sample list of our target URLs appears in Figure 1 below.

```
http://storage.googleapis.com/patents/grant_full_text/1990/pftaps19900102_wk01.zip      19900102
http://storage.googleapis.com/patents/grant_full_text/1990/pftaps19900109_wk02.zip      19900109
http://storage.googleapis.com/patents/grant_full_text/1990/pftaps19900116_wk03.zip      19900116
http://storage.googleapis.com/patents/grant_full_text/1990/pftaps19900123_wk04.zip      19900123
http://storage.googleapis.com/patents/grant_full_text/1990/pftaps19900130_wk05.zip      19900130
http://storage.googleapis.com/patents/grant_full_text/1990/pftaps19900206_wk06.zip      19900206
http://storage.googleapis.com/patents/grant_full_text/1990/pftaps19900213_wk07.zip      19900213
http://storage.googleapis.com/patents/grant_full_text/1990/pftaps19900220_wk08.zip      19900220
http://storage.googleapis.com/patents/grant_full_text/1990/pftaps19900227_wk09.zip      19900227
```

Figure 1: Sample List of URLs

Our URL list was a simple text file with two columns. The first column was the URL for the file we wanted to download, and the second column was the file's publication date in yyyymmdd format. The publication date column was included for file naming purposes to better organize our downloaded files. We then created a Windows® batch script that would read each URL from the list, download each target file using cURL[3] for Windows, and save the resulting .ZIP file to a directory on our server. The batch script code is as follows:

```
for /F "tokens=1,2" %%i in (E:\PatentAnalysis\1990_Google_Links.txt) do curl.exe --output "E:\PatentAnalysis\PatentDownloads\%%j.zip" "%%i"
```

Figure 2: Windows Script Using cURL to Download and Save Compressed Files

In our one-line script above, "E:\PatentAnalysis\1990_Google_Links.txt" is the text file containing our list of URL and date pairs (in this case, one entry for each Tuesday in 1990) and "E:\PatentAnalysis\PatentDownloads\yyyymmdd" is the file path to which we want to save our downloaded file, where yyyymmdd is the formatted publication date (i.e., that particular Tuesday's date as read from the second column of our text file). We actually duplicated this script for each year in our target range and ran each individually to facilitate testing, monitor for errors, and limit impact to our corporate Internet bandwidth. However, the script could easily be combined to process the entire range of years by simply creating a single text file containing all 1,187 URLs and corresponding dates.

The second step in our data preparation was to decompress the 1,187 .ZIP files that we downloaded in the previous step. To accomplish this, we created another short script to automate the extractions using WinZip® and a separate add-on called WinZip Command Line. First, we generated another text file very similar to our list of URLs created during the previous step. This time, instead of containing a URL and a publication date, our two columns contained the path to the downloaded .ZIP file to be decompressed and the target location where we wanted to save the extracted contents, as shown below in Figure 3.

```
E:\PatentAnalysis\PatentDownloads\1990_ZipFiles\19900130.zip     E:\PatentAnalysis\PatentDownloads\ExtractedDownloads1990
E:\PatentAnalysis\PatentDownloads\1990_ZipFiles\19900206.zip     E:\PatentAnalysis\PatentDownloads\ExtractedDownloads1990
E:\PatentAnalysis\PatentDownloads\1990_ZipFiles\19900213.zip     E:\PatentAnalysis\PatentDownloads\ExtractedDownloads1990
E:\PatentAnalysis\PatentDownloads\1990_ZipFiles\19900220.zip     E:\PatentAnalysis\PatentDownloads\ExtractedDownloads1990
E:\PatentAnalysis\PatentDownloads\1990_ZipFiles\19900227.zip     E:\PatentAnalysis\PatentDownloads\ExtractedDownloads1990
E:\PatentAnalysis\PatentDownloads\1990_ZipFiles\19900306.zip     E:\PatentAnalysis\PatentDownloads\ExtractedDownloads1990
E:\PatentAnalysis\PatentDownloads\1990_ZipFiles\19900313.zip     E:\PatentAnalysis\PatentDownloads\ExtractedDownloads1990
E:\PatentAnalysis\PatentDownloads\1990_ZipFiles\19900320.zip     E:\PatentAnalysis\PatentDownloads\ExtractedDownloads1990
E:\PatentAnalysis\PatentDownloads\1990_ZipFiles\19900327.zip     E:\PatentAnalysis\PatentDownloads\ExtractedDownloads1990
E:\PatentAnalysis\PatentDownloads\1990_ZipFiles\19900403.zip     E:\PatentAnalysis\PatentDownloads\ExtractedDownloads1990
```

Figure 3: Text file with two columns; one for file to be decompressed and second with target location

The script code that used this text file is as follows:

```
for /F "tokens=1,2" %%i in (E:\PatentAnalysis\Zip_file_Links_1990-1999.txt) do "C:\Program Files\WinZip\wzunzip" "%%i" "%%j"
```

Figure 4: Windows Script to Decompress Zipped Files and Save to Designated Location

This single line of code combined with the text file above extracted all 1,187 .ZIP files into a series of folders, one for each year. It is important to note that the scripts above do not actually create the directory structure, but rather assume that the desired folder structure has been created beforehand. Again we partitioned this process, this time into two segments of 1990-1999 and 2000-2012, in order to better isolate any errors and to limit the impact to other processes running on the host server.

With all of the desired data downloaded and decompressed, we were ready to start parsing the corpus and converting it to SAS data sets. However, for such a large amount of data, the time and computing resources that would be required to convert all patents since 1990 into SAS data sets was not feasible given our limited computing resources. We had allocated only a single virtual machine with two processors and 2 GB of RAM to execute this research. Since we were only interested in a few companies whose data we wanted to explore, this meant that we were really only interested in a very small percentage of the total patent data. Therefore, converting the entire corpus to SAS data sets would be a waste of time and resources as most of the data would not be used. We decided that a

[3] cURL is a command line tool for transferring data with URL syntax. See http://curl.haxx.se/

better approach would be to leave our source data in its raw format and selectively parse only those patents held by the companies we had selected for analysis. This meant that we needed a way to identify and filter out just those patents held by a specific company.

The third step of our data preparation is where SAS programming came into the process. In order to search for patents held by a specific company, we created a SAS program that would parse all documents in the corpus to catalog the assignees for each patent granted since 1990. An obstacle to achieving this was the variation in format of the patent grant documents over the years. The data from 1976 through 2000 (though we were only concerned with 1990 and later), is provided in text file format. The data for 2001 is provided as Standard Generalized Markup Language (SGML). The data from 2002 to present is provided in XML format. We wrote a SAS program that would read iteratively through our folder structure to identify each source file and then pass each file's path to one of three macros corresponding to each of the three file formats. The three macros performed the same function but each one handled a different format and structure for the source data.

It is important to note here that the XML source data files are concatenated such that the <us-patent-grant> element, which would otherwise be the root element for a patent grant, occurs thousands of times in each source file. Therefore, the XML files in their original state cannot be opened by typical applications that parse XML files, nor can they be utilized by SAS® XML Mapper without prior manipulation.

The code below is the key portion of the GetFileNames macro that accepts a folder path as a parameter and iterates through each file in the target folder. The target folder name ends with the last four characters of the year. These last four characters are extracted to the SAS macro variable &gYear. Based on the value of &gYear, the GetFileNames macro passes the file path to the appropriate macro, either %AssigneesPre2001, %Assigness2001-2004 or %AssigneesPost2004.

```
%LET i = 1;
%DO %WHILE (&i <= &num_items);
    %LET item_name = %SYSFUNC(DREAD(&dir_id, &i));
    %LET filePath = %UNQUOTE(&directoryPath)\&item_name;
    %IF %EVAL(&gYear) < 2001 %THEN %AssigneesPre2001("&filePath");
    %IF %EVAL(&gYear) >= 2001 AND %EVAL(&gYear) <= 2004 %THEN %Assignees2001-
2004("&filePath");
    %IF %EVAL(&gYear) > 2004 %THEN %AssigneesPost2004("&filePath");
    %LET i = %EVAL(&i + 1);
%END;
```

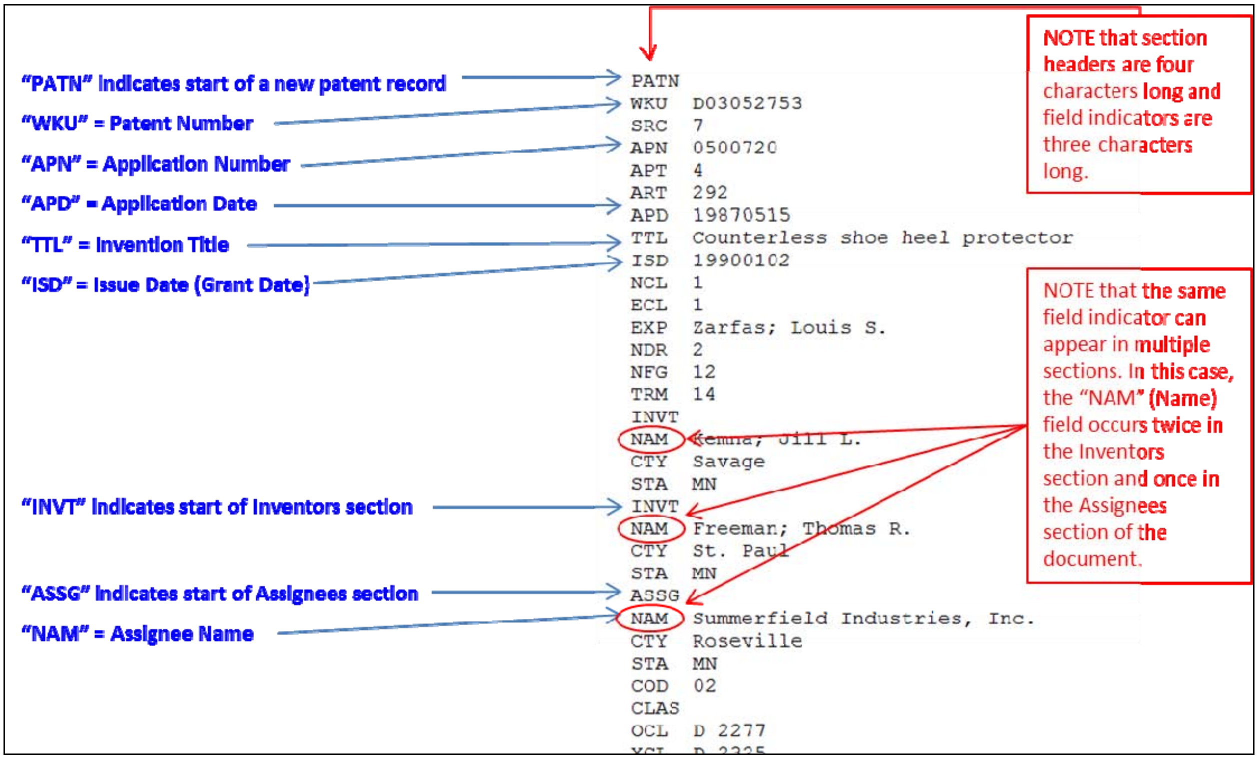

Figure 5: Sample of the Raw Data for a Pre-2001 Patent Grant in Text File Format

The code below is a snippet of the %AssigneesPre2001 macro. This code processes the patent grant files that are provided in text file format.

```
DATA work.u_assignees(keep=grant_num grant_date assignee assn_role bulk_file)
work.d_assignees(keep=grant_num grant_date assignee assn_role bulk_file);
    INFILE srcFile DLM='0A'X LRECL=5000 TRUNCOVER;
    FORMAT sectionTag $5. lineValue $100. docSection $25. grant_num $15. grant_date
mmddyy10. assignee $200. assn_role $2. bulk_file $200.;
    INPUT sectionTag $ 1-5 lineValue $;
    IF docSection = 'Publication Reference' AND LENGTH(sectionTag) = 4 THEN
docSection = 'other';
    IF sectionTag = 'PATN' THEN docSection = 'Publication Reference';
    IF docSection = 'Publication Reference' THEN DO;
           IF sectionTag = 'WKU' THEN grant_num = lineValue;
           IF sectionTag = 'ISD' THEN grant_date = INPUT(lineValue, yymmdd8.);
    END;

    IF docSection = 'Assignees' AND sectionTag = 'NAM' THEN assignee = lineValue;
    IF docSection = 'Assignees' AND sectionTag = 'COD' THEN assn_role = lineValue;
    IF docSection = 'Assignees' AND LENGTH(sectionTag) = 4 THEN DO;
           bulk_file = &assnFilePath;
           IF SUBSTR(grant_num,1,1) = 'D' THEN OUTPUT work.d_assignees;
                  ELSE IF SUBSTR(grant_num,1,1) ^= 'P' THEN OUTPUT work.u_assignees;
           docSection = 'other';
           assignee = .;
           assn_role = .;
    END;
    IF sectionTag = 'ASSG' THEN docSection = 'Assignees';
    RETAIN docSection grant_num grant_date assignee assn_role;
RUN;
```

Due to the nature of the source data, the SAS data step must use indicators in the data to partition each document into sections, such as the Inventors section and the Assignees section. This is important because the field names are not unique and can be repeated in more than one section, as seen with the “NAM” (Name) field in Figure 6 above. Our DATA step uses a variable called docSection to keep track of the section to which each row belongs. In addition to handling multiple occurrences of the same field name, this also allows us to trigger conditional logic or to handle data in various document sections in different ways. For example, when the DATA step above reaches a row that signifies the end of the Assignees section, it executes the OUTPUT statement to write the current observation to the work.u_assignees or work.d_assignees data set based upon the patent number indicating that it is either a design or utility patent, respectively. The end result of parsing all of the assignees in this manner is a SAS data set containing the grant number, grant date, assignee name, assignee code (for potential future use), and directory path of the source file. This data set is then available to filter out just those patent numbers corresponding to a company that we would like to analyze.

ANALYSIS APPROACH

In order to filter out just the patents for a specific company, we created a SAS program called FindCompanySubset. This program allows for manual entry of a company name via a SAS macro variable. This macro variable value is then compared against the indexed assignee data for all patent grants in the entire corpus by using the FINDW function in a subsetting IF statement. The result is the company_subset data set containing all of the patent numbers related to the target company name and another data set called unique_assignees that contains the unique list of assignee names containing the search term(s). There are often many (sometimes hundreds) of variations of a company’s name, some of which may contain the generic search term but are actually not related to the target company. For example, a search for “ASUS” would also return results for a company with “Pegasus” in its name. These variants in assignee values were manually reviewed and assessed for relevance. Any assignee values not deemed relevant were manually deleted from the unique_assignees data set, which was then used in the next step.

A SAS program called ParsePatents was created to extract the individual patent documents from the raw data. This program first creates a hash table based on the modified unique_assignees data set. It then iterates through the company_subset data set, using the hash table to verify that the patent belongs to the target company and not to one of the manually deleted company name variants. If the observation’s assignee value is not found in the hash table, then the observation is discarded leaving only patents belonging to the target company. The ParsePatents program then uses the remaining observations in the company_subset data set to generate the file_paths data set, which contains the unique set of source file names found in the company_subset data. Generating the file_paths dataset allows parsing of only those source files which contain patents granted to the target company. After generating the list of files to be processed, each source file’s path is then passed to one of three SAS macros that parse the text, SGML, and XML variations of the patent grant documents.

The key DATA step in the ParsePatents program:

```
DATA _null_;
    SET pats.file_paths;
    PUT “PROCESSING FILE “ bulk_file;
    IF grant_year < 2001 THEN DO;
           Put “Grant_year < 2001 “;
           CALL EXECUTE(‘%ParsePatentText(‘||bulk_file||’)’);
    END;
    IF grant_year >= 2001 AND grant_year <= 2004 THEN DO;
           put "Grant_year between 2001 and 2004 ";
           CALL EXECUTE('%ParsePatentSGML('||bulk_file||', '||grant_year||')');
    END;
    IF grant_year > 2004 THEN DO;
           CALL EXECUTE('%ParsePatentXML('||bulk_file||')');
           put "executing grant_year > 2004 ";
    END;
RUN;
```

The ParsePatents program iterates through the file_paths dataset, thereby parsing only the raw source files that contain patents for the target company. The program then passes the source file’s path to one of three SAS macros that parse patent grant documents in text, SGML, and XML formats, respectively. The ParsePatents program and the invoked macros generate two SAS data sets called d_patents (design patents) and u_patents (utility patents). These data sets represent all of the patents granted to the selected company between 1990 and September 18, 2012. Some key portions of the DATA step used to extract data from a grant document in text format are described below.

The INPUT statement introduces two important DATA step variables, the sectionTag variable, which is essentially the name of the field, and the lineValue variable for capturing the actual content of the row.

```
INPUT sectionTag $ 1-5 lineValue $;
```

As each new row is read, the first step is to determine whether a new section of the document has been reached, for example when moving from the basic patent publication information (the “PATN” section in Figure 5 above) to the section which lists the names of the inventors (the “INVT” section in Figure 5 above). Based on the sectionTag value that is read from the source file, the docSection variable is set to track which section of the document is currently being processed, as seen in the following example:

```
/*Determine which section of the document is currently being processed (retains
value from line-to-line, only changes when a new section is encountered);*/
IF SUBSTR(sectionTag,1,4) = "PATN" THEN docSection = 'Publication Reference';
IF SUBSTR(sectionTag,1,4) = "ABST" THEN docSection = 'Abstract';
IF SUBSTR(sectionTag,1,4) = "INVT" THEN docSection = 'other';
```

When the end of the Abstract section is reached (i.e., the docSection variable equals “Abstract” and a new section is indicated by a four-character section heading other than “ABST” is indicated by the sectionTag variable), the current observation is written to the d_patents or u_patents data set in the temporary WORK library using the OUTPUT statement, and the grant attributes are set to MISSING in preparation for reaching the next patent grant in the text document.

```
IF docSection = 'Abstract' AND LENGTH(sectionTag) = 4 AND sectionTag ^= 'ABST' THEN
DO;
    IF SUBSTR(grant_num,1,1) = 'D' THEN OUTPUT work.d_patents;
        ELSE IF SUBSTR(grant_num,1,1) ^= 'P' THEN OUTPUT work.u_patents;
    docSection = .;
    grant_num = .;
    grant_date = .;
    app_num = .;
    app_date = .;
    inv_title = .;
    abstract = .;
    rc = .;
END;
```

If the end of the Abstract section has not yet been reached, the macro continues reading the rows of the text file. The following piece of code shows how the basic publication information such as grant date and application number are collected for the patent grant:

```
        IF docSection = 'Publication Reference' THEN DO;
              * Capture the grant date;
              IF sectionTag = 'ISD' THEN grant_date = INPUT(lineValue, yymmdd8.);
              * Capture the application number;
              IF sectionTag = 'APN' THEN app_num = lineValue;
              * Capture the application date;
              IF sectionTag = 'APD' THEN app_date = INPUT(lineValue, yymmdd8.);
```

Once all of the text files have been parsed, the resulting u_patents data set contains the abstracts for all utility patents for the target company. Design patents do not have abstracts and were captured but not utilized for this research project. It is the abstract information in each company's u_patents dataset that formed the basis for the remainder of the research.

As this research effort progressed, insights gained helped shape the final analysis products. As a first step we outlined the general research goal as identify major technology companies' intellectual property trends through data mining and advanced analytics applied to the USPTO holdings. We thought that recent technical publications would provide a starting list of contemporary technical topics that would form the starting list for exploration of the USPTO holdings.

Two major research thrusts were selected, (1) technologies related to smart phones and (2) technologies related to video displays (see **Figure 6**).

Thrust 1: Smart Phones		Thrust 2: Video Displays	
3G	4G	Liquid Crystal Displays (LCDs)	Light Emitting Diodes (LEDs)
Displays	Tablets	Plasma Screens	Organic Light-Emitting Diode (OLED)
Mobile	Wi-Fi	Cathode Ray Tube (CRT)	Digital Light Processing (DLP)
portable media players	pocket video cameras,	Active-Matrix Liquid Crystal Display (AMLCD)	Field Emission Display (FED)
compact digital cameras	GPS navigation	SED-tv	Thin-Film Transistors (TFT)
Personal Digital Assistant (PDA)	battery life	Quantum Dot Display (QLED)	

Figure 6: Key Terms for Smart Phones and Video Displays

Technology Companies			
Apple, Inc.	ASUS®	Dell, Inc.	Foxconn®
Hewlett-Packard® (HP®)	International Business Machines Corporation (IBM)	LG Corporation	Motorola, Inc.
Microsoft Corporation	Panasonic Corporation	Toshiba Corporation	Samsung®
Sony Corporation			

Figure 7: Top Technology Companies by Total Revenue or Market Presence

To focus the research, we selected a small group of leading technology companies to subset our data (see **Figure 7**). We anticipated that in each of these selected thrusts our research would expose interesting association of technical terms that would lead to additional research and data analysis. We use Concept Linking in SAS Enterprise Miner™ software to identify associated terms with the key term list in **Figure 6**.

Distribution analysis of patents by year by assignee (company) was determined using SAS Enterprise Guide 4.3. The distribution for the 13 selected companies is shown in **Figure 8**. IBM has the largest number of utility patents for this period followed by Samsung®, Toshiba and Sony. With the recent, highly publicized patent lawsuits between

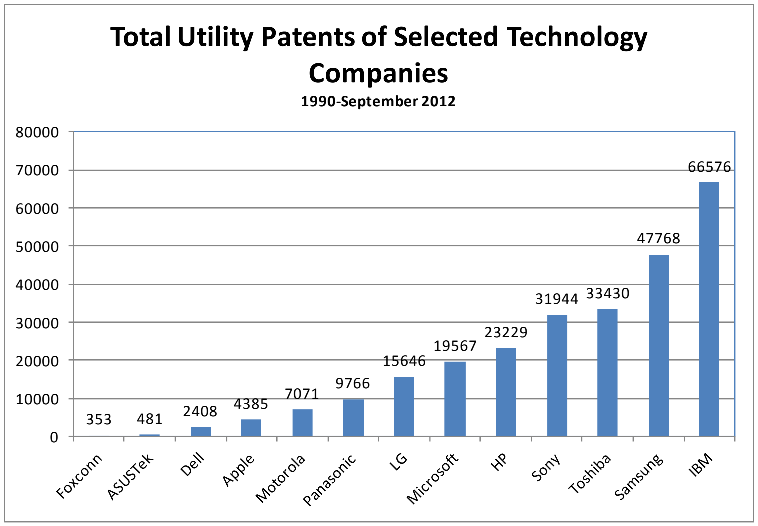

Figure 8: Distribution of Utility Patents by Year for Selected Companies

Apple and Samsung, the authors expected Apple to have an active patent effort and subsequently a much larger number of patents. **Figure 8** indicates Apple is not a major contributor. **Figure 11** shows Apple has significantly increased its patent holding in recent years.

With 13 companies in our subset and such a variation in the number of patents held, we decided to subset the time distribution in approximately thirds. **Figures 9-11** show the time distributions of patents held by our company subset.

Of note in **Figure 9** is the closeness of Samsung and IBM in the number of patents granted from 2007 to 2012.

IBM has often been cited as a leading patent producer. Samsung, at least in the past few years, seems to have nearly matched IBM's number of granted patents.

Figure 9 also shows a significant increase, in the past decade, in the number of grants awarded. Further investigation will seek to determine if this trend is reflected in the key technology areas of interest in this paper.

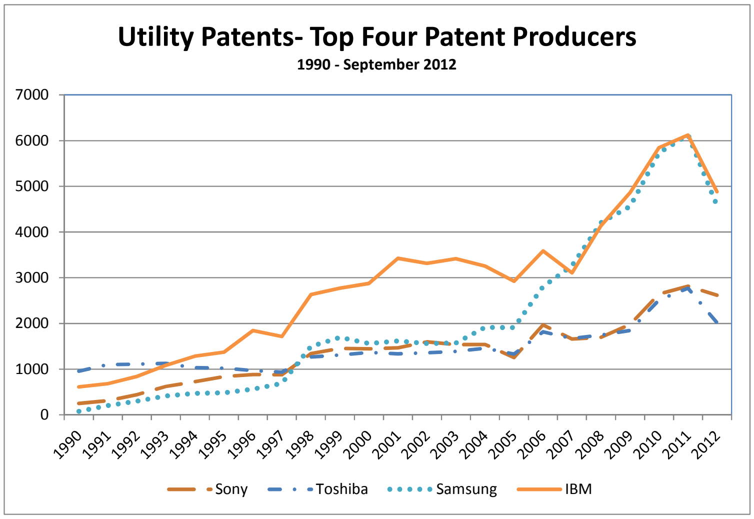

Figure 9: Time Distribution of Top Four Patent Producers

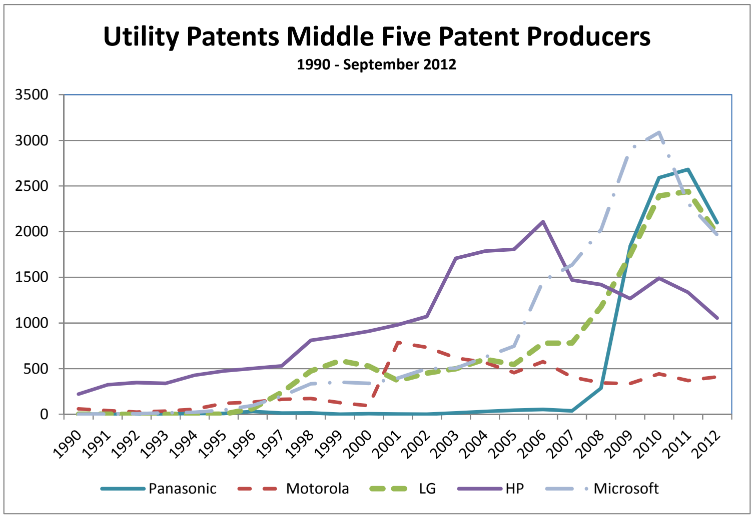

Figure 10: Time Distribution of Middle Five Patent Producers

Of note in **Figure 10** are the relatively late entry (2007) of Panasonic and the significant jump in patents granted to Microsoft, Panasonic, and LG starting in 2006-2007.

Since these three companies have a prominent position in the technologies of interest in this paper, further investigation and analysis will help determine the relative significance of this event.

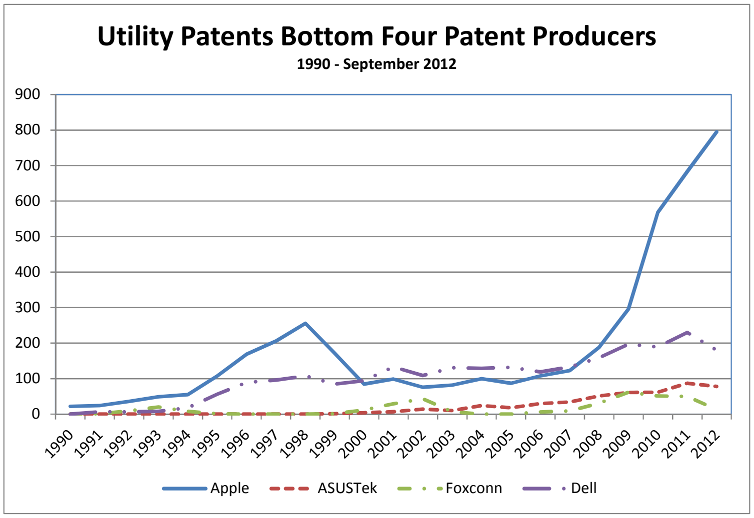

Figure 11: Time Distribution of Bottom Four Patent Producers

Of note in **Figure 11** is the significant increase in patents granted for Apple starting in 2007. The financial fortunes of Apple were sharply lower in the 2001-2004 period followed by double digit profitability from 2005-2007 and an almost doubling of profits from 2008 – present. Clearly Apple was in a better financial position to conduct research after 2004. Additionally, the resurgence of Apple marketing leading to the explosion of iPhone® in 2007 are strongly correlated to the number of patents granted.

With the perspective gained from analyzing the patent time distribution of our companies of interest, we now turn to the technologies of interest. A technology expert from either thrust listed in **Figure 9** could readily create a list of closely related terms. We used the Concept Linking feature in SAS Enterprise Miner and Text Miner to look for close association of terms. Some of the results are shown in **Figure 9**. Next, we created SAS code that allowed the efficient search of our patent holdings. We found some interesting results. In the below paragraphs, we show the more noteworthy ones.

Video display technologies are used in a variety of consumer electronic devices as well as in commercial and military applications.

Our research identified a number of technical research areas relating to large displays (movie theaters), computer displays (desktops and laptops), and handheld devices. We were interested in which companies seemed to have been the most active in these technologies as evidenced in patents granted.

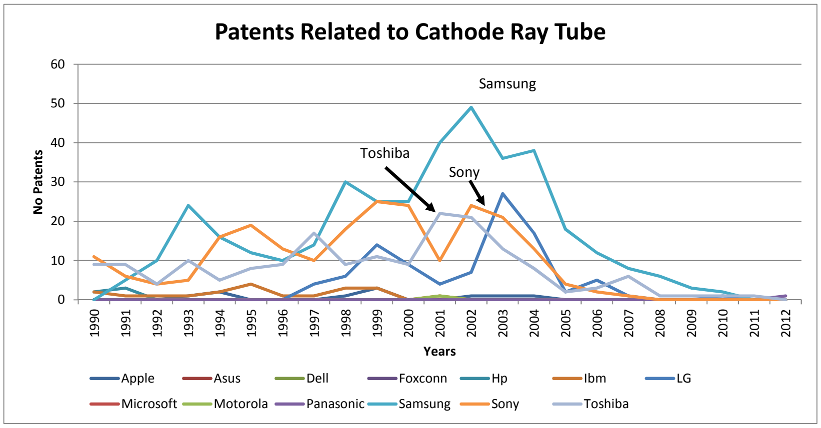

Figure 12: Patent Time Distribution for Cathode Ray Tube

One of the earliest technologies in this area is the cathode ray tube – widely used in televisions and oscilloscopes. **Figure 12** shows decades-long research efforts for many of our selected companies and Samsung's prominence in the early 2000s. As other competing technologies became favored, the number of patents dropped significantly in the latter part of the 2000s.

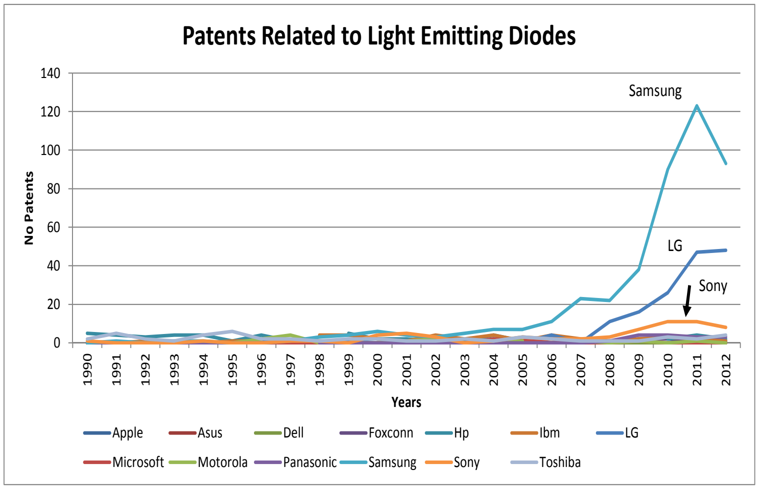

Figure 13: Patent Time Distribution for Light-Emitting Diodes

Another technology that comes to mind is the light-emitting-diode (LED), which has also been used in a variety of applications. As **Figure 13** shows, LED patents span decades but seem to have exploded in the mid 2000s with Samsung a strong contributor in the latter years.

Due to the September cut-off date of our data set, the sharp downward trends for Samsung shown in **Figure 13** is most likely misrepresented.

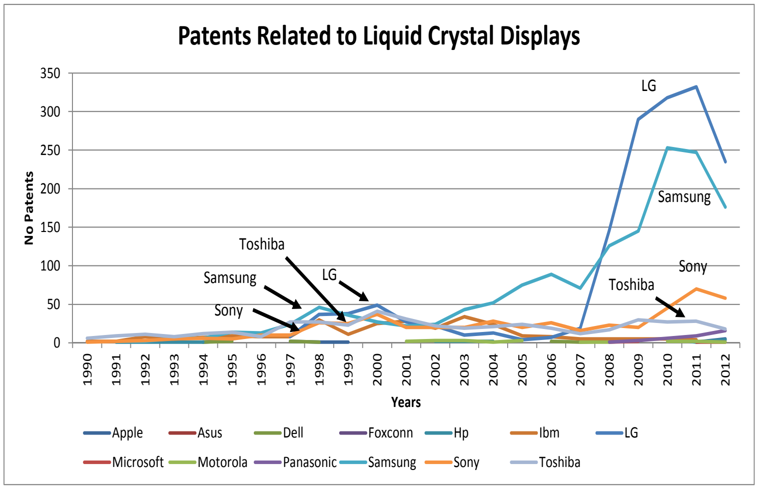

Figure 14: Patent Time Distribution for Liquid Crystal Displays

Liquid crystal displays (LCDs) have also become prominent in marketing for many display devices. Interestingly, **Figure 14** shows a peak in the late 1990s with strong representation by Toshiba, Samsung, LG and Sony. Another peak is indicated in the late 2000s, with LG and Samsung having a dominate role. The first peak can be associated with watches, while the second seems correlated to the explosion in mobile devices.

A related technology active matrix LCDs (AMLCDs) have seen a fairly low number of patents, with LG, Sony and Toshiba being the most prominent patent producers (see **Figure 15**).

AMLCDs have been a technology of choice for many notebook computer manufacturers, due to low weight, very good image quality, wide color gamut, and response time.

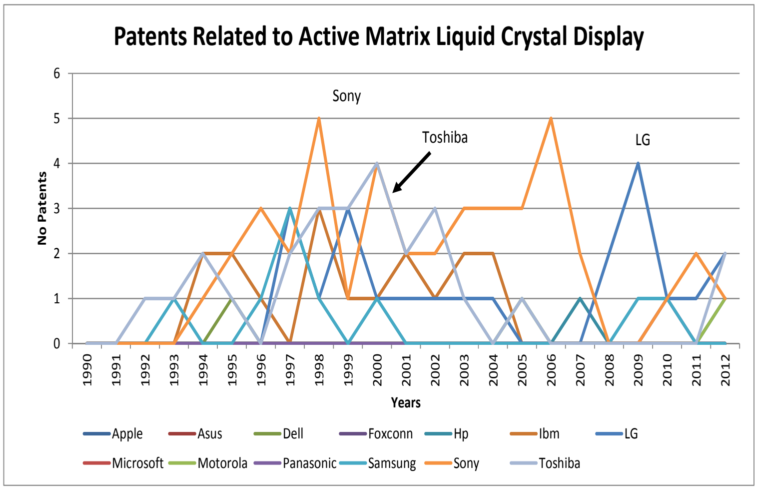

Figure 15: Patent Time Distribution for Active Matrix Liquid Crystal Displays

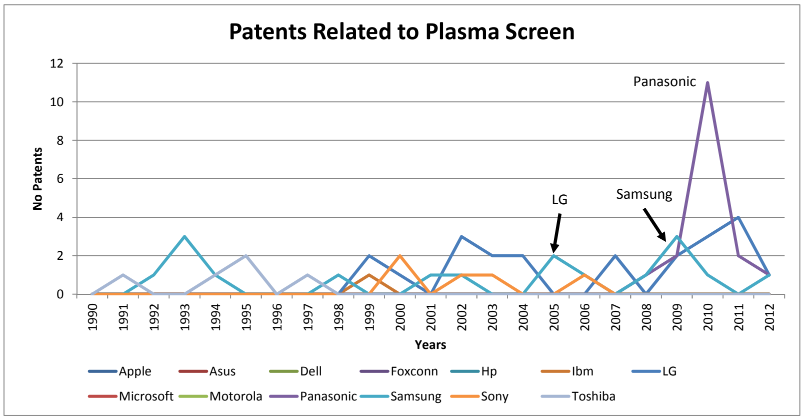

Figure 16 shows the patents related to plasma screens. Although there has been decades-long patent activity the later part of this decade seems to have had the most research. Leaders in this area are Panasonic, Samsung and Sony. This technology is widely used and marketed in high-end, flat panel high-definition televisions.

Figure 16: Patent Time Distribution for Plasma Screen

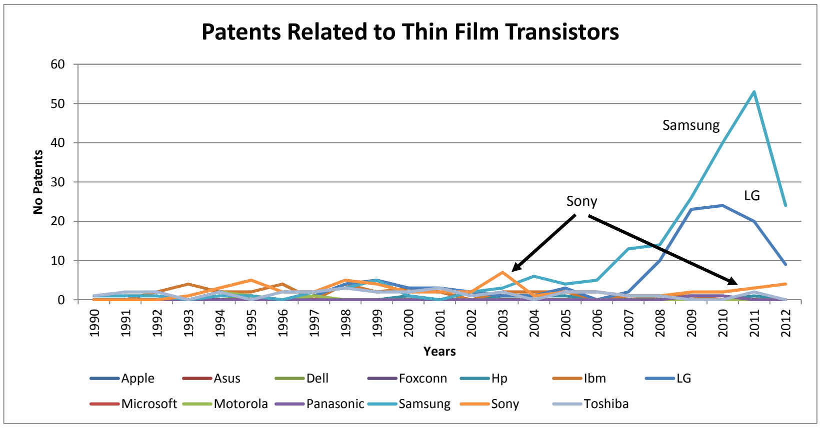

Figure 17: Patent Time Distribution for Thin Film Transistors

Figure 17 shows patent activity for thin film transistors (TFTs) mostly during the late 2000s. Samsung, LG and Sony are the most active in this research technology.

Many color LCD televisions and monitors use this technology.

Figure 18 shows patent activity for digital light processing (DLP). DLP is used in many commercial theaters and has had fairly low but consistent patent activity over the studied period.

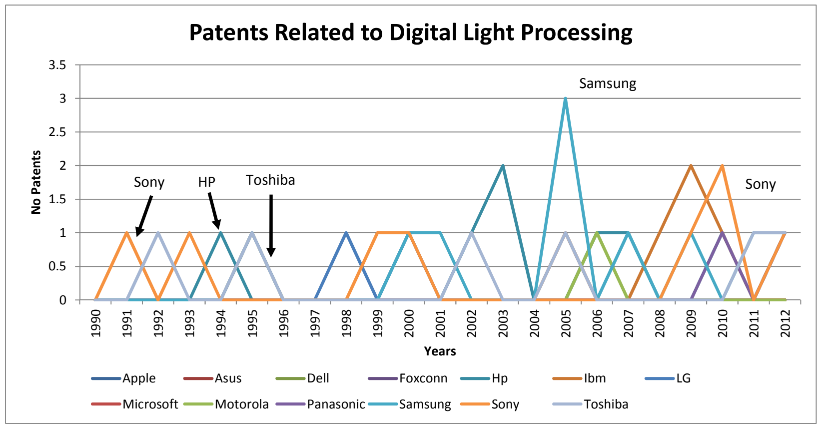

Figure 18: Patent Time Distribution or Digital Light Processing

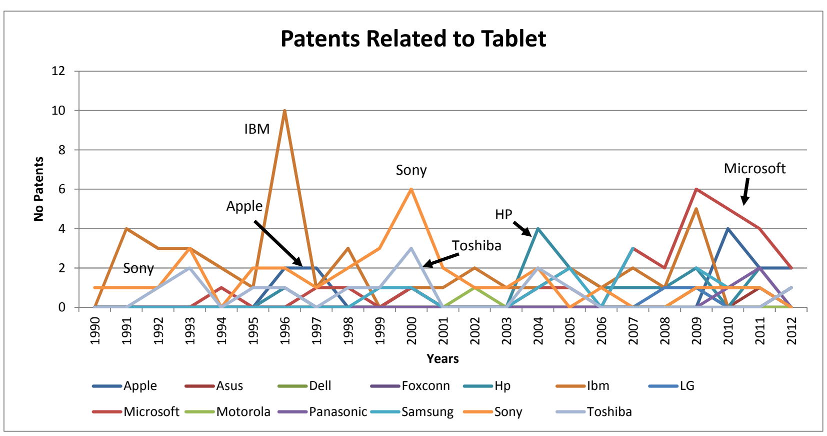

Figure 19: Patent Time Distribution for Tablets

The last two figures (**Figures 19 and 20**) show patent activity related to mobile devices and tablets. Mobile devices seem to have the most patent activity in the 2000-2012 period, with Motorola, LG, Sony, Panasonic and Samsung having large numbers of patents.

Contrasting patents related to mobile devices with those related to tablets shows a remarkably different distribution. Patents for tables span the entire time spectrum, with a larger variety of participants indicated.

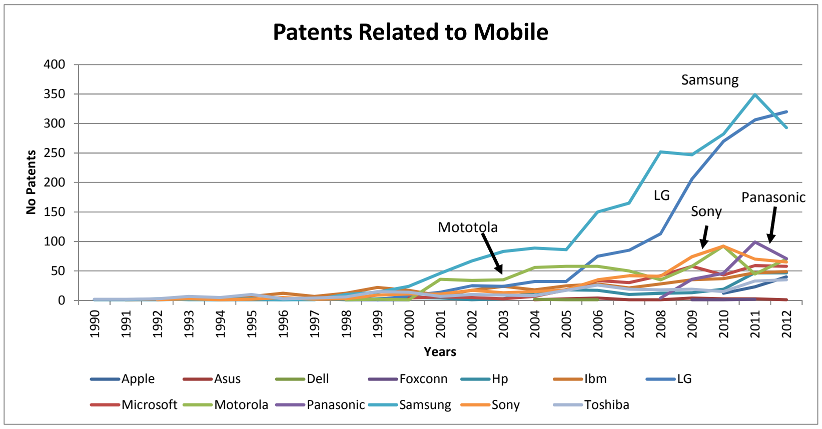

Figure 20: Patent Time Distribution for Mobile

This research also resulted in many much less interesting findings that, for the sake of brevity, are not presented in this paper.

Summary

This paper assessed the vast U.S. Patent Office information repository and an analytical methodology that extracts patent information and produces interesting insights to research development testing & evaluation (RDT&E) trends for several major technology companies.

This research revealed interesting technology associations using SAS Enterprise Mine and Text Miner and expanded the utility of data mining by exposing unknown relationships and by assessing their prevalence in granted patents.

Ultimately, this research helped uncover interesting RDT&E investment trends in two major thrusts. The methodology can be quickly adapted to other technology areas of interest across all industries.

SAS Code

Portions of the SAS code used in this paper are provided at http://www.sascommunity.org/wiki/Data_Mining_of_US_Patents**.**

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Kenneth M Potter
Science Applications International Corporation (SAIC)
6723 Odyssey Dr.
Huntsville, AL 35806
tel: (256) 319-8440
E-mail: kenneth.m.potter@saic.com

Robert N Hatton
Science Applications International Corporation (SAIC)
6723 Odyssey Dr.
Huntsville, AL 35806
tel: (256) 319-8403
E-mail: robert.n.hatton@saic.com