Cosmonautics Patent Analytics Using Cooperative Patent Classification Hierarchy

Rishi Raghav[1], Seshadri Ragavachari[#] and Hari Ravi[#]

[1]West Windsor Plainsboro High School North, Plainsboro Township, NJ, USA
[#]Advisor

ABSTRACT

Patent research is a sector of entrepreneurship that has long been used to understand both technological and financial trends in STEM industries. Such research provides insight into intellectual capital as it relates to entities, countries, and innovation areas, thus being an important factor in research and development decision-making. Yet, the sheer volume of patents filed in conjunction with unorganized attributes within patent databases creates a challenge for effective patent analysis. With this in mind, we create a patent analytic tool that organizes a specific patent class using CPC, applicant, and time attributes.

Introduction

The process of creating a patent is an expensive and time-consuming process. Over the last century, over 16 million patents have been published across more than 100 countries. The accumulation of this vast trove of patents has created a need to search, slice & dice, and analyze these publicly accessible documents. One of the goals for such an exercise is to understand trends and identify patterns in emerging inventions. The current patent documents outlined by various government entities have many common attributes that can be used for this form of analysis. An example of such an attribute is the Cooperative Patent Classification (CPC), which categorizes the data by research subject area at multiple levels of hierarchy (Armour-Gemmen, 2017).

Data aggregation around specific attributes of patent documents is the single most important element to create the tools needed for the patent analysis. In order to produce an effective patent analytics tool, we first identify specific attributes across large volumes of patents that have the potential to be aggregated and reveal useful engineering trends. A primary issue that arises in current patent data is the inefficient organization of patent attributes such as CPC, applicant, and publication event, thus making it highly difficult for patent trends in research areas of high demand to be sufficiently analyzed. With regards to CPCs, patents are frequently associated with a group of research areas rather than being paired with research areas individually (Berry, 2020). As a result of the inability to distinguish between CPCs for a particular patent, analysis becomes increasingly ambiguous and ineffective. Moreover, CPC descriptions are often expressed as a series of clauses that lack the level of clarity necessary to fully understand the research area associated with a specific patent. Just as in the case of CPCs, multiple applicants are often associated with the same patent record, making it challenging to slice and dice by applicant (Lerner & Seru, 2021). An additional issue reflected by commonly used patent analytic tools is the association of patents with multiple publication dates, which causes ambiguities when analyzing patent filing trends with respect to time. We aim to identify a solution by normalizing attribute data at an atomic level such that it can be transformed into an analytics-friendly dashboard. To maximize the effectiveness and feasibility of this analytics platform, we must narrow our subject area to a specific research industry. For our pilot, we select the cosmonautics research area, which encompasses the engineering technology surrounding the design, launch, and control of spacecrafts. This area of research is particularly advantageous as a pilot given that it consists of a broad time frame, unfragmented geographical distribution, and large patent data sets that are readily available.

Concision and Hierarchy Categorization of CPC Master Data

The CPC descriptions contained within the CPC Master data are often a series of clauses. Although such a CPC description is ideal for human interpretation, it is not useful for aggregation and analytics, as indicated in *Table 1.1*. Conversely, a concise CPC description will provide a clean framework to analyze and present data, as shown in *Table 1.2*.

Table 1.1 (Source: US Patent Office): The table below shows the CPC master data compiled by EPO and USPTO entities. The CPC descriptions are less concise and frequently span multiple sentences. This format makes the use of CPC description challenging for data analytics. A further challenge illustrated by *Table 1.1* is that the layout of the CPC hierarchy is not formatted in a way that analytics programs can readily decipher and use to produce graphical reports.

CPC Code	Level	Description
B64G1/24	2	Guiding or controlling apparatus, e.g. for attitude control (jet-propulsion plants F02K; navigation or navigational instruments, see the relevant subclass, e.g. G01C; automatic pilots G05D1/00)
B64G1/36	3	using sensors, e.g. sun-sensors, horizon sensors
B64G1/38	3	damping of oscillations, e.g. nutation dampers
B64G1/32	3	using earth`s magnetic field

Table 1.2: The table below indicates the modified CPC master data that transforms the CPC Description data to concise definitions. Concise definitions can be easily represented in X-Y axis dimensional categorization in analytical reporting tools. In addition, they can be aligned with industrial terms that are quickly understood by a broad spectrum of users in various research areas. For example, "Guiding and Controlling" is a term that can align with both a patent law firm researching trends to evaluate patent IP capital and an educational institution that is analyzing cosmonautics research areas. Further, *Table 1.2* aims to flatten the CPC hierarchy into a single row instead of the multi-row indented representation shown in *Table 1.1*, thus providing an additional advantage for the patent analytics process.

CPC Code	CPC Description	CPC Parent Code	CPC Parent Description	Subject Code	Subject Description
B64G1/24	Guiding or Controlling	B64G1/24	Guiding or Controlling	B64G1	Cosmonautics
B64G1/363	Using Sun sensors	B64G1/24	Guiding or Controlling	B64G1	Cosmonautics
B64G1/38	Damping of Oscillations	B64G1/24	Guiding or Controlling	B64G1	Cosmonautics
B64G1/32	Using Magnetic Field	B64G1/24	Guiding or Controlling	B64G1	Cosmonautics

Normalization of CPC and Applicant Designation in Patent Publications

Within current patent data, each patent publication is not split by atomic level of CPC as indicated in *Table 2.1*. This shortcoming makes data analytics by research area highly difficult.

Table 2.1 (Source: European Patent Office): The table below illustrates a "Spacecraft Servicing Devices" patent published under multiple CPC codes. All CPC codes are listed as a single field in a descriptive format. In addition, the same CPC code is often repeated under multiple countries in which the patent was applied. Patent data stored in this format does not lend itself to data analytics.

Title	Inventors	Applicants	Publication number	CPC	Publication date
SPACECRAFT SERVICING DEVICES AND RELATED ASSEMBLIES, SYSTEMS, AND METHODS	CHOW KENNETH SIU-KIN [US] GLOGOWSKI MICHAEL JOSEPH [US] HEKMAN BENJAMIN MICHAEL [US] LIEBERBAUM MARK [US] LLORENS WILLIAM A [US] NICHOLSON JAMES GARRET	NORTHROP GRUMMAN INNOVATION SYSTEMS INC [US]	US2020223568A1	B64G1/1078 (EP) B64G1/242 (EP) B64G1/244 (EP) B64G1/26 (EP) B64G1/64 (EP) B64G1/646 (EP,US) B64G1/242 (US) B64G1/402 (EP) B64G4/00 (EP)	2020-07-16

An important step in our research undertaking was the normalization of patent databases to split the patent publication record by each CPC code. As illustrated in *Table 2.2*, when the data in *Table 2.1* was transformed, the patent publication for "Spacecraft Servicing Device" by Northrop Grumman is split into 8 atomic CPC level records that can be aggregated along with other patents having similar CPC categorization.

Table 2.2: The table below reflects the 8 atomic CPC level records for the "Spacecraft Servicing Device" by Northrop Grumman. Breaking a single patent record into distinct CPC records provides the ability to analyze patent data by engineering subject area.

CPC Code	Publication Number	Earliest Publication
B64G1/1078	US2020223568A1	2020-07-16
B64G1/244	US2020223568A1	2020-07-16
B64G1/64	US2020223568A1	2020-07-16
B64G1/646	US2020223568A1	2020-07-16
B64G1/242	US2020223568A1	2020-07-16
B64G1/26	US2020223568A1	2020-07-16
B64G1/402	US2020223568A1	2020-07-16

Combining Concision and Normalization for Analytics

By combining the concise and categorized CPC master data with normalized patent publication data as shown in *Table 2.3*, we aim to form a patent publication structure that readily lends itself to analytical processing, thus creating opportunities for researchers in different industries to perform trend analysis.

Table 2.3: The table below illustrates the result of forming a relation between the categorized CPC master data illustrated in *Table 2.1* and normalized Patent publication illustrated in *Table 2.2*. The result is a patent data set that is highly categorized by CPC hierarchy and that can be grouped and analyzed though graphical tools.

Subject Code	Subject Description	CPC Parent Code	CPC Parent Description	CPC Code	CPC Description	Publication Number	Earliest Publication
B64G1	Cosmonautics	B64G1/10	Artificial Satellites	B64G1/1078	Maintenance Satellites	US2020223568A1	2020-07-16
B64G1	Cosmonautics	B64G1/244	Attitude Control	B64G1/244	Attitude Control	US2020223568A1	2020-07-16
B64G1	Cosmonautics	B64G1/64	Coupling or Separating Systems	B64G1/64	Coupling or Separating Systems	US2020223568A1	2020-07-16
B64G1	Cosmonautics	B64G1/64	Coupling or Separating Systems	B64G1/646	Docking or Rendez-vous Systems	US2020223568A1	2020-07-16
B64G1	Cosmonautics	B64G1/24	Guiding or Controlling	B64G1/242	Orbits and Trajectories	US2020223568A1	2020-07-16
B64G1	Cosmonautics	B64G1/24	Guiding or Controlling	B64G1/26	Using jets	US2020223568A1	2020-07-16
B64G1	Cosmonautics	B64G1/40	Propulsion Systems	B64G1/402	Propellant Tanks	US2020223568A1	2020-07-16

Process of Data Extraction, Transformation, and Analysis

Data Extraction

In our research, we use the cosmonautics patent data set as a pilot to illustrate the transformation steps applied to raw publication data in order to create analytics-friendly data sets. The first step we take is to use ESPACENET, an online patent searching tool, to isolate cosmonautic patents for analysis. ESPACENET is advantageous as it consists of a classification search feature that enables users to search patents based on their CPC. In our case, the CPC for cosmonautics is denoted as "B64G1." A search for patents corresponding to the CPC "B64G1" returns 14,658 patents, as of July 2021, as shown in *Figure 3.1*. In addition to the cosmonautics-specific parent CPC, each of the patents consists of an extension to their CPC that defines the subcategory in which they were filed.
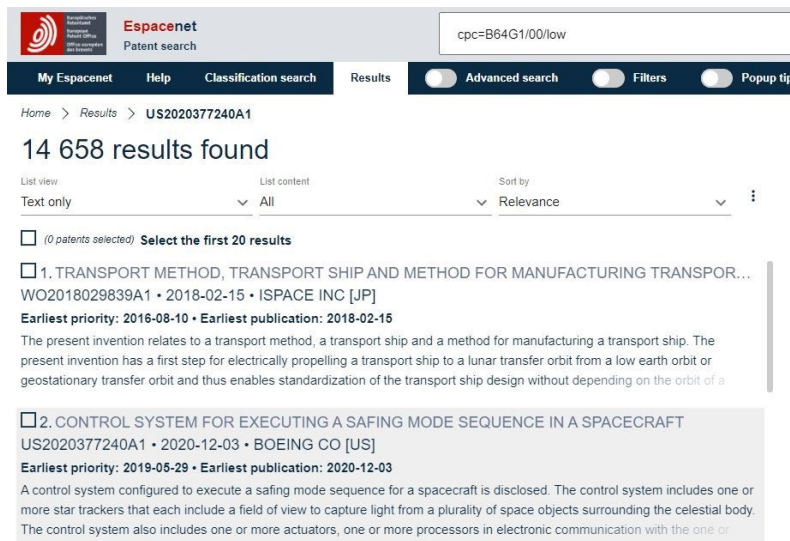

Figure 3.1: The image above reflects that a classification search for cosmonautics patents (B64G1) returns 14,658 patent publications.

Data Manipulation

We export the 14,649 cosmonautics patents into an Excel format and import the patent data into Oracle database, which was used in our research project, following which we create a raw data table for all cosmonautics patents. The raw data table consists of several patent attributes, including patent title, inventors, applicants, publication number, CPC, publication date, earliest publication, and family number. Once the cosmonautics patent data is imported into Oracle database, we use SQL String manipulation functions such as SUBSTR & INSTR to normalize the CPC and applicant attributes. In this process, we transform the data format shown in *Table 2.1* to the data format in *Table 2.2*. Further, we download country code mapping from the USPTO patent office, which is later used for a country-based analysis of patent filings.

Data Concision

Following the normalization of the CPC and applicant attributes of the cosmonautics patents, we must proceed to the concision of the CPC description attribute. We initially download the alphabetical CPC master data from USPTO,

which associates all distinct CPCs with their respective descriptions. This raw data is then uploaded into the Oracle database, where we filter and extract the cosmonautics CPC master data into an Excel document. Within Excel, we establish concise CPC descriptions that are analytics friendly by categorizing each CPC into a higher-level research area for analytics aggregation. As a result, we transform the CPC description format shown in *Table 1.1* to the CPC description format shown in *Table 1.2*. The revised CPC descriptions are then uploaded into Oracle database.

Data Finalization

In order to perform analysis, we use SQL to join the patent data indicated by *Table 2.2* with the CPC master data shown in *Table 1.2* to create the analytical data set shown in *Table 2.3*. We export the analytical data set from Oracle into Excel tables that focus on three distinct patent analytics areas: patent publication by CPC, patent publication by applicant, and patent publication by main subject area. Additionally, all three data sets contain the country and publication date attributes.

Results

We select Zoho analytics as the tool to carry out our data analysis due to its comprehensive data representations, interactive dashboard structure, and predictive analytics. Initially, we upload each dataset to Zoho analytics in the Excel file format. We then establish five distinct patent areas that will make up the cosmonautics analytics dashboard: cosmonautics summary, CPC category summary, CPC deep dive, applicant deep dive, and applicant timeline. As shown by *Figure 4.1*, we can generate reports within each of these areas that visually analyze the correlation between two patent attributes in order to reveal both categorical and time-based trends. Moreover, the dashboard feature of the analytics tool allows us to include widgets that summarize important conclusions from the data report. Such a structure within the cosmonautics dashboard enables users to understand patent filing trends at a broader level while still having the option to drill down the dimensional attributes of the data.
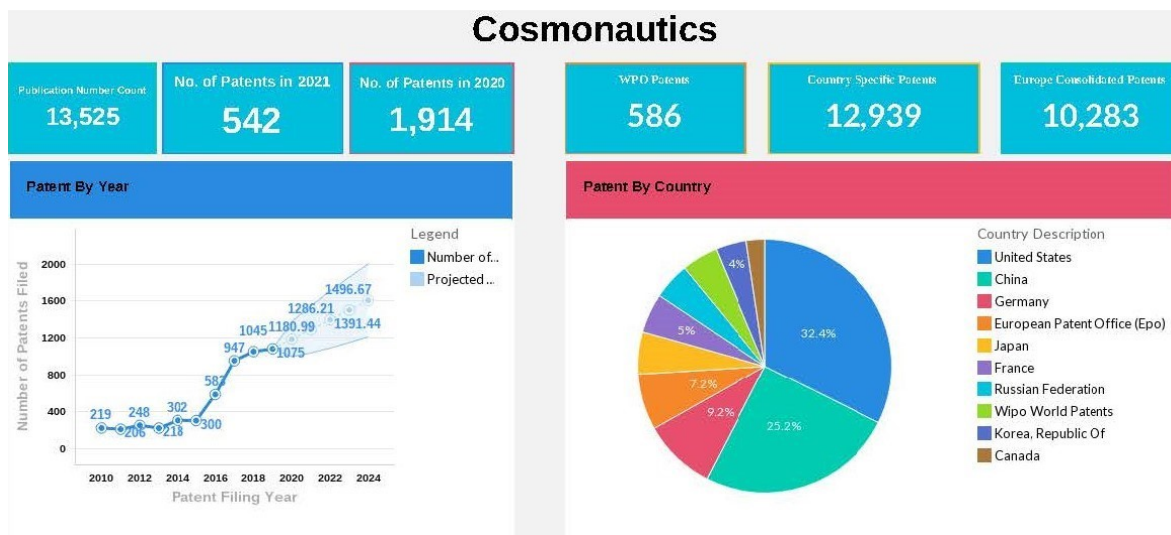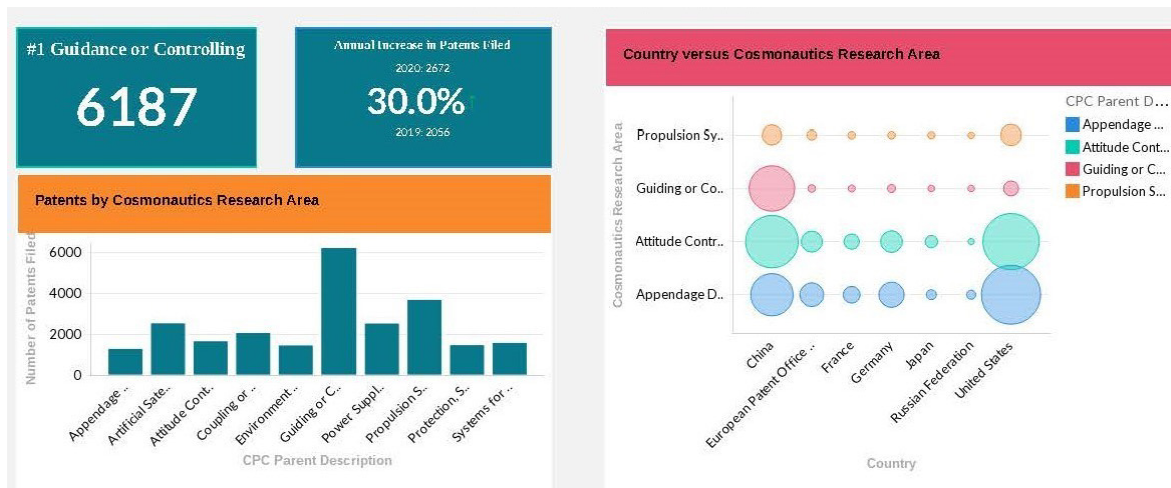
Figure 4.1: The figure above reflects the analytical graphical reports that we generate in Zoho analytics using the transformed cosmonautics patent data. The graphical results show that the CPC and applicant normalization provides insights into patent filing trends across filing years, countries, and research areas. In comparison to the raw cosmonautics patent data, the cosmonautics analytics dashboard provides the greater ability to recognize meaningful trends in various engineering research areas. These trends in turn lend themselves as tools for more appropriate research and commercial investments in the cosmonautics area (Rodriguez & Palacios, 2014).

Conclusion

Through the cosmonautics patent analysis dashboard, we create a pilot of a patent analysis framework that makes patent data significantly more comprehensible, thus providing broader insights of trends in research areas. While our pilot focuses on the cosmonautics research area of the aerospace industry, it can be expanded to other research sectors within the aerospace industry and ultimately scaled to other industries. Through this process, a common analytics platform can be established for patents across multiple industries, thus providing the opportunity for wide usage. As we apply our patent analysis system to various industries, we can use more advanced methods such as machine learning to not only accelerate the scaling process, but also to increase the scope of our analysis. Hence, the patent analytics system will continuously evolve as technology advances and industries develop overtime.

Limitations

Although our research process creates a concise and categorized CPC hierarchy that yields beneficial results, the feasibility of this process is mediated by a smaller dataset for the cosmonautics research area. The CPC master data across all industries contains over 260 K records. Therefore, the manual concision and creation of a hierarchy for such a large CPC dataset would be time consuming and more prone to inaccuracies. This limiting factor could create a barrier to scaling the patent analytics method to other areas. Another limitation is the high level of manual intervention needed to extract and transform the patent publication data from online data bases. This process may limit our ability to include newly filed patents into the analytics tool in real time.

Acknowledgments

References

Armour-Gemmen, M. G. (2017). From Creativity to Classification: A Logical Approach to Patent Searching. The Research Repository @ WVU. researchrepository.wvu.edu/faculty_publications/1133/

Berry, T. (2020, December). The Effects of a Patent Application's Technology on the Patent Prosecution Process. Johns Hopkins University. https://jscholarship.library.jhu.edu/bitstream/handle/1774.2/63601/Berry-Capstone-2020.pdf?sequence=1&isAllowed=y.

Lerner, J., & Seru, A. (2021). The Use and Misuse of Patent Data: Issues for Corporate Finance and Beyond. Harvard Business School. https://www.hbs.edu/ris/Publication%20Files/18-042_c035be94-3f69-4a08-ab2f-5b77e7446df6.pdf

Rodriguez, M., & Palacios, A. (2014). Identifying Recent Developments In Knee Prostheses Through A Patent Analysis. citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1079.1558&rep=rep1&type=pdf

US Patent Office. (2021). Guide to the CPC (Cooperative Patent Classification). Cooperative Patent Classification. ARPN Journal of Engineering and Applied Science. https://www.cooperativepatentclassification.org/wcm/connect/cpc/212f75e9-e9d4-4446-ad7f-b8e943588d1b/Guide+to+the+CPC.pdf?MOD=AJPERES&CVID=.

US Patent Office. (2016). Seven Step Strategy. United States Patent and Trademark Office - An Agency of the Department of Commerce. https://www.uspto.gov/learning-and-resources/support-centers/patent-and-trademark-resource-centers-ptrc/resources/seven.

WIPO. (2020). World Intellectual Property Indicators 2020. Geneva: World Intellectual Property Organization. https://www.wipo.int/edocs/pubdocs/en/wipo_pub_941_2020.pdf