

Long Island University Digital Commons @ LIU

Undergraduate Honors College Theses

LIU Post

2016

The Application of Big Data Analytics to Patent Litigation

Chloé Margulis

Long Island University, musikhan15@gmail.com

Follow this and additional works at: http://digitalcommons.liu.edu/post_honors_theses



Part of the [Computer Sciences Commons](#)

Recommended Citation

Margulis, Chloé, "The Application of Big Data Analytics to Patent Litigation" (2016). *Undergraduate Honors College Theses*. 5.
http://digitalcommons.liu.edu/post_honors_theses/5

This Thesis is brought to you for free and open access by the LIU Post at Digital Commons @ LIU. It has been accepted for inclusion in Undergraduate Honors College Theses by an authorized administrator of Digital Commons @ LIU. For more information, please contact natalia.tomlin@liu.edu.

The Application of Big Data Analytics to Patent Litigation

An Honors College Thesis

By

Chloé Margulis

Fall, 2016

Computer Science Department

Dr. Brian Galli, Advisor

Dr. Stephanie White, Reader

Date

Abstract

This research defines the current gap between big data analytics and patent litigation. It discovers how big data analytics can be applied to the patent industry in order to create more effective risk analysis, an early warning system, and preventative strategies for inside and outside of the courtroom. Big data has the potential to modify current practices in the patent industry, namely geared towards aiding patent examiners, attorneys, inventors, jurors, and judges. It also offers a solution to the threat that patent monetizers pose on smaller companies and inventors, who often lose rights to their patents and other assets in these sometimes unavoidable lawsuits. This research examines the application of big data in the healthcare industry for real-time results and preventative measures. These actions set a good precedent for further diffusion into other industries, specifically patent legal. Features for future implementation and project development are presented as a roadmap to create a universal big data analytics system for the patent industry. Finally, this research will touch upon a case study of *Apple v. Samsung* and identify how the case might have yielded different results in the event that big data analytics had been applied to the legal proceedings.

Table of Contents

1	Introduction	4
2	Literature Review	6
2.1	What is Big Data?	6
2.2	What is Patent Law?.....	8
2.3	What is Apple v. Samsung?	11
2.4	What Software is used for Data Analytics	12
2.5	Programming Languages that work for Data Analytics.....	15
2.6	How Analytics are implemented in Healthcare.....	18
2.7	How Has the Healthcare Industry Used Big Data?	20
2.8	The Application of Big Data to Healthcare R&D.....	25
2.9	How Big Data is currently used in Litigation	27
2.10	How Patent Litigation currently uses Big Data.....	30
2.11	Gap Analysis	31
2.12	Literature Review Conclusion.....	33
3	Analysis	35
3.1	Existing Gaps in Big Data’s Application to Litigation	35
3.2	The Relationship between Big Data, Healthcare, and Litigation	40
3.3	Apple v. Samsung: The Decision and its Applicability in the U.S., Japan, and China.....	43
3.4	The Beneficial Features of a Big Data Risk Analysis	46
3.5	Thoughts on the Outcome of Apple v. Samsung	47
3.6	The Reasons Reform is Necessary in the Patent Industry.....	49
3.7	The Added Problem of Patent Monetization in Smartphone Warfare	53
3.8	Apple v. Samsung’s Effects on Current Calls for Change.....	59
3.9	How Big Data Analytics Might Have Changed the Outcome of Apple v. Samsung.....	62
3.10	A Conclusion on Healthcare’s Big Data Application and It’s Correlation to The Patent Industry.....	64
3.11	A Suggestion for Big Data Analytics Implementation.....	66
3.12	Big Data’s Connection between Healthcare and Patent Litigation, and Apple v. Samsung Revisited.....	76

4	Conclusion.....	80
4.1	Research Contribution.....	80
4.2	Limitations and General Barriers for Implementation	83
4.3	Suggestions for Future Research.....	85
4.4	Final Summary	87
	Bibliography	90
	Additional Sources.....	95

1 Introduction

This research intends to create a bridge between big data analytics and the patent legal industry. The industry can reap significant benefits from a suggested implementation of big data. This is demonstrated with examples of big data usage in the healthcare industry. Combined with a case study of *Apple v. Samsung*, we will analyze how decisions in the courtroom can be enhanced with big data's predictive capabilities. Finally, we suggest a higher-level table of required features for future implementation.

For several years, big data has been a strategic tool in healthcare with the intention to reduce costs and save lives. Recently, these same tools have infiltrated the legal industry, primarily to differentiate competitive rates and judicial bias. Big data should, and can, be applied more effectively to the industry, specifically pertaining to intellectual property.

Benefits similar to those in healthcare can be mirrored in the patent legal industry. This is possible because the industries have two things in common: required proactive planning and real-time access to crucial information. Since big data facilitates healthcare to accomplish these tasks, the same can be achieved in the patent industry. Effective application can provide users in the patent industry with predictive analytics, risk management, and an early warning system that can yield more preventative litigation strategies.

A big data service will offer assistance to a variety of patent practitioners. With such a service, US Patent Office examiners can determine the validity of patents during the application process. If an application overlaps with an existing patent's credentials, then the application can be denied. As a direct result, future legal and licensing obstacles may be eliminated. Patent holders and inventors can use the service to determine whether or not it is beneficial to pursue patenting ideas. Furthermore, big data will present the best ways they can protect their ideas from patent

monetizers and other plaintiffs. A lawyer would use big data to strategize in the event that his client is sued for patent infringement. Finally, judges and jurors can use the service to get a better understanding of the circumstances that a lawsuit is brought against a patent holder. An educated, non-biased jury, for example, has the potential to change the outcome of a case. By analyzing *Apple v. Samsung*, this study highlights how beneficial big data analytics can be to the above listed patent practitioners.

One of the primary purposes of this research is to unearth big data analytics ability to cut litigation costs, create effective strategies, cases, and witness testimonies, and predict case outcomes. This research will also study and apply the current technology used for implementing big data services in healthcare to the patent industry. A brief analysis of services, such as Hadoop, and programming languages, such as Python and R, will help formulate the ideal big data analytics platform and its required features.

The Literature Review will present existing research regarding big data, healthcare, and the patent industry. It will provide basic definitions for key concepts and components of this research. The chapter will conclude by demonstrating the gap that exists between big data and the patent industry. The Analysis will discuss a bridge over the portrayed gap. It will further analyze the case *Apple v. Samsung* and present an alternative to the final verdict based on the application of big data analytics. Both chapters highlight the effects of big data analytics on the healthcare industry in order to prove its effectiveness and cross-applicability. The Conclusion presents the contribution this research has to the legal and technology fields, the limitations and general barriers for implementing the ideal big data analytics system, suggestions for future research, and a final summary of our findings.

2 Literature Review

2.1 What is Big Data?

Businesses increasingly turn towards big data because of its strength in predictive analytics and high return on investments. In simple terms, big data encompasses a massive amount of information in different data formats. Analytics make data point correlations visible when, at first glance, they may appear nonexistent (EMC, 2015, p.4). Big data also has a high presence in internet activity. Companies, such as Amazon, benefit from collecting this type of information to make predictions and trend analysis that further organizational goals and maintain consumer satisfaction. Because of big data analytics, many businesses can make suggestions about what to buy, what websites to visit, and what social media accounts to follow.

A 2011 McKinsey Global Report officially defines big data as “data whose scale, distribution, diversity, and/or timeliness require the use of new technical architectures and analytics to enable insights that unlock new sources of business value” (EMC, 2015, p.5). McKinsey suggests that while big data can accomplish many complex functions for its users, it is just as crucial that big data optimization systems are created and maintained. Otherwise, minimal value will be gained. For businesses to keep up with the technologically changing global market, it is imperative that they update and expand their resources.

Storing information collected over the internet requires a business to have an abundance of disk space. Solid State Hard Drives that accommodate these needs are expensive, so businesses turn to cloud storage. Hypothetically, cloud storage does not run out of space, making the continuous acquisition of mass information easier and less expensive. The only disadvantage of cloud storage is security. The user cannot customize or enhance the security system that comes

with the cloud, making sensitive information more vulnerable to potential fraud. Companies with more funds can afford developing custom clouds with enhanced security features. Amazon and Google are two companies that maintain their reputation by securing user information and user confidentiality. This is achieved by hiring experts who develop customized cloud storage systems with built-in data encryption. These experts will regularly test the security wall to ensure that it cannot be breached by any type of hacker. If needed, they make necessary improvements to increase the cloud's level of security. This is an important aspect to consider when developing a big data system that will be on a cloud. When storing and accessing sensitive information, security and data encryption should be top-level priorities.

Big data is composed of three characteristics: volume, variety, and velocity. Because of volume, mass information is accessible and useable. Facebook, for example, analyzes and stores all user activity in data warehouses. Variety implies that available data appears in different forms. Although a positive attribute, it guarantees a level of complexity, because businesses now experience an extra step in the data manipulation process—they must first convert the data into a single, readable type. Once the data is converted, it can be used for predictive analytics. Finally, velocity exemplifies the rapid speed at which new data can be created and for trends to change. This attribute presents a challenge, but can be overcome with support, and possibly machine learning, that accommodates to the changing environment.

These attributes make big data more difficult to analyze than other data types. Because of its distinctiveness, “it [was] predicted the market of big data technology and services will reach \$16.9 billion in 2015, up from \$3.2 billion in 2010” (Vanarse, 2014, p.4). This prediction is affirmed by big data's 40% annual growth rate, which is 7 times the growth rate of general

information and communication technology (Vanarse, 2014, p.5). Big data acquisition is a developing field that many industries can tap into in order to further organizational goals.

Recent studies demonstrate a similar trend in big data growth. According to the International Data Corporation, in November 2015, “big data technology and services market [will grow] at a compound annual growth rate (CAGR) of 23.1% over the 2014-2019 forecast period with annual spending reaching \$48.6 billion in 2019” (IDC, 2015). The three V’s and these statistics are some reasons businesses are increasingly attracted to invest in big data. Velocity, variety, and volume are advantageous attributes, but they are hard to capitalize on unless properly addressed in an up-to-date big data system. Therefore, the reports and information that is procured by the three V’s reveal the most benefits to a business.

2.2 What is Patent Law?

In January 1975, the United States Patent and Trademark Office (USPTO) was established to oversee the process of granting utility and design patents. Several requirements must be met before a patent can be accepted. First, the idea must be patentable subject matter with utility, novelty, nonobviousness, and enablement (USPTO, 2014). Patentable subject matter is broadly defined by past court decisions. In 1980, the Supreme Court decided in *Diamond v. Chakrabarty* that “Congress intended patentable subject matter to include anything under the sun that is made by man” (Cornell University Law School, 2015). This case set the precedent for future patentable subject matter to include anything except for abstract ideas, natural laws, and physical phenomena.

The utility requirement ensures that the invention or idea is useful and applicable to the real world. Someone with ordinary skill, or lack of knowledge about the subject matter, should understand how to use the invention without difficulty. In the USPTO’s Utility Guidelines, “Utility

must be specific to the subject matter claimed; not a general utility that [can] apply to a broad class of inventions” (Cornell University Law School, 2015). This presents a challenge to mobile phone developers, such as Apple and Samsung. This will be discussed later in the research. Despite such outlined rules, these companies race to patent general modifications that are applicable to most cellphone development. This was a major issue in *Apple v. Samsung*. The primary reason companies do this is for monetary gain, with the intent to push competitors out of the market. Such actions result in costly litigation.

The novelty requirement implies that the invention is new and/or unknown to the country that it is being filed in. This requirement is tested with a prior art comparison, in which the idea is supposed to be compared to existing patents and ideas, however this does present a major loophole (Cornell University Law School, 2015). In the past several years, there has been an influx in cellphone modification patents. The cellphone companies scramble to patent any improvements, therefore suggesting that they are novel. On the contrary, they are not, because every other company is patenting the same improvements. For example, both Apple and Samsung claimed rights to the slide-to-unlock feature on smartphones. As a result, lawsuits are filed left and right, to determine which improvement was patented first.

The fourth requirement is that the invention cannot be “obvious to a person having ordinary skill in the art at the time the invention was made” (Cornell University Law School, 2015). Usually this requirement is questioned in the courtroom *after* the patent is granted. An example is *Apple v. Samsung*, in which the presiding jury was technologically educated individuals with a bias towards one company’s product over the other. The composition of the jury raised many questions. First, was the jury composed of people with ordinary skill and understanding of the technology in

question? Second, can there ever be an unbiased jury in a case that relates to every-day-use technologies? These types of questions illuminate some of the implications in the patent process.

The final requirement is enablement, which requires a thoroughly written report in the patent application. The filer must present *all* details, including applicability, development, production, and maintenance. If it is not thorough enough, the USPTO has the right to deny the patent (Cornell University Law School, 2015). This requirement is weakly enforced, which is evidenced by the abundance of broadly-scoped patents. Many patents are mere skeletons comprising of only an idea and no substance to support the idea. This may be caused by the overflow of applications to the USPTO. It stresses the patent examiner to crunch in more application analysis in the same amount of time. Therefore, duplicate or overlapping patents can slip through the system, which presents a challenge and potential legal rift between patent holders.

The flaws presented above impact the legality of intellectual property. A patent holder can accuse and sue those he believes infringes on his patent, even for the vaguest reasons. If an inventor fears being sued, he can file a declaratory judgement to argue that the patent at issue is not invalid or that there is no reasonable justification for infringement (Cornell University Law School, 2015). Unfortunately, resolving these issues is not as easy as it may appear and will require a lot more time and money. These are only some of the ways in which patent law is effected by the flaws in the patent acceptance process.

For these reasons, we conclude that the patent system is flawed, beginning with the patent application process. A plan for improvements is necessary to prevent continuous legal debates, as seen between Apple and Samsung. Big data analytics presents itself as a solution to this dilemma, as will be discussed later in the research.

2.3 *What is Apple v. Samsung?*

Apple and Samsung continue to fight over their patents since 2011. Apple was the first to bring about legal charges by informing the public that Samsung “systematically copied Apple’s innovative technology and products, features, and designs, and deluged [the] markets with infringing devices in an effort to usurp market share from Apple” (Tibken, 2014). These bold claims were the beginning of what would become known as one of the most historic technology lawsuits of our time.

The companies argued about 5 Apple patents and 2 Samsung patents. Some of the debated features include universal searching, background syncing, quick links, slide-to-unlock, and automatic word correction (Tibken, 2014). One presumption is that Apple intentionally filed patents that were broad in scope in order to make patent overlap unavoidable. This goes against one of the primary guidelines for patent acceptance in the USPTO. Since Apple filed allegations first, it had more leverage in the courtroom. Even if the patented improvements were obvious and fundamental to keep afloat in the market, a company that implemented these would be held liable to Apple. Apple’s legal aggressiveness demonstrates its objective to maintain an industry clout. Apple achieving a monopoly, however, would not be beneficial, because it would reduce market competition. This leads to a lack of innovation along with slowed improvement and development processes. A competitive market, on the other hand, will force all companies to constantly release new products and improvements to uphold business and consumer satisfaction.

The 2012 decision for *Apple v. Samsung* dictated a win and loss for both parties. The Samsung Galaxy phones and the iPhone 4, 4s, and 5 were said to infringe on each other (Tibken, 2014). Samsung was found more at fault than Apple. As of December 2015, Samsung brought to

court a plea to receive \$548 million in settlement *back* from Apple. It argues that “while Apple may hold the cash for now, if patents are deemed invalid or it were to win a case on appeal, it should be allowed to retrieve at least part of its payment” (Reisinger, 2015). In other words, Samsung believes that it can prove in court that *Apple’s* patents are invalid. This could potentially reverse parts of the 2012 decision. Furthermore, recent research demonstrates that several of Samsung’s patents in fact did not infringe on Apple’s intellectual property.

Back in July 2015, Facebook, Google, and eBay joined forces with Samsung to appeal in court for the reimbursement from Apple. They argued, “if Apple’s victory is allowed to stand, the possibility of subsequent patent lawsuits utilizing the same intellectual property could negatively impact the development of useful modern technologies and have a devastating impact on companies” (Reisinger, 2015). Just as these allies suggest, continuous debate in court can hinder growth in the technology industry by limiting resources for innovation. It may be that the future of patent litigation and technology development depends on how *Apple v. Samsung* unfolds and shapes the legal playing field.

As long as companies conduct basic improvements on their smartphones and patent them, infringement wars will not cease. Patent monetizers, such as Apple, are ruthless when it comes to suing others for supposed violation. *Apple v. Samsung* is the major precedent case that does not set a good example, but acts as a fruitful study in this research.

2.4 What Software is used for Data Analytics

Big data is challenging to process and analyze because it is composed of structured and unstructured data. There are other platforms, such as a traditional relational database, that are used to analyze and manipulate data. This database, however, reaps value from structured, instead of

unstructured, data. The raw, unstructured, data is more challenging to work with because it lacks uniformity in format and type (EMC, 2015, p.8). Big data is known for being unstructured, which means that it requires more processes and algorithms for conversion and manipulation than a traditional database.

Before big data became popular, data warehouses were introduced in a system called MonetDB. These large databases require extensive querying that provides businesses with the intelligence and decision support they need. Healthcare still implements MonetDB because of its speediness in producing real-time results. MonetDB has a hierarchical memory system that decreases its task completion time more than traditional relational databases. Another reason MonetDB is still in use is because it is constantly modified to remain up to date. It offers additional modules that increase functionality and applicability. Unlike many other databases, MonetDB implements its own language, called MAL, and is compatible with C languages. This makes it easier for programmers to develop extensively customized modules (Idreos, Groffen, Nes, Manegold, Mullender, 2012, p.2).

There are some key ways traditional relational databases (RDBMS) differentiate from big data. Most simply, big data is not stored in database structures. Databases rely on tables, columns, and rows. The data is keyed in different entities, which can represent a person, thing, place, or concept. The entities are composed of attributes, which give value. The attributes are further broken down for tuples, which represent the value of the attributes in a specific instance of an entity. Relationships are then used to demonstrate correlations between the tuples in related tables, therefore creating structure.

Data warehouses are popular ways to store mass amounts of information. The Health Catalyst Late-Binding Data Warehouse is an example, which bridges the gap between big data and

traditional relational databases. A RDBMS relies on numerous system layers to conduct necessary tasks. The top layer is the application layer, which creates the user interface. This is the only layer that interacts with the user. This layer does contain SQL, which is used for querying and accessing data, however the user is not aware of this continuous action. In this layer, the user's selections and actions are translated to SQL.

The Health Catalyst Warehouse, on the other hand, puts as much as possible of the accessing, processing, and manipulating functions of a system into the application layer. Most other big data warehouses now do the same. Data transformation, business rules application, and semantically binding the data is fulfilled in one layer, instead of across many (Adamson, 2015). This makes the system easier and cheaper to maintain, contrary to a RDBMS structure, however it is similar to putting all your eggs in one basket—if there is an error, the entire system may be inaccessible until the issue is resolved.

Hadoop, on the other hand, is an example of a big data platform. It offers a de-structured filing system, which allows it to store thousands of data across nodes by using file directories. These directories are organized hierarchically. In contrast, a RDBMS stores the data in tables with assigned columns and rows (Adamson, 2015). With Hadoop, the data can be accessed across the nodes more quickly to create or modify relationships.

Hadoop is relatively inexpensive to use, which is a slight byproduct of it abandoning the RDBMS structure. Most relational databases use a Storage Area Network (SAN) configuration for the disk drives, which is more expensive than the Direct Attached Configuration (DAC) that Hadoop utilizes. In terms of conceptual design, big data requires less intricate administration because maintenance of complex tuple relations no longer exists. With big data, relationships are created on demand and are not stored in columns and rows (Adamson, 2015). The fact that big

data is simplified also makes solving storage redundancy issues easier on individual nodes than throughout entire databases.

Hadoop is popular because it is easy to use and is open source. This is another reason it is less expensive than other platforms. Companies with a tighter budget use it to avoid investing in extensive infrastructure. If there is a failure, fixing the issue is more feasible because there is an online community offering constant support. If a company using Hadoop has substantial hardware or access to the cloud, then it can store mass amounts of data. To put it into perspective, in 2011, each Hadoop cluster could store 200,000 terabytes of data (Adamson, 2015). Since then, it has grown significantly. It becomes the ultimate playground for firms that seek to expand their businesses without investing in expensive custom-built systems. Large social media outlets use Hadoop, benefiting from its unique ability to produce predictive trends about user interests and business objectives (EMC, 2015, p.24).

When choosing the optimal system to implement big data analytics, a firm should have an understanding of the existing analytics technology. In this case, patent firms must see if what currently exists can be applied to their needs. If they need more than what is available, then it is feasible to develop a central system for the patent domain that offers universal access to the information and analytics. In so doing, firms would be saved from the expense of creating and maintaining their own systems.

2.5 Programming Languages that work for Data Analytics

Choosing the best language to implement big data analytics can be challenging. This is because of big data's reliance on historical information. Predictive analytics of this type of data can be achieved with only certain languages. Two popular languages for this task include Python

and R. Unlike others, these languages offer learning libraries that make understanding how to code more straightforward (Adamson, 2015). Furthermore, when choosing the right language, it is important that “a data scientist be capable enough to run machine learning algorithms on a dataset to derive meaningful insights” (Jeevan, 2015). Some may argue this is better accomplished with Python, since Python makes it easier to incorporate machine learning algorithms into functions.

A look at figure 1 demonstrates that although R is more popular, many organizations are Python advocates. Professionals prefer Python because it has a strong foundation in online learning and educational sources that makes it easier to learn. Data scientists have a larger network to communicate with when creating individualized analytics systems. Python is also more flexible to write linear and logistic regression algorithms, which are vital components for success in big data analytics. These algorithms require more knowledge in machine learning. Machine learning pertains to the ability of computer programs to teach themselves to grow and change when dealing with new data. Other languages, such as Python, have instructions made by individuals that use words from our dictionary to represent functions. Fortunately, because of Python’s extensive libraries, data scientists can learn how to apply machine learning to Python algorithms (Nicolauo, 2014).

The learning libraries offer an open community for users to share ideas and questions and receive guidance. There is more interest to expand Python capabilities. This is one reason there is a trend to use Python for statistical analytics, and researchers note that “over the past two years, there’s been a noticeable shift away from R and towards Python” (Nicolauo, 2014). Python can help the patent industry fulfill its big data needs. Whether it be “wiring together network applications, programming for the web, scripting and automating data processing jobs, if you are looking for one programming language to do all these tasks, then Python is the answer” (Jeevan,

2015). Python has the capability to perform more than one function for its users. Because of its ease in applicability, it is a versatile language that should be used for the patent industry’s big data analytics system.

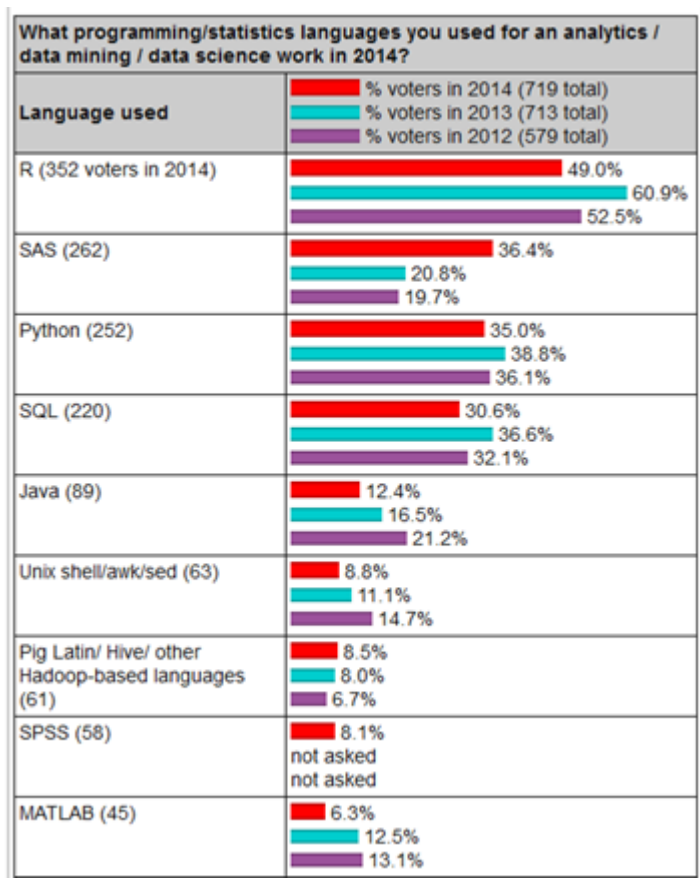


Figure 1: Jeevan, M. (2015, September 04)

There are some reasons programmers might prefer R over Python. Whereas Python can complete many different tasks, R is geared specifically to statistical analytics. Although programmers can still do this in Python, it requires more lines of code than with R. A few extra lines of code can translate into hours of additional programming and higher costs. Businesses may not want to take on these added costs. The statistical, research, and data science applications of R make it a better candidate for those seeking to develop big data systems from ground zero. It would

not come as a surprise if “increasing numbers of business analysts and other non-programmers arm themselves with the R language” (Asay, 2015). Several industries in the business world adopt R as their language, such as Google, Facebook, and Wall Street traders (Nicolauo, 2014). They can afford investing to build a big data system from the ground up. For an industry with a smaller budget, it is not as feasible to use R.

When relating big data to the patent industry, programming in Python is our recommendation. Most big data services that currently exist for litigation are written in Python, which makes the creation of additional systems easier and less expensive for the industry. The Analysis section of the paper studies this conclusion in greater depth.

2.6 How Analytics are implemented in Healthcare

The healthcare industry relies heavily on a traditional database structure because of a hesitancy to transfer to a new system. One example is MUMPS, a file structure and language that is specifically used for storing EHRs (Electronic Health Records). MUMPS stands out from most traditional database infrastructure because it combines an accessing and manipulating programming language with a structure compatible with big data. It also can scale in size to accommodate for mass amounts of information. Fortunately, MUMPS handles querying data quickly, which is not common for traditional database platforms, thus making it a successful predecessor for big data processing (Byrne, 2015).

MUMPS was advanced for its time because it “has its own collection of global arrays stored in nonvolatile memory...A MUMPS global sticks around on a server, accessible at any given time to a computer within the system by the addition of [one] character” (Byrne, 2015). MUMPS globals are variables stored in structured data files. They are automatically and transparently kept

on the server, even when a program or process reaches completion. Since users can directly access the database without having to go through abstraction layers that facilitate communication of instructions between the application and database, querying is faster. Adding and updating information becomes an easier process. This construction can grow, to develop more relationships, without reducing overall turnaround time when it comes to aggregation and predictive analytics. These are important components to consider when conducting real-time analytics.

The healthcare industry is reluctant to leave MUMPS after so many years of relying on it. Transitioning to a big data system requires a different skillset, which most “Hospital IT experts familiar with SQL programming languages and traditional relational databases aren’t prepared for [because of] complexities surrounding big data” (Adamson, 2015). Transferring to a new system will require increased training and education efforts. The industry may not be interested in taking on these added costs.

Security remains the most prominent issue. MUMPS was made specifically for healthcare, to be compliant with HIPAA regulations and patient confidentiality. If installing, configuring, and administering big data is not done carefully or correctly, there is an increased risk in privacy breaches. Already, big data presents a security challenge to every industry, but in healthcare, this challenge becomes a major roadblock for implementation (Raghupathi, 2014, p.6). Most big data platforms run on open source technology, which means data security is only as good as the security of the service provider. The healthcare industry cannot, and will not, take this risk. Therefore, many in the industry remain with the technology they already know, which guarantees security and HIPAA compliance. That platform is MUMPS.

To avoid privacy concerns, it is important that healthcare remains with the status quo or selectively chooses a vendor that will build a custom and secure platform. If healthcare is interested

in plunging deeper into the pool of big data, then the most viable candidate may be Cloudera. This service is geared towards increased security measures through “authentication, authorization, data protection, and auditing. Other commercial distributors are working to add more sophisticated security that will be well-suited for HIPAA compliance and other security requirements unique to the healthcare industry” (Adamson, 2015). Security is one of the most important aspects to healthcare when transferring to technologically advanced services. It is promising that other cloud based services are following in Cloudera’s footsteps and trying to implement more security measures in their platforms. This way, the healthcare industry can be certain that the service they use is HIPAA compliant. As a result, the threat of security breaches and loss of patient information will not be a factor that deters healthcare professionals from using more advanced big data analytics.

In the meantime, healthcare has merit in its reluctance to transfer to purely big data systems. The industry can accomplish predictive analytics with MUMPS, although the system’s capabilities are slightly more limited than they would be with a modern big data platform. Although transferring to big data can be proven fruitful in patient treatment, the risk of exposing private information and violating HIPAA compliance is too great a risk to take. As this research will demonstrate, the patent industry faces the same challenge in security but no challenge in regulation.

2.7 How Has the Healthcare Industry Used Big Data?

Since 2010, healthcare reduced spending by \$300 to \$450 billion. This is in part due to the application of big data. A drop this large in spending accounts for 12 to 17% of annual healthcare costs (Groves, Kayyali, Knott, Van Kuiken, 2013, p.11). It has been proven that big data directly

reduces inefficiency in “clinical operations, research and development, public health, evidence based medicine, genomic analytics, pre-adjudication fraud analysis, device and remote monitoring, and patient profile analytics” (Raghupathi, 2014, p.2). Manipulating unstructured data for trend analysis is one way to reduce inefficiency. For example, by using big data analytics’ predictive features in Hadoop, the Center for Medicare and Medicaid Services detected and prevented more than “\$210.7 million in healthcare fraud in one year” (McDonald, 2016). This occurred because big data gives users the ability to go back in history and analyze all types of data to uncover anomalies and/or patterns. These findings are used to achieve business objectives and strategize plans, efficiency, and productivity.

A reason healthcare is adopting big data is because of Obama’s desire to reform the system. In 2009, the Obama Administration presented the idea of Health 2.0, a plan to improve patient, institutional, and insurance management. The plan’s success depends on big data. Part of the plan is *Pillbox*, which is a service that provides patients with important information about their medications. The user inputs details about the medication, such as name, color, or even shape. Then, the service will return important information about the medication that is patient-specific. Doctor Gang-Hoon Kim explains the benefits of this service: “Pillbox is designed to be user-friendly even for the elderly, or those who are unfamiliar with the internet. [It] is expected to reduce the cost of identifying pills and their effects, offer information about medicine using big data, and help maintain a clean medical system by checking the sale of medicine and medical records” (Jee, Kim, 2013). This use of big data is beneficial to the industry and patients for two reasons. First, a patient does not have to travel to a facility to consult a doctor about his medication. Second, doctors can devote their resources to in-person treatment of patients who experience

immediate and/or fatal medical issues. Unfortunately, the Obama Administration's budget did not allow further pursuance of Pillbox development.

Health 2.0 encouraged hospitals to engage in more information sharing. This is important to healthcare because it allows patient information to be distributed quickly to technologically connected facilities. If a patient transfers to a new facility, his records can be shared electronically. This process saves time and money since patients do not have to complete redundant paperwork and doctors can study patient records in advance before making a proper diagnosis.

Indiana readily adopted Health Information Exchanges, also known as HIEs. Over 80 hospitals in the state are connected to transfer and share "information on more than ten million patients. Now, over 18,000 physicians can take advantage of this data" (Groves, Kayyali, Knott, Van Kuiken, 2013, p.3). They use the data to figure out what treatments work best for specific patients, when cures might not be as effective, and where fraud or poor use of resources exist. On a basic level, physicians use HIEs to easily and quickly transfer patients from office to office. The state reaps the most benefits in the emergency room. When patients are rushed to the ER, doctors immediately obtain patient information and can adjust treatment methods to avoid exacerbating the situation (Healthit.gov, 2014). HIEs save time, money, and lives when implemented in geographically broad healthcare systems.

Although Indiana readily adopted electronic record sharing, other states struggle with this task. Through government funding and incentives, some organizations were created to help this transition. *Premier* is a group-purchasing organization that compiles information from hospitals and healthcare providers. As a result, it "provides data-driven informatics specific for each group member" (Groves, Kayyali, Knott, Van Kuiken, 2013, p.6). This service improves patient

information sharing and ensures more cost effective and timely patient diagnosis. Already, more than 2,700 facilities and 400,000 physicians are signed up with *Premier* (Raghupathi, 2014, p.8).

Because of easy access to an abundance of information, *Premier* generates “clinical outcome measures, resource utilization reports, and transaction level cost data [reports]. This information improved healthcare processes at approximately 330 hospitals, saving an estimated 29,000 lives and reducing healthcare spending by nearly \$7 billion” (Raghupathi, 2014, p.8). *Premier* accomplished these tasks because of big data’s speedy ability to tap into mass amounts of unstructured data and seek out patterns and/or anomalies.

In 2009, the Obama Administration passed the Health Information Technology for Economic and Clinical Health Act to make EHR transfers more attractive to hospitals. The administration offered incentives of \$40 billion and an additional \$2 billion to train employees to use EHRs. It was predicted that by 2019, 70 to 90% of all providers and hospitals will make this switch (Groves, Kayyali, Knott, Van Kuiken, 2013, p.7). According to *HealthIT.gov*, as of January 2016, “over 90% of Regional Extension Center enrolled providers are live on an EHR, and 3/4 of enrolled providers have demonstrated meaningful use of certified EHR Technology” (Healthit.gov, 2016). Figure 2 below demonstrates this success. The X-axis notes healthcare providers and facilities that use EHRs since January, 2016. The Y-axis demonstrates the percentage of beneficial use the participating centers experience (Healthit.gov, 2016). It is conclusive that the Obama Administration’s initiative had a positive impact on making the use of, and transfer to, EHRs more widely accepted.

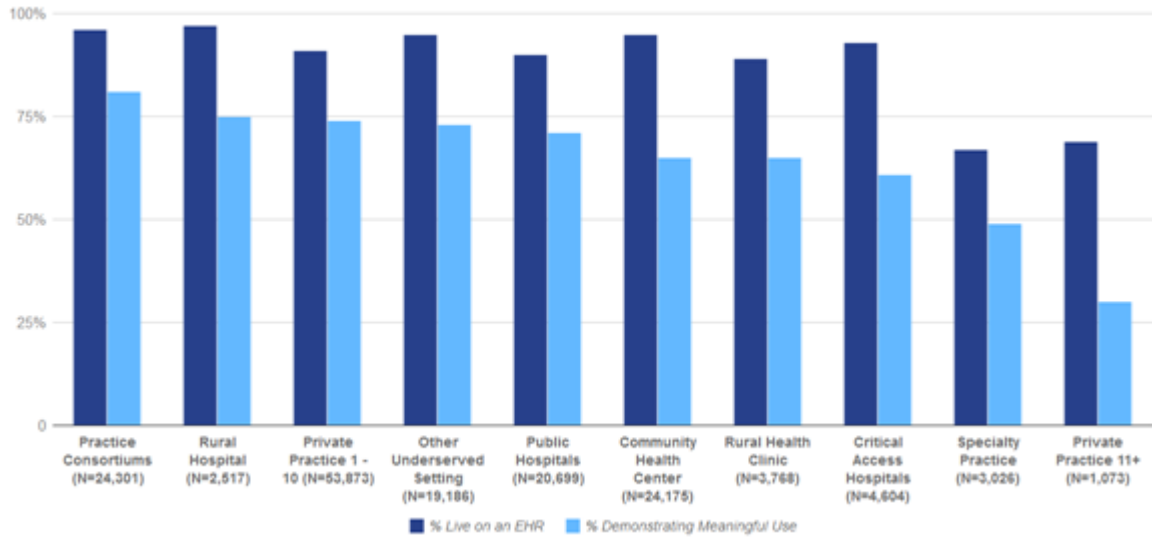


Figure 2: Health IT Quick Stats. (2016, January)

Aside from government promotion, individual organizations create similar initiatives. Managing Care Company *Kaiser Permanente* developed and implemented an information sharing system called *HealthConnect*. This database aids companies to exchange information with affiliated parties. Kaiser boasts a 26.2% reduction in total clinical office visits since the implementation of *HealthConnect* (Groves, Kayyali, Knott, Van Kuiken, 2013, p.10). Decreasing traffic allows the doctors to spend more quality time consulting patients with urgent cases.

Furthermore, the Obama Administration made it easier for facilities to share information with Common Exchange Representation, or CER. It is nationally defined as “an artificial language for representing the information in EHRs, which has well defined syntax and semantics, and is capable of unambiguously representing the information in any EHR from a typical EHR system” (Pasquale, 2013, p.61). Now, EHRs can be transferred to national facilities without the worry of data format incompatibility. The Administration helped eliminate some of the excuses that facilities would make about accepting electronic sharing.

Big data acceptance has become more widespread with the Obama Administration reforms. EHRs make it possible for patients to transfer amongst facilities with ease. Practitioners can access patient information to create more individualized treatment plans and save time, money, and lives. These benefits are widespread in the healthcare industry, and perhaps can be carried over similarly into other industries, such as the legal industry.

2.8 The Application of Big Data to Healthcare R&D

Whereas the healthcare section of this research presented how big data's application to patients, hospitals, and doctors is beneficial, this section highlights other benefits seen in healthcare research and development and pharmaceuticals. As will be demonstrated in the Analysis, parallel benefits can be reaped in the patent industry.

The Obama Administration continued to impose initiatives on medical research and development. It began with a 2012 \$200 million initiative to expand biomedical research specifically through big data analytics (Jee, Kim, 2013). Because of this emphasis on big data application, there are several applications of big data to R&D that are worth noting.

In 2009, IBM and the European Union combined resources to find a more effective combatant to AIDs. It is difficult to find a lasting treatment for HIV because of its morphing capabilities. Big data analytics makes this task easier because the data is constantly updated whenever changes in HIV resistivity are identified (Tomasco, 2009). As part of IBM and the EU's agreement, IBM contributed the *EuResist* project to this cause. *EuResist* is composed of big data tools that the scientists use in their research. Because of this new access, scientists conduct more effective genomic analysis than ever before (Jee, Kim, 2013). *EuResist* Scientific Coordinator Maurizio Zazzi noted, "The ability to analyze clinical, laboratory and demographic data

accumulated over the years significantly improves prediction of the right combination of drugs that works for the maximum amount of time” (Tomasco, 2009). *EuResist* maintains a success rate of 76%. This rate is unprecedented by traditional databases and research methods (Jee, Kim, 2013). The analytic power of big data produces these real-time results, which is crucial for healthcare R&D. As you will see later, this is just as important in the legal industry.

Other research that benefits from big data analytics pertains to asthma prevention. Researchers developed a device called the *Asthmapolis* to monitor inhaler usage. Big data is applied in the GPS tracker which sends geolocation information to a remote database. The database then combines the collected information with data from the internet about the location and population trends. With compiled *Center for Disease Control (CDC)* information about asthma catalysts, researchers procure a clearer understanding of what and where asthma is more prevalent, make trends and predictions about the development of asthma in certain regions, and determine how to create more patient-specific treatment and preventative plans (Groves, Kayyali, Knott, Van Kuiken, 2013, p.12). Without big data analytics, this would not be a feasible task and individual health would be at a higher risk.

As of April 2016, big data analytics also helps with the production of numerous vaccines. According to the *Industrial Engineer*, McKinsey & Company constantly uses big data to create patient specific vaccines. Their analytics provide in-depth reports on patient’s blood components as well as individualized predictions of patient reactions to certain shots. Vaccine production increased by 50%, resulting in savings of \$5 to \$10 million per year for McKinsey & Company (Industrial Engineer, 2016, p.13).

The successful application of big data analytics to real time research and development demonstrates its effectiveness and accuracy in predictions. It positively impacts people's lives

through preventive medical strategies and individualized patient treatment. Success seen from real-time application and manipulation of big data in healthcare R&D makes it a promising service for the patent industry to implement, especially since they have similar needs.

2.9 How Big Data is currently used in Litigation

Currently, big data currently is implemented to compare competing law firms. By analyzing information about costs and legal fees, law firms can adjust their rates against those of competitors. One big data service is TyMetrix Analytics, which is implemented for the above reason. Managing director, Craig Raeburn, describes the way this service increases firm competitiveness: “Law firms have utilized [it] to benchmark themselves against the industry. Many have used our LegalView data warehouse to compare their rates and understand the best way to position themselves with clients—low-cost provider or high-end value player” (Dysart, 2013). Services such as TyMetrix can change the legal landscape for the reasons mentioned above. Further developing these services will only increase the benefits that firms can receive in different aspects of the business.

Lawyer Metrics is another big data service that predicts and compares lawyer success rates. This is achieved by sifting through biographical data, including information about sports history and non-legal publications. Analyzing these factors ensures that the system “explores the associations between performance and several dozen success traits...that firms overlook or give less weight to.... It reveals that law firms are often systematically overvaluing some attributes, ignoring others that really matter, and generally making bad tradeoffs in both entry level and lateral lawyer drafts” (Katz, 2013, p.27). Evaluating the quality of lawyers and their ability to perform is

beneficial to any firm. With this knowledge, a firm can hire lawyers that are versatile, but also assign lawyers to specific cases that cater to their personality and work ethic.

There is also a steady rise in case prediction service, such as LexisNexis. This platform stores domestic and international cases on record. Its intention is to help lawyers find similar suits that can be used as precedent cases. Firms will hire chief knowledge officers, also known as data scientists, to extrapolate data that can prove useful in case development. One North Carolina officer, Bill Turner, explains, “The ability to learn in real-time and gain insights from meaningful, predictive data is increasingly important to delivering new levels of value to clients” (Dysart, 2013). It is promising that firms currently use big data, although its usage is limited. Analytics can be further expanded to predict case outcomes, just as it is used in healthcare to predict patient treatment and recovery procedures.

Big data is a key to unlocking effective case prediction. Intellectual property can require differently formatted data, images, and videos, which presents a challenge to lawyers who spend valuable time searching the internet for useful information. Identifying information that appears in different formats is not easy because of possible inconsistencies. Some of these inconsistencies include differently designed data structures (for example, an article may write a date in mm/dd/yy or dd/mm/yyyy format, a court case may write a date in month, date, year format, and an image might have the date in mm/dd/yy format on the bottom right corner), having no clear links between data and sources (for example, a court case references a patent in the USPTO but there is no obvious link, such as a patent ID or a website, to make the connection between the two data points easier to understand and access), data that is not organized (for example, a patent may include images or videos on its file that are necessary for case analysis, however they are not conveniently organized within the patent file to be quickly available and useful), and source data that does not

match its target (for example, a patent file references a prototype video on a webpage, but the link is not updated in the event that the prototype video is taken down or relocated on the web. This may increase the difficulty and time to create an analysis, because the proper link must be traced between the two) (Ryan, n.d., p.6).

To compensate for inconsistent data as described above, a conversion program is necessary. It is a problematic and costly addition to any system that must be held accountable for all possible variations of data inconsistencies, however it is necessary in the patent industry, in which lawyers need to analyze and make predictions while accessing all the data types used as evidence (Ryan, n.d., p.6). This is part of the data mapping process in developing any system, and it is crucial to understand beforehand what information is worth converting so as not to exacerbate costs. Using services like LexisNexis and hiring data scientists will help speed up the process of evidence acquisition and removing data inconsistencies so that all data formats can be beneficial to a lawyer.

The secret to discovering discrete information that can make a case strong is through semantic searches. This is a unique process in which a lawyer searches for concepts or words that relate to the case. The database will return an extensive list of data as well as previous cases that are useful for case development (Martin, 2013). Another way semantic searches are beneficial is to understand different legal systems laws, jargon, and patents. This type of search will translate these information points into understandable sets. As a result, barriers between international legal systems will be diminished.

Despite these benefits from big data analytics in the legal industry, there are limitations. The legal industry has not fully used all of big data's capabilities. It uses big data now merely for competitive pricing, which only brushes the tip of the iceberg. Big data can be used so much more

efficiently in valuable fields and applications, including risk management and early warning prediction. These are some of the limitations that this research plans to eradicate. Big data enables real-time insights for better strategic decision-making, risk management, and case prediction. These features would be invaluable to the patent industry.

2.10 How Patent Litigation currently uses Big Data

The patent industry does use big data to an extent. The United States Patent Utility (USPU) is a web-based licensing platform for inventors and growing businesses. The platform sifts through millions of patents from the US Patent and Trademark Office and analytically matches companies with concepts and/or inventions with similar business objectives (Litan, 2015). Although beneficial, it does not resolve some of the key issues inherent in the patent industry. The USPU offers only basic services that do not help detect patent overlaps or make predictions about future cases. This research will demonstrate how the industry can reap more benefits by implementing a proper big data analytics system that capitalizes on all of its capabilities.

Some firms utilize big data to identify cases that will yield a high success rate. LexisNexis is known for its new service Lexis Advance MedMal Navigator, a platform that predicts the outcomes of medical malpractice suits. Studying past cases enables it to identify potential witnesses and experts to be consulted during case development (Dysart, 2013). Lex Machina is also similar to LexisNexis, but it is used primarily for intellectual property law. With more than 4 million documents pertaining to patents, lawyers have access to the most lucrative information in the market. Backed by Apple, Cisco, Intel, Microsoft, and Oracle, Lex Machina can map “every electronically available patent litigation event and outcome to bring openness and transparency to IP law” (Katz, 2013, p.32). Lex Machina’s limitation is its lack of applicability to many firms. It

is a beneficial service, but only those who can afford paying for it have access to it. It also does not offer early warning and predictive analytics. It is more of an online repository of intellectual property court cases and patent records than anything else. We suggest, however, using it as a base to create the ideal universal big data system.

Lex Machina's primary programming language is Python, with additional Java processing capabilities. Lex Machina relies on the Pyramid Web Framework, which is open source for Python programmers. It offers simplicity, documentation, speed, reliability, and openness for developers and/or users with minimal programming experience. Therefore, it is not costly to invest, because as the Pyramid creators say, "You can get results even if you have only a partial understanding of Pyramid. It doesn't force you to use any particular technology to produce an application, and we try to keep the core set of concepts that you need to understand to a minimum" (Pyramid, 2016).

Although existing, the current implementation of big data services in the legal industry is severely limited. This implementation deficit is what we intend to fix, specifically in patent litigation. This can be achieved by implementing a universal version similar to Lex Machina that amends the patent filing process, produces early warning systems, and develops more cost effective preventative strategies inside and outside of the courtroom. This type of system would be available to all patent industry practitioners.

2.11 Gap Analysis

We have the ability to resolve some of, if not all, the issues in the patent industry. One of the key dilemmas is that an influx of minor and repetitive patents are granted. As a result, there is a rise in the number of lawsuits. This creates a trickle-down effect, because now it is much easier to patent an idea. This means fewer ideas may be competitively innovative. This is not a good

repercussion, since innovation is a key factor that provokes companies to make revolutionary advancements to technology. The types of patents that drive innovation and competition are termed “homerun” patents. Because of the ease in granting patents for almost any idea, homerun patents are often brought to court against the newer, less innovative and competitive patents, which wastes time and money for all parties. When many generic patents flood the USPTO, resources are diverted away from innovative patents that can be more beneficial to society (Litan, 2015).

Our research will demonstrate big data’s ability to alleviate this industrial burden. An analytics system can sift through existing patents to see if there are matches with incoming applications. If overlaps exist, the system can weigh the impact the patents may have on the business environment and whether or not granting a patent is just in the first place. Aside from helping companies avoid costly settlements, resources would be saved for USPTO examiners who need to focus on homerun patent reviewal. With this respect, a big data system would provide optimal benefits to the USPTO in making the application process swifter and less prone to mistakes, such as lawsuits about patent overlap.

The benefits of a big data system are in fact plentiful. For example, Harvard Business School Professor Karim R. Lakhani notes that businesses can save money by hiring data scientists who conduct patent and case searches, because they would avoid high legal licensing fees. Many professionals and analysts concede that “there’s no better time to unlock and democratize the US patent portfolio [than now]. When it comes to expanding US business opportunities, [big data], this oft-misunderstood resource, proves itself to be a gentle giant, not a Goliath (Litan, 2015). It is crucial that businesses and the USPTO understand the power of big data and embrace it. Big data is not something to fear, but rather something to embrace.

The benefits of using big data outweigh the costs, making the use of a big data system in the patent industry more viable than before. This research will add more to what we already know and don't know about big data analytics and its application to the patent industry. It will help fill the gaps to answer questions such as, why isn't the patent industry already using universal big data services to alleviate industry burdens? Finally, this research will present a solution to the issues present in the industry by suggesting the required features for a universal big data system.

2.12 Literature Review Conclusion

The research reviewed can be applied to the final synthesis of big data and patent law. Big data's application to healthcare and its current position in the legal industry provide a basis for its future in patent law. Currently, it helps decrease costs, save lives, and promote R&D in healthcare. In the legal industry, it cuts costs, helps firms compare competitor fees and rates, and predicts suit outcomes. The healthcare and patent industries are very similar in that they both require proactive planning and real-time access to strategic information. It is unclear why the patent industry does not implement more big data analytics, other than the cost of doing so. Big data can be used effectively if implemented in a universal system available to all patent practitioners.

Early signs of big data adaptation exist in the patent industry, such as Lex Machina. However, a gap exists in the pre and post patent application process. Pre-application, the USPTO is flooded with an overabundance of applications. This makes it difficult for examiners to study each idea and grant patents to those that fulfill all the requirements. Post-application, patent holders seek avenues to sue another patent holder for infringement. Going to court is an exorbitant expense which exhausts many resources that can be used for more crucial cases in the legal system, as well as for innovation. The case study for this research is *Apple v. Samsung*, a legal situation that molds

the industry's future. The case encompasses many of the issues discussed in this research, and therefore is a strong predictor of our future in patent litigation.

The gaps discussed in the patent industry can be amended with effective big data analytics implementation. Such change can lead to an early warning system and preventative strategies for inside and outside the courtroom. Big data can dictate which patents should be granted and which ones overlap. The Analysis will demonstrate how and why big data can be applied as a strategic tool to patent litigation and the patenting process. This will undoubtedly cut costs, create more effective strategies and cases, and predict case outcomes to avoid such a debacle as *Apple v. Samsung* from occurring again.

The cost of implementing and maintaining a big data service is a small price to pay in comparison to the thousands of dollars that go into filing and settling patent infringement suits. It is important that the industry understands what current platforms exist and what the best languages are for implementation. The Analysis chapter will also produce a recommendation for the best platform, language, and other features for the patent industry.

3 Analysis

3.1 Existing Gaps in Big Data's Application to Litigation

A key aspect of intellectual property litigation is risk analysis. Because it is “a matter of critical importance for parties involved in IP litigation to continually assess their respective risks during the entire progression of the case, starting from its filing time, or even prior to filing,” proper risk analysis should be available to diminish the possibility of exacerbating costs (Surdeanu, Nallapati, Gregory, Walker, Manning, 2011, p.1). As a result of not performing proper and thorough risk analysis, inventors may lose patent ownership and become indebted to patent monetizers. In some instances, damages paid by the defendant can amount to millions of dollars, as seen with *Apple v. Samsung*. It is usually hard to settle IP litigation because while the law says one thing, consumer use and favoritism might say another about the technology in question.

Risk analysis yields several key benefits. Its immediate effect is to help decide whether to file a case or not. Involved parties can predict which direction the case will take and what costs they will incur. Ideally, this would deter long term and expensive cases from being filed. The proper risk analysis would be composed of “a logistic regression classifier to capture historical features and a novel relational model using conditional random fields to jointly predict the outcomes of concurrent and related cases” (Surdeanu, Nallapati, Gregory, Walker, Manning, 2011, p.5). In other words, the logistic regression classifier seeks out anomalies and patterns by analyzing historic, unstructured data and applying these findings to the current cases and data that might be related. This algorithm has the capability to produce 22% more accurate outcomes than when a lawyer conducts manual research (Surdeanu, Nallapati, Gregory, Walker, Manning, 2011, p.6). In so doing, the user can determine if the current case has the potential to be as expensive and lengthy

in court as some past cases. The logistic regression classifier will offer the user a prediction of the costs and length of the case. This is one reason big data is important; it can be used to access past information and create value from it. This is something lawyers have not been able to do in the past.

A risk analysis also demonstrates if the costs of arguing in the courtroom outweigh the benefits of settling outside. Settling outside requires different mitigation techniques and an agreement where both parties compromise. Usually, one party will not suffer indefinitely at the hands of the other. A risk analysis that compares the two mitigation strategies for individual cases will ease burdens on the judicial system, judges, and juries, resulting in a lighter caseload and more time for more serious lawsuits. Therefore, minor patent monetization filings should not expend as many valuable resources as they currently do, especially in districts such as the Eastern District of Texas.

One challenge with a big data risk analysis is keeping it up-to-date. IP cases and patents are filed daily alongside the general trend of technological growth. An analysis platform derived from a traditional relational database would not solve this challenge like big data. Data warehouses are automatically scheduled to constantly update the stored information. Since big data performs functions automatically, involved parties would not have to expend valuable time searching for evidence. A user may only need to know keywords to perform big data fact-finding in case development.

The ideal risk analysis would consider the case's merit and prior factors. Case merit should yield any conflicts that may currently exist or arise in the future between the questionable patents. If there is no relation between the two, then the case has no merit and can be dismissed. A big data platform accelerates this process by reporting about similarities and differences through the

implementation of an algorithm. This algorithm can be as simple as a semantic search, which is discussed in further detail later.

Prior factors will determine the likelihood for success in the courtroom. Factors under consideration include opposition success rates, potential bias of the presiding judges, and circumstances that may result in unprecedented outcomes (Surdeanu, Nallapati, Gregory, Walker, Manning, 2011, p.2). This research tool can be implemented through an algorithm, similar to those comprising LexisNexis. However, it is noted that the algorithms in LexisNexis are only advanced enough to compare competitor costs and success rates. The ideal tool would build upon these capabilities, and further detect relationships amongst cases and judge bias.

Another issue industry professionals face is a language barrier. Patents may implement differing legal syntax, especially if they originate from different countries. Differences in global legal systems make arbitration challenging for universally accepted patents. Because technology is widespread and patented globally for the international market, infringements are often unavoidable. Resolving these cases is more challenging because they need to comply with the differing legal systems, and yet they are often resolved in a country that is not their filing country. Patent holders will do this because the decision might be more favorable than in another legal system.

Apple v. Samsung is one case that was brought to different countries, each offering somewhat contradictory decisions. Ideally, a big data platform should adapt patents to every legal system. It would translate languages and legal procedures, finding similarities amongst the different legal systems. This can help avoid international litigation that does not reach the same decisions and deter companies from choosing some legal systems over others (Martin, 2013). At first, we considered a big data platform as a means to fix the American patent industry. However,

technology reaches beyond global boundaries, and therefore it is not feasible to isolate the patent industry, its litigation, and its implications to one country. A universal big data platform that we propose should also be available to international legal systems.

Discovering overlaps in patents is possibly the most important task a big data platform can achieve. This is because it can prevent a clog in the USPTO's application process and prevent future suits in court. Such a use of big data does not currently exist in the patent industry, therefore creating a gap. As previously mentioned, this gap can be bridged with a semantic search in big data. A semantic search is officially defined as "a data searching technique in a which a search query aims to not only find keywords, but to determine the intent and contextual meaning of the words a person is using for search" (Redondo, 2014). In 2013, Google announced its semantic search algorithm called *Hummingbird*, which is intended to produce more specific search results in web browsers. For example, if a user types "Asian cuisine" into the search engine, closes the window, opens a new one, and types in "Chinese," Google will return a list of Chinese cuisine. *Hummingbird* allows Google to access and create historic-based trends to predict exactly what the user would like to see.

In our big data platform, thorough analysis of what the user enters can yield more specific results, quite like Google's *Hummingbird*. This would be considered a domain-specific big data service because it "unleashes the power of previously unknowable insights into the trajectory of technology innovation and core company-specific product plans" (Martin, 2013). This type of big data implementation has the potential to advance strategic decision making, thus becoming a deterrent for patent monetization and costly legal warfare.

Something lawyers currently are not privy to is metadata. Digging into patent metadata can uncover trails to other patents and other useful information for case development. The user may

be overwhelmed with information overload, however big data brings together the structured and unstructured data, therefore saving time in research. This also avoids alienating the user with an abundance of information. With big data, lawyers can capitalize on metadata trails (Martin, 2013). For example, creating these links will expose the frequency that similar cases go to court. As a direct result, lawyers can make more informed decisions when preparing for court so that they are not there for a long time. Lawyers can study in more depth the patents' contentious parts and become aware of what the future may hold for the case.

As mentioned in the Literature Review, LexisNexis is a service that some firms use. One California based firm, *Dummit, Buchholz & Trapp*, uses it because it “allows [one] to determine in 20 minutes—versus 20 days—if a case is worth taking on. This tool is a powerful, one-stop solution for attorneys” (Dysart, 2013). Malpractice attorneys are LexisNexis' biggest clients. The algorithms used can be beneficially integrated into our system for IP professionals. For example, LexisNexis stores domestic *and* international cases. This feature is crucial for reasons previously mentioned. Chief Architect of the company, Ian Koenig, says, “We have the tools and analytics to look across [all] this data in a way that helps attorneys glean insights into potential case outcomes. There are things law firms don't know today simply because they've never stored and correlated all this data” (Dysart, 2013). This is important because it opens many doors for predicting case outcomes and finding the most efficient way to resolve a case. In the IP industry, it will be incredibly beneficial since patents reach beyond domestic boundaries and question patents around the world. LexisNexis will be one of the good starting points for the development of our big data system.

Alongside LexisNexis is Lex Machina, offered only to IP firms to reap similar benefits as malpractice attorneys do. Lex Machina searches the internet for information that can be used for

existing cases. Collected data is stored in a database, where basic algorithms are implemented to map out possible correlations. The service can predict the direction a case may take. Silicon Valley IP attorney Vicki Veenker expressed her faith in Lex Machina: “My prediction is that in the long term, others will join me and big data will bring more transparency to the IP marketplace” (Dysart, 2013). Her support for big data implies that it can be very helpful to the patent industry. This provides hope for a future in this field in which big data is widely accepted and used for all its capabilities. As previously mentioned, Lex Machina is only available to firms that can afford an expensive bill for using the service. The ideal big data system would be universal with more advanced features than Lex Machina.

In conclusion, services do exist to enhance case development. Although the patent industry has access to Lex Machina, it only provides limited resources and not all firms and general industry practitioners can access it. Much can be done to improve what currently exists, and to make big data analytics more adaptable for different cases. The following sections will annotate how healthcare uses big data analytics services, how this relates to the patent industry, and how its potential application to *Apple v. Samsung* might have changed the overall decision and the way patent cases are currently resolved in the courtroom.

3.2 The Relationship between Big Data, Healthcare, and Litigation

The healthcare industry sets a good example with its use of big data early warning systems. One notable system is MediSys by the European Commission. This system identifies potential threats to public health by pooling information from internet reports and other sources. Search and analysis is limited to current and pressing public health issues. This is because if there is a national health issue, professionals will need access to only the most recent information. Analysis based

off past information is not helpful to these professionals when having to make real-time decisions about pressing health issues. The emphasis on recent information is made possible with an algorithm that analyzes severity, frequency, and novelty of article content. When the system finds breaking news, it alerts public health officials, who then take the matter into their hands and notify the public (European Commission, 2014). This is one example of an effective and immediate early warning system. Its success makes it a good forerunner for the early warning system that this research outlines for the patent industry.

Other applications of big data to biomedical R&D are seen across the board. Columbia University utilizes big data to better understand patients with traumatic brain injuries. Combining experimentally collected physiological data with internet sources helps to identify “critical and timely information to aggressively treat complications. The advanced analytics is reported to diagnose serious complications as much as 48 hours sooner than previously in patients who have suffered a bleeding stroke from a ruptured brain aneurysm” (Raghupathi, 2014, p.8).

If big data can be applied to healthcare, where real-time is crucial for saving lives, it is reasonable to suggest that it can also be applied to the patent industry, where real-time is also crucial to resolve issues in and out of the courtroom. Both healthcare and IP rely on analysis of structured and unstructured data, access to constantly updated information from an abundance of sources, and predictions about the future of individuals and patents. Since healthcare has already laid the foundation for big data usage in existing platforms, the patent industry can use these as a foundation to develop a patent specific system.

Pharmaceutical R&D also implements big data analytics when it comes to patenting drugs. A big data service called PubMed is offered for researchers to search through existing patents (Magalhães, Quoniam, Ferreira, Boechat, 2013, p.2). By using this service, scientists will discover

if similar drugs are already patented and on the market. This can avoid costly litigation. Secondly, by using what research already exists, scientists can produce competitively new and improved drugs. This promotes innovation without the threat of litigation. PubMed also offers what we would define as predictive trend analysis. These trends aid scientists to foresee the direction that medicine and epidemics may take. These features would be highly beneficial to complete many of the patent industry's big data tasks. Therefore, PubMed's algorithms and implementation can act as another foundation.

A reason scientists and doctors prefer access to big data versus traditional data stored in relational databases is because it facilitates understanding global health “that is fundamentally different from [information] yielded by the disease reporting of the traditional public health infrastructure” (Magalhães, Quoniam, Ferreira, Boechat, 2013, p.9). The traditional infrastructure does not necessarily yield real-time results, which can make it more prone to errors. But since big data returns results in real-time, scientists make more accurate and timely decisions in drug development. As previously mentioned, the same feature is crucial for the patent industry.

The few examples above provide a backbone for understanding the direction we would like to take in terms of building a big data platform for the patent industry. It is important to remember that both the healthcare and patent industries require proactive planning and real-time access to information. Just as PubMed can prevent drug patent infringement, a similar platform can be as effective in preventing unnecessary technology patent litigation. In both industries, this layer of protection makes the invention and patenting process easier with less fear for legal repercussions. PubMed is the poster child for our big data analytics system. Now, we can analyze *Apple v. Samsung*, and apply a big data early warning system to it in order to yield an alternative outcome.

3.3 Apple v. Samsung: The Decision and its Applicability in the U.S., Japan, and China

Although *Apple v. Samsung* was brought to several international courts, the United States decision held the most weight. The American jury decided that Samsung infringed on more of Apple's patents than Apple infringed on Samsung's patents (Tibken, 2014). They reached the consensus that Samsung did not copy some features, such as slide-to-unlock, which appears on almost every smartphone. The jury's reasoning was that one company cannot hold a monopoly over such a universal mechanism (Tibken, 2014). This decision confirms the requirement that a patent cannot protect something deemed common or necessary. A majority of smartphones, if not all, require a version of the slide-to-unlock feature for security purposes. Therefore, it cannot be defined as a unique characteristic that only one developer has the rights to own.

Apple v. Samsung was also settled in Japan, which has a specialized court called the IP High Court. The High Court's judges are more experienced in IP than those presiding in the U.S. system. Furthermore, these judges are aided by former patent examiners who yield a more thorough understanding of technology limitations and jargon. Therefore, "armed with complete knowledge of technical details and the patent application examination procedures, the system ensures that the IP judges clearly understand the technology and the law when deciding the scope of a patent claim and whether an infringement has occurred" (Bajwa, 2014, p.97). As a result, there is more informed decision making and less biased results. On the other hand, in the U.S. system, the responsibility rests on the judge, or a "non-biased" jury, that must apply their limited understanding of IP to preside over cases. The judges and/or jury may not receive adequate professional guidance as they do in Japan, making final decisions rather questionable.

Whereas the U.S. jury favored Apple, the Japanese IP High Court favored Samsung. Research can only speculate as to why Japan supported Samsung's allegations. Some suggest that the U.S. jury was predominantly composed of Apple users and that Japan was biased towards Samsung because it is an Asian company, headquartered in Korea. Regardless, both decisions demonstrate the difficulty a legal system has in reaching an unbiased, educated decision. It is even more challenging to reach such a conclusion regarding technology, because everyone relies on it in different countries around the world; we all experience favoritism towards one product over another. This leads to an inherent bias and obvious divide amongst tech users and future jurors that may preside over patent suits in the U.S.

Intellectual property disputes are also settled differently in China. Although China did not settle on *Apple v. Samsung*, they did bring claims against Apple for selling iPhones that compete with Chinese phones. Many thought China's *Shenzhen Baili Marketing Services Co.* would lose to Apple, however they were proven wrong. China comes out strong in patent litigation because of its recent reforms. Before 1985, patent law did not exist. Since then, China developed specialized courts with seasoned judges for intellectual property disputes. They follow the same change as Japan undertook. China also offer rewards for noteworthy inventions and ideas, which is their version of "homerun" patents (Chin, Nicas, 2016). At first, there was an influx of patent filings, because people jumped at the notion of receiving cash rewards. There was also a proportional increase in overlapping ideas. As a result, the Chinese legal system tightened the application process, making it more difficult to patent ideas, and therefore significantly diminishing the rate of infringement. Even with these system changes, China demonstrates competitive innovation unlike any other country.

Although infringement damages in the Chinese legal system are astronomical, the Chinese courts constantly deal with patent suits, especially ones involving foreign parties. American companies, for example, prefer settling IP suits in China because they are more likely to win abroad. To put it into perspective, from 2006 to 2014, 81% of foreign plaintiffs won in Chinese courts. The reasons for this are that “foreign firms...only sue in China if they are confident they can win” and it is extremely inexpensive to file compared to filing in America (Chin, Nicas, 2016). A firm can spend less than \$30,000 on a case while in the U.S., it would cost between \$650,000 and \$5 million for a claim of up to \$25 million (Kerstetter, 2012). In China, as the number of patent infringement cases increased four times over the past 3 years, so did China’s enforcement of the law. The American system does not practice the same way, and that is another reason we see many issues inherent in our system.

A strong patent legal system, as seen in China, may not be feasible to implement in America, mostly because of cost. To implement China’s system here, we would have to reconstruct the entire American legal system, including passing new amendments and bills to install specific IP courts and judges, incentives, rules, and regulations. The cost to train judges to become professionals in IP would be astronomical. Therefore, to achieve similar success rates in China, the American patent system ought to adopt a more affordable route to change. This route is building a universal and global big data analytics system.

Had the American system already been built with the legal ramifications that China and Japan possess, Samsung might not have received the backend of the settlement in the U.S. This brings into question just how effective a big data risk analysis could have been. With such a tool, Samsung could have studied the composition of the jury to foresee the direction the case may take. If the company saw beforehand that a majority of the jury was Apple users, it might have

approached the trial differently. The legal costs of pursuing such a case in court might have been more clearly outlined, resulting in a pursuance of alternate settlement. Or, Samsung might have taken advantage of a big data risk analysis to press charges in a different country, such as Japan or China, where its word may have more weight over Apple's word. Without a doubt, the costs of litigation could have been compromised if Samsung had access to a big data risk analytics system.

3.4 The Beneficial Features of a Big Data Risk Analysis

Several features in big data make a risk analysis effective in gleaning speedy insights into a scenario. Management of the data is important to ensure that it is reliable. Processes that are repeated when new data is brought into the system are required to maintain a high level of quality in the data being stored. Once there is quality assurance, the data can be mined. Data mining software helps the user quickly sift through mass amounts of information in order to uncover patterns. These patterns comprise the first stages of analysis. In addition, this software will decipher what is relevant and not to the question at hand, ignore repetitive data, and finally, assess likely outcomes and answers of the issue. Although all this can be achieved by a single individual, it becomes an overwhelming process, making it difficult to create informed decisions in a timely manner. That is where big data analytics comes in handy.

Another feature of big data risk analysis is in-memory analytics. The system may store data in the memory, instead of on a hard drive. This makes the process of accessing and working with data more immediate. This is an attractive feature to users because of its ability to run iterative and interactive analytics scenarios. Therefore, the prepping data that was once used to test new scenarios and create models on a hard drive is eliminated. Efficiency and accuracy again are exemplified with this feature of big data risk analysis.

The statistical algorithms and machine learning techniques of predictive technology in big data composes a fourth feature in risk analysis. This relies completely on accessing historical data that may be stored on a server, database, or hard drive (SAS, 2016). Because of cloud storage, storing mass amounts of information is made possible and more cost effective. In regards to the legal industry, storing all past cases on a cloud would be the most feasible option, and therefore would greatly help to glean beneficial insights about risk.

Finally, the text mining capabilities of big data encompasses everything else involved in successful risk analysis. Software coded in Python or other programming languages makes it easy to search the web, emails, social media feeds, blogs, and other intelligence to reap insights that the human eye might not have seen otherwise. With this newly collected information, a proper big data risk analysis can be made.

Every one of these features could have been taken advantage of in the *Apple v. Samsung* case. An effective analysis and mining of historical data, past cases, and online articles and information about smartphone technology and the two companies would have been made possible. Informed results would be presented to the judge, jury, and the two parties in less time than it took to settle the case in court. Imagine the costs saved from such a service. And then imagine the effects that this risk analysis would have had on the future of smartphones and patenting technology in general.

3.5 Thoughts on the Outcome of Apple v. Samsung

In 2012, Apple received upwards of \$1 billion in settlement from Samsung. As of December 2015, Samsung agreed to pay the remainder, amounting to \$548 million. However, Samsung continues to fight for reimbursement (Tibken, 2014). Despite the costs Samsung owes

Apple, it is otherwise relieved about the case turnout. Sales did not diminish, although they took a plunge after the Samsung Galaxy Note 7 caught on fire, was banned from aircrafts, and recalled. If anything, following the suit, Samsung began producing bigger and better devices for Apple to compete with.

Apple, on the other hand, is not as pleased with the decision. Its representatives' adamant opinions were expressed in a press release shortly after the 2012 verdict: "Today's ruling reinforces what courts around the world have already found: that Samsung willfully stole our ideas and copied our products. We are fighting to defend the hard work that goes into beloved products like the iPhone, which our employees devote their lives to designing and delivering for our customers" (Tibken, 2014). What Apple does not acknowledge is that Samsung developers put in the same effort and creativity to create successful and advanced devices. Creators do not receive the proper recognition and protection for their creations, and it does not matter what company they work for. With big data analytics, however, creators can be ensured a level of protection, whether they are backed by a large company like Apple or they are independent inventors. The ideal big data system would help creators identify possible overlaps, gaps, and/or similarities amongst existing patents. As a result, creators would not have to fight in court to defend their work, and lawsuits such as *Apple v. Samsung* would not be necessary to assert creative and design rights.

Apple and Samsung continue to pester each other with infringement suits so long as reform is not made to the patent industry. Unfortunately, this means that smartphone patent wars are unavoidable in the short term. We can implement a big data system to resolve the current issues. This would be a necessary precaution in order to prevent smartphone wars from overpowering the global market and hindering growth and innovation. Already, *Apple v. Samsung* shaped the

landscape for patent creation *and* infringement. We do not want the landscape to keep changing, so it is necessary that we stabilize it with big data analytics.

3.6 The Reasons Reform is Necessary in the Patent Industry

The decision in *Apple v. Samsung* has the potential to shape how companies design and protect products in the future. We are well aware that the smartphone market is rapidly growing. In 2013, U.S. smartphone revenue was \$58.2 billion, and by 2015 it reached \$83.5 billion (Marr, 2015). At the same time, there has been an increase in the number of smartphone patent cases filed in the Eastern District of Texas. Despite this fact, the overall growth rate has not suffered. It is ironic that after seeing the astronomical costs involved in *Apple v. Samsung*, companies continue to go to court for supposed infringement. This is one of the many reasons that researchers urge reform to be instigated in the patent industry.

The problems that envelope the patent industry make the entire system flawed. One flaw is that the USPTO is not selective enough when granting patents. Stressed with the massive amount of patents that flood the office, examiners are more likely to pass over details in the applications that may already be patented. This creates another issue in the form of inevitable overlaps, therefore leading to increased lawsuit filings. This frequency of patent filings puts pressure on the judicial system, creating more problems that could have otherwise been avoided with reform to the initial application process. As Professor Sheppard of University of Nebraska College of Law says, “Software patents are clogging the system at every possible point,” thus presenting challenges to not only patent holders but also the judges and jurors down the line in the patent industry (Vascellaro, 2012).

A second flaw is the inability of the jury to come to a knowledgeable and unbiased decision. As previously mentioned, patents are complex in nature. Settling on technology-related disputes requires a jury that shows no preference to the technology in question but also has enough knowledge to make an educated decision. The fact that a “federal judge recently threw out a high-profile case between Apple and Motorola, saying the patent system was in chaos” proves that knowledge, patience, and reform is key to legal success (Vascellaro, 2012, April 12). It is not beneficial for the system nor society if cases are thrown out because judges and/or jurors do not comprehend the technological and legal jargon.

In the U.S., judges are not adequately educated to make informative decisions for such cases. At least in the Japanese and Chinese judicial systems, patent specialists aid the judges. If there are no patent aids, then at least the judges are educated on the subject matter. Therefore, an IP case is not disregarded because a judge deems it too complex. Since modern technology continues to grow as a daily necessity, it is crucial that the legal system and laws keep up with the changes.

Apple filed over 250,000 patents to protect the continuous improvements it makes in design and utility features (Inmon, 2015, p.1). Some of these modifications, as deemed by the jury, cannot be patented because they are necessary for every smartphone. Yet, Apple continues to fight for these supposed rights. According to an article published by the *Wall Street Journal*, the goal of smartphone companies is “to find a patent that sticks, and to force competitors to work around it or strike a licensing deal” (Feldman, Ewing, Jeruss, 2013). This is exactly what Apple does, even if it goes against the 2012 *Apple v. Samsung* verdict. This in turn has a negative impact on the industry, because it deters other companies from pursuing innovative ideas. Those companies either do not have enough money to buy out Apple’s stake on an idea, or they do not want to lose

their rights by agreeing to a licensing pact with Apple. This is an industry where one company should not have a monopoly over the others, especially if we would like to see continuous growth in smartphone technology as well as the overall market.

Since the decision, Apple expanded its patent count to encompass ideas that global competitors might have on different systems. Steve Jobs admitted in a press conference that Apple would “spend every penny to fight copycats” (Feldman, Ewing, Jeruss, 2013). Tim Cook, Apple’s Chief Executive Officer, affirmed Jobs’s opinion, telling investors in October 2015 that “Apple spends a lot of time and money and resources in coming up with incredible innovations. We don’t like it when someone else takes those” (Feldman, Ewing, Jeruss, 2013). Apple fears that other companies will take advantage of loopholes in the patent system to claim Apple’s ideas as their own. Two major questions formulate as a result. First, is Apple technology more innovative than that of its competitors? And second, is waging patent warfare the best route to affirm market supremacy? Answering these questions presents a challenge to professionals. What we do know is that during *Apple v. Samsung*, Apple controlled only 19% of the global smartphone market while Android dominated an astounding 64% (Inmon, 2015, p.1). This statistic demonstrates that Apple may be desperate to hold on to its market share, no matter the costs. Unfortunately, Apple and Samsung’s fight for dominance threatens smaller companies from joining the market, which has a negative impact in the form of stunted market growth and innovation.

It is unsure what route smartphone wars will take in the future. As a Rutgers School of Law professor said, “When you have companies spending hundreds of millions in [minor] litigation, something is seriously wrong with our patent system. You’ve got to wonder whether it’s doing more harm than good” (Feldman, Ewing, Jeruss, 2013). Some IP experts believe parties will avoid litigation with cross-licensing agreements. In this alternative, an idea is shared by many, but the

involved parties often pay exorbitant prices to use it. Resources may be better allocated in this option, however it does not fix the inherent issues in the patent system.

In recent years, Apple pursued cross-licensing with Motorola and Google, despite its continuous presence in court with other companies. Negotiations made outside the courtroom guaranteed that Apple would “be able to take better advantage of the innovations they and their competitors produce, while giving proper credit to the innovators and allowing them a fair return on investment” (Inmon, 2015, p.4). Although this alternative does not guarantee a remedy for all issues we outline in this research, it can be beneficial. One of the purposes of the ideal big data system is to create preventative strategies inside and outside of the courtroom. If the system’s early warning analytics alerts a creator that costs will be raised if he pursues trial in court, then it will motivate the creator to resolve the issue *outside* of the courtroom. Since Apple admits it is more advantageous to negotiate agreements outside of court, cross-licensing agreements will become more popular and considered a solution to lengthy suits in court. The parties involved will benefit: less resources are expended to maintain a court case on all sides, judges and juries can allocate their time to more pressing cases, and creators can still maintain their rights and use of the questionable patents.

Unfortunately, patent trolling still remains an issue if negotiations are pushed outside of the courtroom. Patent trolls are known as “entities [that] concentrate on generating income by licensing or litigating patents, rather than by producing an actual product” (Cusumano, 2013, p.1). If anything, cross-licensing agreements may increase the power and influence of patent trolls in the industry. Although Obama declared patent trolls to be a major obstacle in 2013, no reforms have been instigated to diminish the trolling. Patent trolls make litigation messy, worsening the problems we already face. Essentially, patent trolls attempt to monopolize existing ideas and profit

at the expense of the inventors. Small startups and individuals lack the knowledge and funds to buy out existing patents as leverage against competing trolls. Therefore, individuals become susceptible to lawsuits that are filed against them.

As previously mentioned, the system also lacks ample information and resources to potential and existing innovators about the patent world. As a direct result, parties struggle in understanding what patents exist and what litigation has already occurred. Services like Lex Machina are offered for this purpose, but not necessarily affordable nor advanced enough for all patent industry practitioners. Unfortunately, inventors and small companies will enter the courtroom blind, lacking an understanding of the intentions and severity of trolls' allegations against them (Cusumano, 2013, p.65). The only defense these parties have is a lawyer who will spend hours and money searching for ways to protect the client. This manual protection can be replaced by an automatic and affordable universal big data analytics service that lawyers can use. A lawyer would decrease case development time and make more well-informed decisions for his client. He might even have the potential to avoid litigation in the courtroom as a whole. A big data system would further diminish the growth of patent trolls' monetization, as discussed in the subsequent section.

3.7 The Added Problem of Patent Monetization in Smartphone Warfare

Recent studies indicate that technology patents are more likely to be litigated than other types of patents. According to researchers John R. Allison, Emerson H. Tiller, Samantha Zyontz, and Tristan Bligh, internet-related patents are "litigated 7.5 to 9.5 times more frequently than [other] patents" (Cusumano, 2013, p.13). In fact, when studying tech startups, they discovered that

“79 of [223] had been ‘trolled,’ that is, threatened with a patent monetization lawsuit unless they acceded to a licensing agreement” (Cusumano, 2013, p.13). Patent trolls predominantly target smaller companies, individuals, and/or startups with minimal assets, because they have less legal knowledge and money to defend themselves. As a result, patent monetization continues to grow. In just 2015, 67% of all patent lawsuits were filed by trolls alone (Mullin, 2016).

The government took precautionary steps against monetization suits with the 2011 America Invents Act. Its primary purpose is to deter plaintiffs from combining defendants into a single lawsuit (Cusumano, 2013, p.15). Prior to the act, monetizers pursued this route because it cut costs and increased efficiency. In other words, a monetizer could group as many inventors as he wanted in one filing. In the end, the plaintiff, known as the patent troll, receives more money in the settlement than if he filed individual suits. Usually, a troll will settle between \$100,000 to \$250,000 *per* defendant (Mullin, 2016). Now, if the troll sues several patent holders at once, his earnings will be greater than if he had to pay the filing fees to pursue each defendant individually. Most of these group lawsuits occur in the Eastern District of Texas, which is explained later in the research. The 2011 act makes an attempt to prevent this litigation strategy. It only permits trolls to join defendants *if* they share the same products or patents. The likelihood that they do is minimal, which makes this a good deterrence of monetization suits. The premise for this decision is that it is impossible for differing products to violate different patents on the same grounds (Macri, 2015). Despite this act, monetizing trolls have not diminished their pursuit of patent holders.

The Eastern District of Texas (E.D. Tex.) is notorious for its abundance of monetizing trolls. After the America Invents Act, the district witnessed a 166% increase in filed patent cases (Nazer, Ranieri, 2014). Since trolls face heavy penalties if they attempt to sue multiple parties at once, they began suing individually. Unfortunately, the act was not a remedy to this flaw in the

system. Furthermore, during June 2015, the majority of all patent cases were filed in the E.D. Tex. In just one month, Oberalis LLC filed 50 cases and Genaville LLC filed 25. Eleven other major trolling companies filed a total of 173 in the district. In Comparison, there were only 15 patent cases filed amongst 8 other districts (Nazer, Ranieri, 2014). The E.D. Tex. accounts for 95% of all patent cases started by trolls, and 44% of all the nation's patent lawsuits (Mullin, 2016).

These statistics highlight why patent trolls gravitate towards the E.D. Tex. Despite current legislation, they are more likely to win a case here than in any other district. Figure 3 highlights this reality. Since 2014, other districts have seen a relative decrease in the number of patent cases filed. The E.D. Tex., on the other hand, nearly doubled, with a 165% increase from 522 cases to 1,387. Figure 4 reports how companies prefer filing in the E.D. Tex. versus other districts. Of the 15 listed, 13 companies file all their cases in this one district. Finally, figure 5 compares the E.D. Tex. with 5 other districts. It shows this district's popularity, exemplifying that if you are to combine all the other districts together, the total number of their filed cases does not even compare to the astronomical number filed in the E.D. Tex. alone.

Patent monetizers succeed in the E.D. Tex. for several reasons. A major reason is because presiding judges rarely accept summary judgements. According to Federal Rules of Civil Procedure Rule 56, "Defendants have a right to file a summary judgement motion and to have that motion decided" (Redondo, 2014). A defendant makes a summary judgement if he believes allegations of supposed infringement are not valid. In the E.D. Tex., judges actually deny defendants this right, even though they are guaranteed it under a Federal rule. These judges approve less than 18% of summary judgement motions compared to upwards of 31% granted nationwide (Redondo, 2014). As a result, defendants are dissuaded and sometimes not even offered

the opportunity to fight allegations made against them. Therefore, the defendant has even less defense in the courtroom and incurs greater costs by continuing to fight.

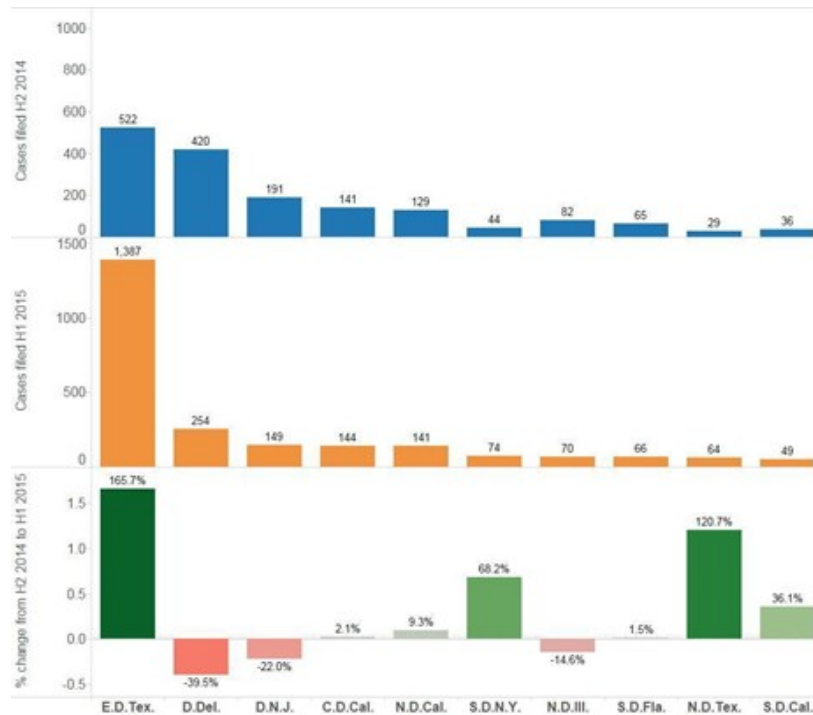


Figure 3: Howard, B. (2015, July 14)

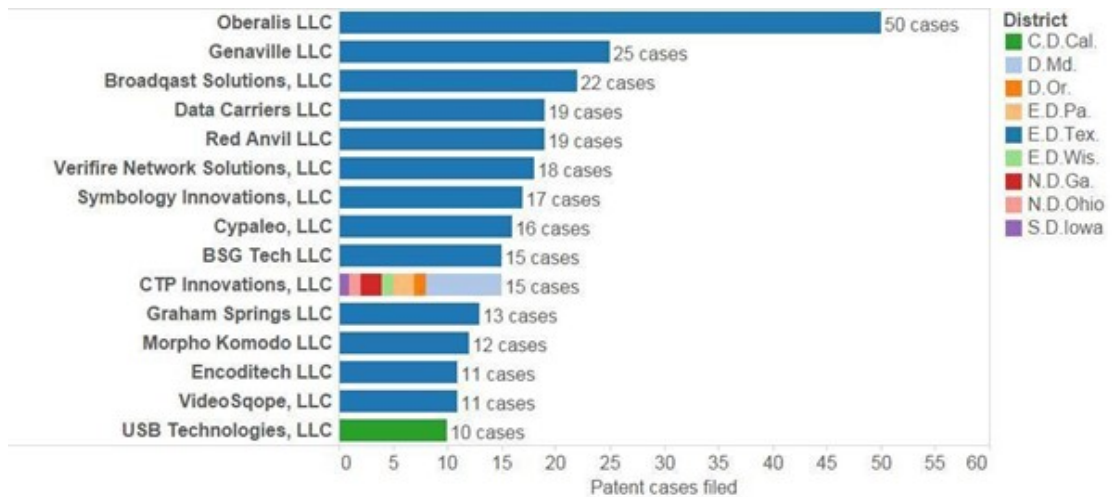


Figure 4: Howard, B. (2015, July 14)

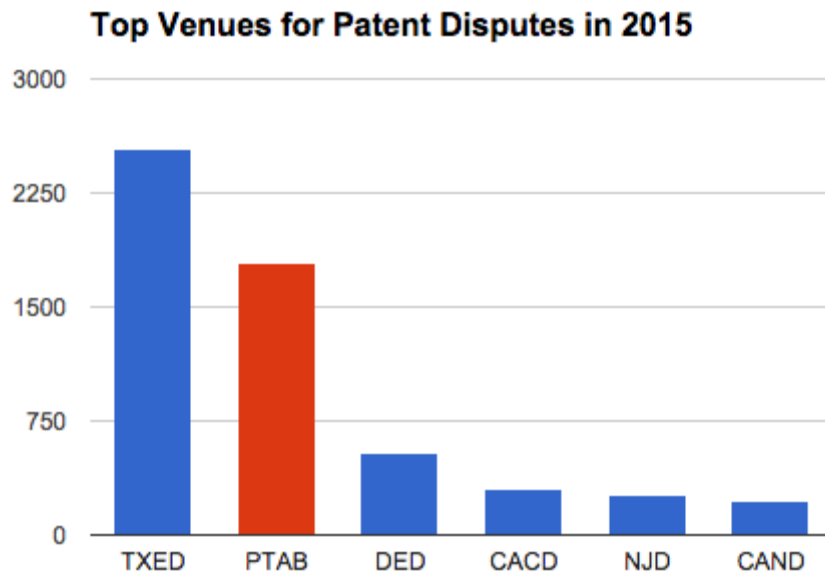


Figure 5: Mullin, J. (2016, January 5)

In some cases, the defendant can fight for a summary judgement, but it becomes another long and costly process. The district purposely elongated the procedure in order to dissuade defendants from choosing this route, therefore benefiting patent trolls. Even if the defendant takes on these extra costs and the judge accepts the motion, there is no guarantee that the defendant will win the case (Redondo, 2014). Regardless, the defendant does not benefit when he is brought to court in the Eastern District of Texas. Only a big data analytics system can provide a defendant with the information needed to fight a lawsuit in this district.

With regards to summary judgements, if the defendant decides to pursue this motion, then the time in court is extended, which increases litigation fees. This can turn into a waste of time, money, and efforts, especially since the defendant has no way of predicting the case outcome and whether or not his summary judgement will be granted. This can increase settlement pressure. When the judge does preside over the case, the defendant is more willing to pay the monetizer

whatever it demands because the defendant does not want to sit in court any longer. The defendant will do this to avoid wasting more time and money on a case in which the patent may not even be worth half the cost of damages incurred during the suit. These are some of the hurdles that deter a defendant's success and attract monetizing trolls to file lawsuits in the E.D. Tex.

Judges in the E.D. Tex. are given more patent cases to rule over than in any other district. For example, in 2013, a judge would settle 900 patent cases on average. Many of the patent holders return to court with similar, if not repeating, cases. Because of this repetition, the judges are more likely to make the same verdicts. Some lawyers will speak before the same judge, and therefore can take advantage of knowing how the judge will sway. They can alter their strategies to appease to a judge's bias (Redondo, 2014). Small, independent defendants do not have the same advantage, but would benefit with a big data analytics system that helps them make more informative legal decisions.

Whereas patent monetization can sometimes replace costly lawsuits with cross-licensing agreements, this solution will not resolve the major issues we noted in patent overlaps and the system as a whole. With the implementation of a big data system, smaller companies can better prepare themselves should trolls sue them in any district, but especially in the Eastern District of Texas, where decisions rarely favor the defendant. Patent holders can use the features of a big data system in order to make informed and educated decisions in cross-licensing agreements and monetization settlements. There is also the possibility that a big data system will rid the industry of patent trolls once and for all. Analytics can expose the truest intentions of the trolls, that being for individual profit, and demonstrate that there is no proof or merit in suing a patent holder for infringement. Furthermore, big data analytics would show that the patent trolls are not even suing on grounds of infringement, that in fact, there are no overlaps in the patents in question. However

without such affordable and universal services, small companies are sentenced to continuous suits and settlements in which their pockets are emptied and their patents are taken away from them.

3.8 Apple v. Samsung's Effects on Current Calls for Change

Apple v. Samsung was decided by a jury comprised of smartphone users and patent holders. Researchers question if this knowledgeable jury did make an unbiased decision. Despite its educational background in technology, the jury struggled to understand the legal jargon and instructions. That is one reason that experts believe the jury reached a hasty and biased decision (Burney, 2015, p.90).

Member of the jury and patent holder Ilagan Hogan admitted that the jurors skipped over several key questions because they did not know enough about the topic nor understand the instructions. One question the jury ignored was whether Apple patents were considered prior art. As previously mentioned, prior art is a key component used to validate patents in the preliminary stages of the patent acceptance process. To gloss over such a key factor is a large mistake on part of the jury (Burney, 2015, p.93). If the jury had been knowledgeable enough in legal jargon and all the characteristics regarding patents, then it may have changed its verdict. It might have recognized that Apple in fact copied existing patents. In this regard, a jury educated in legal concepts and technology is required in order to reach a non-biased, thorough decision on patent related cases. Ideally, a big data system would fill in the gaps of knowledge for the jurors, thus easing the decision making process. A big data system would act similarly to the patent examiners who accompany and aid judges in the Japanese and Chinese courts. Major factors, such as prior art, will be fully accounted for and hasty decisions can be avoided.

In the *Apple v. Samsung* case, when the jury submitted its decision to the overseeing Judge Koh, it was immediately returned to them. Judge Koh found there were careless mistakes in monetary damages and other aspects of the suggested settlement. The errors are a direct result of their hastiness to reach a verdict and to skip important questions. The complicated jargon and instructions deterred the jurors from seeking additional sources to help them better understand the issue, and so they resorted to their own bias to decide on the matter. As previously mentioned, such prevalent cases ought to be dealt with by a technologically and legally educated jury, whose “brand loyalty may [not] cloud the jury’s judgement” (Burney, 2015, p.94). This is a flaw that can be mended with a big data system that educates the jurors throughout the decision making process. If all the facts are laid out and legal jargon is simplified, the jury will not have a reason to resort to bias.

The big data system should serve an initial purpose of preventing these cases from being brought to court. However, if this fails and a jury must preside, the system would be advantageous to the jurors. Just like Japan has IP professionals aiding judges, a big data system would be the jurors’ aid. Complicated legal jargon and instructions can be translated into simpler terms. The questions and case details also can be simplified so that the jury would not feel pressured to skip important decision factors. A big data system would improve court efficiency, thus deterring mistakes from occurring in the decision making process.

Understanding how different countries presided over *Apple v. Samsung* can help us understand what changes should be made to the U.S. system. Japan’s IP High Court ensures a level of increased efficiency so that the judges who are “armed with complete knowledge of technical details and the patent application examination procedures...clearly understand the technology and the law when deciding the scope of a patent claim and whether an infringement has occurred”

(Burney, 2015, p.97). On the other hand, U.S. judges do not receive adequate professional guidance, and often have to rely on their own interpretation of the technology in question. As previously mentioned, U.S. judges and jurors can be professionally guided with the use of big data analytics.

Judges ought to have a blend of technical and legal expertise if they are expected to preside over patent suits. Over the long run, the easiest way to achieve this is with a big data system which will take the place of a legal assistant. Big data can store mass amounts of information and then apply algorithms in order to create trends, highlight patterns, defragment complex information, and predict future trends.

There are several reasons physical changes to the U.S. court system are not feasible. For example, we can hire law clerks with a background in technology and patents. However, this would be a costly change and require excess time to find qualified candidates. Creating patent-specific courts is also out of the question. They would most likely be instantiated in the Northern District of California, where *Apple v. Samsung* took place, and the Eastern District of Texas, where the majority of patent cases are filed. As a result, these districts may yield heavier caseloads and run into resource depletion while other districts experience an imbalance between assigned cases and available resources. Finally, there would be an overall tipping of the scale in the judicial system from the poor resource distribution. Solving such an imbalance would require creating new courts that specialize in patent law and spreading them out across many districts. But even doing this would require altering district lines, making costly changes to jurisdictional statutes, and hiring and training of new judges (Burney, 2015, p.104). These are the reasons we cannot pursue physical change to the U.S. patent system.

The above alternatives are expensive and timely, offering no guarantee of efficacy. On the other hand, big data analytics that is available to lawyers, patent industry workers, and now judges and jurors can help avoid the discussed implications. There would be no need to change the physical judicial system, nor violate constitutional rights or require new amendments to be drafted. The only change would require building a big data system, off a platform of existing systems such as Lex Machina, and educating patent industry practitioners how to take advantage of this service.

3.9 How Big Data Analytics Might Have Changed the Outcome of Apple v. Samsung

Access to a big data system has the potential to change the landscape of *Apple v. Samsung*. On the most fundamental level, when Apple first sued Samsung, Samsung could have used big data analytics to identify if Apple's allegations were in fact valid. If they were, then Samsung would have reason to worry when brought to court. Even if Apple was truthful about Samsung's infringements, Samsung could have used the predictive analytics component of big data to identify what the outcome might be and what were the best preventative strategies to pursue in court. If, on the other hand, the analytics demonstrated no evidence in favor of Apple's claims, then the case could have avoided court, saving millions of dollars for both parties.

When in court, big data could have helped both parties with extensive research in case development. The early warning component would not have benefitted either party by this point, but the risk analysis would. The risk analysis would demonstrate trends, patterns, and anomalies based off relationships it could identify between all the stored structured and unstructured data. Finally, the insights gleaned from this type of analysis would better help the lawyers formulate the most concrete evidence-based responses to allegations made against them. If the two companies

could present more clear and concise cases and evidence that did not require back and forth debate, perhaps the presiding jury would have had less difficulty breaking apart the evidence and to reach a conclusion. In addition, the case could have been settled in less time.

On the judicial level, big data analytics would have helped judges *and* jurors. Analytics can perform searches and provide information within seconds, saving the judicial system the 4 years that Apple and Samsung sat in court. When the parties were in court, the decision and evidence searches were conducted without the use of technology. Sifting through data about existing and historic patents to find clues about potential overlaps is a timely procedure. In an applied big data scenario, the judges would have access to limitless resources and would not have to pick their own brains to procure shallow assumptions. Furthermore, the jury would be more educated on legal jargon, decision questions, and details about the case and patents. There would be no justification for the jury to skip important questions. The fear that the jury would make a biased decision would be diminished. If the jurors have the facts laid out before them in simpler terms, there is no reason for favoritism and loyalty towards one company to weigh into the decision process.

Additionally, patent trolls, who became more rampant since *Apple v. Samsung*, can be stopped. The courts would use big data analytics to quickly identify if infringements actually exist. If they do not, then the trolls have no right to sue, and therefore will lose some of their influence in the industry. Soon enough, they would have little to no power in court, and will only be wasting their time and money by pursuing patent holders who are protected by big data analytics. Small startups and other companies would have an extra layer of protection from big data in the event that they must defend themselves against these monetizers, namely in the Eastern District of Texas. The early warning system and risk analysis would be most beneficial to these parties. It would

warn them in advance if someone is seeking to sue them, and then they would have more time to pursue preventative measures and create a solid and educated defense.

Furthermore, with big data, we can make it more difficult for trolls to sue defendants in groups. The judges would also have less reason to favor plaintiffs, specifically in the Eastern District of Texas. This is possible primarily because judges would now have all the required support and facts about the case and supposed infringement. In this regard, big data acts as a type of check on the judicial system and the verdict of judges and juries.

Overall, the ideal big data system would have saved time, money, and efforts during *Apple v. Samsung*. In the long run, it would reform the current issues outlined above in the patent industry. Implementing a universal big data system would be less expensive than applying physical changes to the legal system, therefore making it a more reasonable option that is also available to all patent industry practitioners. Big data has the potential to decrease the amount of smartphone wars that fill up the courts, especially in the Eastern District of Texas. In so doing, a big data system will not diminish innovation or the competitive nature of companies in the global market; it will merely make companies and inventors more precautionous and proactive in their decision making and pursuit of new ideas.

3.10 A Conclusion on Healthcare's Big Data Application and It's Correlation to The Patent Industry

All the previously discussed information is necessary to understand how big data analytics can be applied to the patent industry. Big data's application in healthcare is a concrete precedent, primarily because real-time information and decision making is necessary to both industries. Doctors and researchers rely on constant updates to save money, time, and lives, just as lawyers,

patent holders, and patent examiners rely on constant updates to save money, time, and ideas. Any mistake in predictions can lead to astronomical costs incurred by the healthcare industry. Big data yields a smaller percentage of error in predictions based off historic information than most other analytics. Big data has this success rate because it is intentionally designed to support the task of working with complex data and resolving anomalies that may arise as a result of storing mass amounts of information.

In parallel, the patent industry relies on real-time updates to grant patents, as well as to develop cases and preventative strategies. The massive influx of patent applications may be overwhelming to examiners. Because of this stress, examiners are more likely to overlook patent details, which results in patent overlaps and future litigation. This can be avoided with big data analytics, just as similar obstacles were avoided with big data's application to the healthcare industry. Patent applications can be analyzed by a computer instead of an individual, which would make the examiner's schedule more open to review "homerun" patent applications. If overlaps are discovered, the system would return a warning to the examiner so that the filer can be notified and the patent can be denied.

Numerous examples of this feature in healthcare's big data application were discussed in this research. The most similar example is the use of big data in the emergency room. ER doctors are under high levels of stress when dealing with the influx of emergency patients. With big data, electronically transmitted health records are presented to the doctor instantaneously. The EHRs even offer suggestions to the doctor about what treatments may or may not work for the incoming patient based on the patient's medical history. Then, the doctor does not have to spend valuable time manually searching the patient's health records or making assumptions about the best

treatment. Decision making functions are served in a big data application to simplify the doctor's job, alleviate some of the stress, and help the patient in a more timely manner.

Spending valuable time sifting through health records is analogous to reviewing every application that comes into the patent office. A "homerun" patent is parallel to an incoming patient in the ER. Automated processes to enhance the decision making process will save time and money for both industries, innovative patents for the patent industry, and lives for healthcare. A doctor being informed about the most important details composing a patient's records, such as medicinal allergies, is similar to a patent examiner being notified of overlaps in existing patents. The needs in both industries are very much alike, with regards to access to real-time information, early warning systems, and risk analysis. Therefore, we use the healthcare industry to relate big data requirements to the patent industry.

3.11 A Suggestion for Big Data Analytics Implementation

This research presents support to make Python the programming language for the ideal big data system. Lex Machina implements Python, which acts as a viable platform for our universal system. The contending language was R. However, the main reason R should not be used is because it "is more about sketching and not building. You won't find R at the core of Google's page rank or Facebook's friend suggestion algorithms. Engineers will prototype in R, then hand off the model to be written in Java or Python" (Nicolauo, 2014). This demonstrates that Python, rather than R, is supported more for the actual programming of applications. Professionals will revert to one of the two more popular coding languages, Java and Python. They are more versatile and supported on different platforms than R is. Python can easily achieve two of the major functions for this big data application: predictive modeling and risk management. This makes it a

good option for implementation. If desired, the building team can sketch out the basics for the big data analytics algorithms using R, then transfer it to Python for actual implementation.

The costs to produce a universal system are not as astronomical as one might think. This is due to the fact that there is an abundance of learning libraries to facilitate Python programming for beginners and experts. Building the system may take more time than anticipated, especially if pursuing several key project development steps. The objective to build a big data analytics platform for all patent industry practitioners must be defined. Next, a team must be composed of individuals with different backgrounds in business, software, and analytics. Team diversity simplifies data and project complexity since more questions are likely to be addressed during the development process. Costs can be diminished in the next step, which is establishing key performance indicators and setting goals and deadlines (Patel, Sundararajan, Marangos, Zimmers, 2016, p.38). This is an important step in any project because we do not want to waste resources or time. Should the team not meet its goals by the agreed upon due dates, then it risks increasing time, money, and efforts during the project cycle.

The next crucial step involves working with the data that will potentially be stored. The first thing is to create a data dictionary. A data dictionary is used to store descriptions and definitions of the type of data items that will be stored, what the values may contain, how the data fits into the functions that will be built, and what the data means in real world application. Then, once the team outlines the different components as well as their purposes, they can create a master data file that acts “as the data repository for [future] analysis” (Patel, Sundararajan, Marangos, Zimmers, 2016, p.39). As the team collects potentially useful information, it may uncover trails that lead to more information and metadata. The master data file will hold all of these links. Sometimes, the master data file will come in the form of a traditional relational database.

Generally, when creating a master data file, information is combined into a flat file. A flat file contains records of information that have no structured interrelationship. In the later system development stages, the information in the flat file can be separated based on relations to each other. A challenge we will be presented with is increased complexity in the flat file as more information is added to it. Later in the process as we use different analysis procedures, these complexities will be diminished. This is because duplicates of information can be deleted and anomalies, such as spelling errors, can be resolved.

Developers will come in contact with some issues when accessing a flat file. These issues include information duplicates and data inconsistencies. Although resolving such hurdles is common and timely, it is a crucial step in order to create a valid and efficient system. Removing data redundancies, such as copies of an article that appears more than once in the file, is an important obstacle to overcome before the data can be manipulated through analytics. Once complete, the team can further identify correlations and define relationships.

The next step is to perform a basic mining analysis, such as importing the flat file into Hadoop. Brought to life through visualized trends, the team can see relationships come to life between the different variables and data points. Furthermore, relationships the team did not even think of will be illuminated. One difficulty that may arise from this process is increased data complexity. The developers may be presented with common issues such as duplicated data points and/or careless mistakes in spelling or numbers. In some instances, this can prevent an analysis from running, but in most cases, the developers can capture the issue, resolve it, and get the analysis running to perform its task. The only major barrier this presents to a team is the loss of valuable time and possibly money, if the team has to hire external data scientists to resolve any discrepancies and get the data mining analysis functioning properly.

The next major step is to develop the algorithms that will achieve the most important components of the system: the semantic search, risk analysis, and early warning detection. The semantic search involves “the application of natural language processing, computational linguistics, and text analytics to identify and extract subjective information in source materials” (Batinca, Treleaven, 2015, p.90). Suggestions for implementation do exist, however they are narrowly defined for handling structured text. It is not as effective with non-text data types, such as image and video. Lawyers and other industry professionals may require the compilation and analysis of diagrams, images, and videos for a case. They do not have the time to sift through hundreds, if not more, of these data types and create trends about them. A big data system should accomplish this function because not only will it perform a basic comparison of these non-text data types, it will take the analysis a step further by combining these findings with those regarding other data types and predicting trends. A big data non-text data analysis accomplishes far more than a computerized system that merely compares diagrams, images, or videos.

Therefore, the team ought to develop an algorithm that also analyzes these different data types. When developing these algorithms, the team might be presented with the barrier of having to program with more than one language. For example, machine learning is more often used in sentiment analysis, which means the team might have to learn how to apply machine learning implementation to a system that is predominantly programmed in Python. A sentiment analysis is a means to identify and categorize opinions that are expressed in text and even videos and images. Machine learning facilitates speedier and more automatic processing of large collections of data. Another barrier appears in the task of manipulating non-text data types. Programmers will have to access the Python learning libraries in order to better understand how to approach this difficulty. Compared to other programming languages, Python is more compatible with sentiment analysis.

This is possible because of the existing tools and classifiers that further the programming learning process. A sentiment analysis algorithm in Python and machine learning can be used by lawyers to better understand the underlying meanings of reports, preceding cases, product and patent reviews, and potential witnesses. This would act as a truth test, therefore aiding in the development of more accurate cases and witness testimonies.

Another aspect to consider is whether the developed algorithms should be rule-based or expert-based. The question is, can one set of rules be applied to all patent practitioners' needs, or should the rule sets be defined by experts that are flexible to the different needs in the system? For example, there will need to be algorithms that seek out keywords or metadata in the stored data. Whether they are rule-based or expert-based will impact our ability to quantify the required resources. If the algorithms are expert-based, a requirement will be to hire additional experts to define individual algorithm rule sets. With regards to our research, expert-based algorithms may be the best option, because they can be more specific to meet all the needs in this industry; the experts can make separate decisions for patent examiners, patent holders, lawyers, judges, and jurors. Should we recommend rule-based, then all the rules must be followed and there is no opportunity for change. In other words, rule-based is very black and white. Expert-based can still be standard with a degree of customization, especially since not all of the same rules will apply to the differing patent practitioners. Rule-based can be easier to monitor and maintain, because it is only one set of rules. However, it will be harder to define and implement since the legal industry is not a one-size-fits-all industry. On the other hand, expert-based can be accomplished more efficiently and easier, but its need for constant updates will require continuous hiring of experts to review and possibly modify the system. Therefore, although the costs may end up equaling each

other, expert-based is the best option because of its flexibility to comply with the differing needs of all the patent practitioners.

The suggested requirements for the ideal system are outlined at the end of this section. One must understand that there are many types of data that need to be analyzed. This includes, but is not limited to, historic datasets, real-time feeds, raw data, and cleaned data. Most of these are text, which may require uniform reformatting. Reformatting becomes a simple task when achieved through services like Google's Refine3, conveniently coded in Python (Batinca, Treleven, 2015, p.101). Unfortunately, manipulating unstructured data is not as easy. This requires additional algorithms to transform the data through preprocessing, tagging, and parsing. One can use a data-mining algorithm that searches for patterns amongst data points to achieve tagging. It is outside the scope of this research to delve deeper into the different algorithms used to transform data into a uniform type.

It is also important to understand what additional resources are required to maintain the system. Back-end services are the more obvious, providing for basic connectivity between the platform, data providers, and users. It permits multi-user environments, "prefab" models, and APIs. APIs are condensed learning libraries that will come hand-in-hand with a simplified graphic user interface, also known as GUI. Users can customize their IDE, Integrated Development Environment, to make it more geared towards their specific preferences and needs.

It may also be beneficial to build an aggregation database into the system in order to filter and restructure incoming data. Otherwise, to complete this task, we may need to outsource the data. This requires more money and time in the long run. Integrating an aggregation database, on the other hand, makes it easier and faster to store, extract, and manipulate mass amounts of data. It will further improve our ability to perform complete risk analysis and predictive models based

on historic trends. Pursuing the route of developing an aggregation database will increase costs in the short run. We will need to hire professionals with experience in database modeling, and more time will be partitioned for building this database. This can delay the use of the actual system, but the long term benefits discussed above outweigh the short term costs.

We also take into consideration purchasing hard drive space and memory. Whether through the cloud or actual hardware, it is a necessary cost to maintain and store mass amounts of information. The cheaper and more popular route is cloud storage, but there comes a greater security risk. The protection of your data in the cloud is only as good as the security measures built into the cloud. Those storing information have no control in security customization, *unless* the cloud storage is built specifically for the user. Seeing that we would like to build a universal big data system, we should have no problem investing in building a customized cloud-based system. Companies such as Apple and Google maintain these customized and security enhanced clouds. We suggest staying away from a local hard drive because, as of now, there are millions of patents around the world. Storing all this information would require terabytes of space, all of which cost more than a cloud that can be easily expanded whenever requested. Furthermore, in the long run, more money will be saved by building a custom cloud than purchasing additional hard drive space whenever more information needs to be added.

Because of the way technology constantly changes, it is difficult for us to present a concrete plan for how to exactly develop a big data analytics system. At the end of the day, building a big data analytics system requires careful consideration on the part of the development team and all the system's potential users. We provide a table of required features at the end of this section for a development plan, of which can be accepted in full or in part. The team will need to conduct testing of the system before it is made public for the use of IP practitioners. This may delay the

availability of the service and increase costs, but it is a necessary step in the development process. We want to ensure that no difficulties arise when the system is used constantly by thousands of people. The team will also have to incorporate into the budget the maintenance cost. This task may include establishing a secure yet accessible network and ensuring that the stored information is up to date and accurate.

Our suggested plan will meet the needs for predictive analytics, risk analysis, and early warning systems if enough time and resources are appropriated for such a project. The benefits, as discussed in the entirety of this research, will prove most optimal when a custom-built big data analytics system is instituted for all patent industry practitioners to use.

The features outlined below are required in order to develop the predictive, early warning, and risk analysis features of a big data analytics system. No one requirement is too difficult to accomplish, however some will take longer to complete than others. The ability to complete a requirement on time will depend on the budget and amount of knowledgeable experts that can be hired. The level of difficulty to complete a feature was determined relative to all of the other features. Some features may be considered easier than others because there are more resources available to facilitate development. A feature can also be considered easier to implement because of its reliance on existing data and services. A feature may be considered difficult because there are not enough resources available, and it is more likely that knowledgeable experts will need to be hired to accomplish the feature. In this respect, costs of production, including time and money, will increase.

The timeframe at which any one of the features can be accomplished is difficult to quantify. Therefore, we predict the possibility that the feature will be complete within the near, intermediate, or long term. If a feature is in the near term, there are ample resources and experts to accomplish

the feature quickly. If a feature is in the intermediate term, then there may not be enough resources and experts to presently tackle the project. The difference between this and the long term is that in the intermediate term, there is discussion and development underway. An example of a feature in the intermediate term is Lex Machina’s system availability for commercial, bankruptcy, and employment law (Lex Machina, 2014). Such a feature will not be accomplished in the short term, however it will be available in the intermediate term because production is already occurring. Finally, a feature that falls into the long term will not be available for an unquantifiable amount of time. Furthermore, a task gets categorized as long term because there is no research and/or development, nor do we yet know how to solve and accomplish the feature. With respect to our big data analytics system, none of the features will be long term.

Below is the table outlining the required features for our big data analytics system. The second column is used to demonstrate the level of difficulty to create the feature. The third column demonstrates the relative timeframe that the feature can be accomplished.

Table of Required Features

Required Feature	Level of difficulty considering current technology and resources	How soon can the feature be met?
Data Encryption & Enhanced Security Features	Somewhat difficult: this is an important feature for any cloud storage system, and companies such as Amazon and Google already implement this. However, making it secure and difficult to breach presents a challenge to companies new to the field.	Intermediate term: This will require more resources (money, time, and knowledgeable experts), so it will take longer to implement than if the developers went with a pre-built cloud system. This is already evident in healthcare, because of HIPAA Compliance Regulations. As a result, healthcare’s implementation is a good foundation.

<p>Data Redundancy Check</p>	<p>Somewhat difficult: this is very important for the analytics features to run quickly and efficiently. Without it, an analysis may run and bring in the same article (for ex.) more than once, because the data storage was never checked for repeating pieces of information. This wastes time, money, and efforts.</p>	<p>Near term: This will be one of the latter steps, but should be completed before mining, early warning, risk, and predictive analyses. This feature is still in the near term because of the availability of resources and knowledgeable experts. It will be completed later in the process because it takes more time to process existing data than it would to complete other required features.</p>
<p>Python Programming Implementation</p>	<p>Easy: extensive learning libraries and pre-built programs/algorithms are open source, which makes for a responsive and active community. Lex Machina is built on the Python Pyramid Platform, which can be considered the foundation for this system.</p>	<p>Near term: This will be the first feature to accomplish. Setting the programming language is crucial to be able to write the algorithms and platforms for the different user needs.</p>
<p>Machine Learning Incorporation for Semantic Search</p>	<p>Somewhat difficult: semantic searches are crucial to make predictive trends that combine different data formats. Finding experts who know how to write a semantic search in machine learning is the biggest challenge.</p>	<p>Intermediate term: This will take some time to develop because we must find experts to program in machine learning for semantic searches and sentiment analysis. They also must be able to incorporate this into the Python foundation. This is not long term because there are some available resources and experts.</p>
<p>Cloud Storage Space</p>	<p>Easy: there are many available cloud services that can be purchased. One has to be chosen that is not overly priced, can be easily expanded at a reasonable cost, and permits customization for security and encryption.</p>	<p>Near term: This is one of the first steps that needs to be accomplished. The system requires a place to store all the data and have quick access to it.</p>

Data Conversion	Difficult: this requires a conversion program that takes into account all possible variations of data inconsistencies. This is required since the users need to be able to analyze all different data types/formats (images, videos, articles, court cases, etc.).	Near term: This should be completed once the Python language is implemented and a cloud system is purchased or built. It will require more money, time, and experts to create a conversion program, but is necessary.
-----------------	--	---

Figure 6

3.12 Big Data’s Connection between Healthcare and Patent Litigation, and Apple v. Samsung Revisited

As mentioned throughout this research, a bridge is built between the healthcare industry and patent litigation through the application of big data analytics. Both industries require proactive planning and analysis, which makes them very similar. They require real-time access to information to reach decisions that benefit industry professionals. Big data analytics can achieve three of the same features needed in both healthcare and patent litigation. These features include predictive analytics, risk management, and strategic planning in the form of early warning systems.

Big data capitalizes on these features to provide benefits in the healthcare industry. Time sensitive barriers are eliminated. Predictive analytics provide benefits in pharmaceutical research and development, as it helps foresee what drug development ought to be made in response to epidemics. Big data’s risk management capabilities help reduce fraud and poor spending of funds in hospitals (Tomasco, 2009). It also helps with allocation of doctors and other resources based on patient density. And finally, among the many different ways strategic planning aids healthcare professionals to prepare for epidemics, it also is advantageous in understanding the effect that drugs have on different patients, and the chance that a patient will survive a treatment specific to

his medical needs. Strategic planning also comes in the form of reliance on Electronic Health Records, which aids different facilities to take care of new, incoming patients more efficiently and effectively. Other applications of these big data features are detailed in the Literature Review and Analysis.

Just as the three major features of big data solve problems and breach real-time barriers in healthcare, they can achieve similar success in the patent industry. The specific application of big data's three features to *Apple v. Samsung* would have resolved the underlying issues in the patent industry. Revisiting *Apple v. Samsung* and applying big data's features exemplifies the bridge we seek to build between patent litigation and big data. In this case, predictive analytics would have illuminated the direction the case would take in court. It would have prepared both parties for the astronomical fees they would incur during the 4 years of courtroom proceedings. Big data's risk management would have helped both parties allocate resources more efficiently and perhaps would have urged Samsung to bring the case to another country, such as China or Japan. In China, it would have been cheaper to pursue, and the legal system might have demonstrated more favor towards Samsung versus Apple. This is because China's IP specific court has educated judges in intellectual property and patent law who take into account *all* the factors of a case. No question is left unresolved. The decision might have been different, resulting in a smaller settlement payout than in America. Based on some facts from the case, foreign legal systems may have deemed Apple more liable for infringement than Samsung.

If this big data feature was brought to Japan's patent system, the decision in Japan would not have changed. Big data would have no major impact on the *Apple v. Samsung* decision because Japan already has a specific court for IP law. This country focuses its resources on developing the IP court, which is led by educated judges and patent experts. Therefore, Japan would not even

require a big data system, because the system is not as flawed and uneducated as the U.S. intellectual property judicial system. The Japanese court is precise in its decision making and is unlikely to oversee important case details like the American jury did for *Apple v. Samsung*. America, however, does not have the same resource allocation in its judicial system, and therefore the best option to reform the flaws of the patent system is to develop a big data analytics platform.

Finally big data's early warning and risk management functions might have helped Samsung avoid a lawsuit in the first case. Courtroom litigation would be avoidable and the two parties might have had the opportunity to alter their patents so that they did not overlap. With the application of big data's three major features, *Apple v. Samsung* could have been resolved from the get go. Such real-time application and proactive planning would save money and resources, just as money, lives, and other resources are saved daily in healthcare because of big data analytics.

Furthermore, we can revisit the impact of these three features on different facets of the patent industry. Big data's predictive analytics would help a firm or company foresee the direction an existing case may take in the courtroom. Risk management helps the involved parties weigh the costs and benefits of pursuing litigation in the courtroom to see if it is feasible or if they should pursue negotiation outside of the courtroom, such as cross-licensing agreements. And, big data's early warning capabilities will help notify patent holders if they may be sued down the road. Patent examiners can use this strategic planning characteristic to determine if there will be overlaps in existing patents and incoming patent applications. Finally, judges and jurors can use big data's plethora of stored information, patterns, and trends to discern the details of a case and better understand legal jargon and instructions.

As one can see, the barriers of real-time information access and saving time, money, and patents in the patent industry are as easily eliminated with big data's three major features as they

are eliminated in healthcare. Both industries require proactive planning and real-time application, which presents big data analytics as the best and most optimal solution for both. Since healthcare already institutes big data analytics, the patent industry should do the same with this research and its suggestion for implementation.

4 Conclusion

4.1 Research Contribution

This research was conducted with the intent of shedding light on an existing gap between patent litigation and big data analytics. Minimal research exists about the topic, making this contribution a good foundation for understanding the gap and conducting future studies. This body of knowledge can have positive effects on all the fields addressed in this paper, including patent litigation, patent examination, big data analytics, computer science, healthcare, predictive analytics, and business intelligence.

The patent litigation field reaps the most obvious benefits, as it is the subject of this research. The concepts discussed in the Analysis illuminate the issues tainting the patent litigation system. Prior to this research, it was possible that professionals in the industry were not aware of the threats and cost of having patent trolls and monetizers in the system. They were not aware that if *Apple v. Samsung* was resolved in another country such as Japan or China, the outcome and impact on smartphone wars would have been different. As previously mentioned, American technology companies have brought lawsuits to foreign countries because of less expensive filing fees and increased possibility of winning cases. They also were not aware that technology and analytics would be the flawed system's savior. Therefore, this research contributes intelligence to the patent litigation field, and plants a seed for reform in the form of implementing a big data analytics system.

This research separates patent examination from patent litigation because both require different processes and big data application. The patent examination field will experience vast benefits from this research, primarily since big data analytics can speed up the application process

without compromising on efficacy. This research sheds light on the apparent flaws in the entire patent examination process. It does not blame individual patent examiners, but rather the lack of regulation in the actual process. Because change to the legal system is timely and costly, involving amendments, laws, and bills, this research ought to shed light to the patent examination field that the best way to fix the current wrongs is to adopt a big data analytics system.

Big data analytics is another field that benefits from this research. What is presented in the Literature Review and Analysis demonstrates how advantageous big data implementation is for different fields. As seen, big data reaps success in the healthcare industry. This research may also prompt professionals in the big data analytics field to explore its application to different industries. Big data's predictive trends makes it an incredibly beneficial source for all types of users. The flexibility of big data makes it a dynamic tool that is constantly in research and development. The research presented can urge big data scientists to develop more algorithms, accessibility, use, and manipulation of mass amounts of data. The growth of big data does not stop here. Perhaps in the near future, professionals will discover other beneficial uses of big data that might not have been as obvious without access to research such as this.

The field of computer science benefits from this research as it demonstrates the wide applicability of the field's professionals to different businesses. Programmers, data scientists, and IT professionals are hired all the time to meet the technological needs of varying industries. Now, the notion presented is that computer science professionals can offer their services to the patent litigation industry. This research contributes to the computer science field by promoting the education of CS professionals to become more knowledgeable in developing algorithms and platforms that capitalize on big data's features.

The healthcare industry also can benefit from this research. We present healthcare as an example compared to an industry that lacks the same level of technological progress. This research, however, may further prompt healthcare professionals to see where they ought to be taking more advantage of big data analytics. In so doing, the current system can be improved and updated, more healthcare facilities can be connected through EHRs, and more medical health issues can be resolved.

Predictive analysis, although the core of success for big data analytics, will also benefit from this research. Similarly to the healthcare industry, professionals in the field of predictive analytics can apply this research to understand where improvements should be made. We hope to see uniform and higher success rates across the board in all applied industries, including healthcare, natural disaster, global financial business, and legal.

Finally, the field of business intelligence can gain insight from this research. This research demonstrates that big data analytics can be applied to industries that may not appear to have anything in common at first. Although the healthcare and patent industries are very different, they both require and utilize technology and big data analytics for the same purposes. As business intelligence also relies on real-time access to information and early warning strategies, it can relate to a similar application of big data analytics as seen in healthcare and the patent industry. In this respect, any industry that requires similar features, access to mass amounts of data, and predictive capabilities will benefit from these research findings.

In this research, we proposed a framework of required features for the ideal big data analytics system. The table of required features can be viewed in the Analysis section “A Suggestion for Big Data Analytics Implementation.” This contribution adds value to the research field because our framework provides a higher level structure to facilitate understanding how to

implement big data analytics in the patent industry. Before our table of required features was created, there were no basic suggestions for building a big data system for the patent legal industry. Therefore, this framework can be utilized by future research in which a big data analytics system is built for patent industry practitioners.

4.2 Limitations and General Barriers for Implementation

There are existing barriers that must be considered before accepting implementation of a universal big data analytics system. They have the effect of increasing costs during the development stages. One barrier to consider is integrity of the stored information. Data integrity relates to the security, accuracy, and consistency of any data that is stored somewhere and accessed. If we use Python as the coding language, how can we ensure that the data will maintain its integrity? This is an important barrier to take into account because if the data loses its integrity, then it increases the risk of predictive analytics yielding a higher percent of error. For predictive analytics to be optimized, data must be secure, up to date, and free of any type of error. With regards to the security aspect of the universal system, it must be taken into consideration how individuals will be granted permission to add information to the system database. Can anyone who uses the service add to it, similar to Wikipedia? Or are a select few individuals permitted to add information and cases from around the world?

If the system is open source, then anyone can provide input and add information to the overall database. Although this is beneficial because it means the information will be constantly updated, it does increase security and integrity risks. Data can be added that is inaccurate, which would result in IP professionals creating reports and decisions based on faulty information. In some instances, this can be worse than not having access to the right amount of information.

Furthermore, a user can attach a virus to some of the inputted data. This can result in security breaches and threats, and a compromise on personal, and even government, information. This is a risk that heavily impacts all the system users, and therefore is considered a major barrier should the big data system be open source.

Another barrier pertains to the implementation of the system. We cannot quantify how many developers are required to build the platform. This is an aspect of project management that needs to be further studied. The maximum expenses allocated to the development team will dictate how many professionals can be hired and what tools can be used to achieve the ideal system. To put it into perspective, some of the required roles include analysts, visualizers, data collectors, data scientists, open source professionals, testers, programmers with experience in Python, security and network professionals, those with knowledge about the patent industry, and business professionals. A well-rounded A-team should encompass all the necessary roles to perform the project in stages in the most efficient manner. It might be beneficial to start this project by building subsystems geared towards the specific users in the patent industry.

These barriers can be overcome with a proactive team and proper project management. This research merely presents the barriers and suggests that they can be eliminated. It is not within our scope to outline exactly how the barriers can be eliminated, the difficulties in eliminating them, and what the definite costs for building and implementation consist of.

There are general limitations inherent in the suggested implementation. These limitations include, and were discussed in the table of required features, data and system verification, validation, and debugging, data integrity and conversion, redundancy, encryption, and security. Another general limitation is the maintenance of the system to ensure that it is the most real-time

as possible. Not all of these limitations are necessarily bad, however they do require more budget, time, and experts to be properly addressed.

Several limitations existed during our research. Because of a small budget and limited time, we could not build an actual analytics system to test. The major limitation while conducting this research is that there was not an abundance of available research and literature on big data analytics and its application to the legal industry. Because of this gap, we had to reach out to the healthcare industry, which is similar enough because of its need for real-time information, to learn from and create a clear line of sight between big data analytics and patent litigation. Furthermore, as a result of the lack of existing sources, this paper acts more as a foundation comprised of predictions and our thoughts on creating a bridge between big data analytics and patent litigation than a compilation of existing research proceeded by a hypothesis. In this respect, this is novel research that has little foundation in existing studies and applications. It is primarily a study composed of assumptions and suggestions for future research and implementation.

4.3 Suggestions for Future Research

Since the research we conducted is so new and creates its own foundation, there are several suggestions we can make for future research. The limitations we were faced with also make it easier to suggest how future research should be conducted to see more concrete results and bridge the gap between big data analytics and the patent industry.

Researchers can apply the findings in this study to the cases that may arise concerning technology patents. They can see if our predictions about big data's effects on *Apple v. Samsung* would be similar for other cases. This presents a more solid foundation to support the implementation and creation of a big data analytics system. Future research can also be conducted

in the actual development of a big data analytics system. Data and computer scientists can try creating the system in Python and apply the types of algorithms discussed in the study. Then, the scientists can conduct real-world application with patent law firms to test whether or not the system is as beneficial as we predict it to be.

There is value in continuing the research and concepts presented in this study, primarily because of its novelty. Little to no studies exist to bridge the gap between patent litigation and big data analytics. Therefore, this presents a good foundation for future research to develop the concept and suggestions we present. It is important that we understand the applicability and effectiveness of big data to improve business intelligence in more than just one industry. That is why we conducted this research in the first place, and what we hope will be continued in the future.

This research may be extended to encompass other forms of litigation. Perhaps benefits can be reaped in bankruptcy law. If individuals can foresee the future of their financial standing, then maybe they can prevent having to file for bankruptcy. Currently, Lex Machina is working on expanding its analytics to encompass bankruptcy law (Lex Machina, 2014). Another form of litigation that may benefit from big data analytics is general trial law. Regardless of the lawyer's specialty, if he is brought to court, it would be easier for him to access mass amounts of information that could potentially change the outcome of the case, predict how this one would unfold, collect the proper evidence and witnesses to counter the opponent, and favor the judge's bias. In general, the research conducted that pertains to the patent industry is broad enough to also be applied to other forms of litigation. There is potential for growth, acceptance, and use of big data analytics in every legal specialty.

The ultimate goal of this research is to lead to an actual implementation of universal big data analytics in the patent industry. The table of required features that we created in the Analysis

is a condensed framework that suggests how someone in the future can build the proposed big data analytics system. We outlined the high level requirements in the table, which act like a road map that a team can use for implementation. Researchers can explore this supposed framework and required features in the future.

We hope this research served the purpose of educating practitioners in the legal, data, business, and healthcare industries about the applicability of big data analytics to different facets of a business's objectives. This is where some of the greatest value from this study will stem from, and we anticipate that professionals and scientists will conduct further research without facing the same barriers and limitations as we did. Our goal is that this research will lay the foundation for the development of a universal big data system for use by all patent industry practitioners.

4.4 Final Summary

This paper achieved the goal of providing a basis for why and what should be done about implementing a big data analytics system in the patent industry. Because of the existing application in healthcare and the similarities between the two industries through proactive planning and real-time access to information, we can bridge this gap. Therefore, it is probable that benefits reaped in healthcare from a big data system will trickle into the patent industry. This research also utilized *Apple v. Samsung* as a case study to highlight the flaws in the patent system. We use the case to demonstrate how its decision and the future of smartphone wars could have been altered by the use of big data analytics. Finally, suggestions for implementation were made, detailing some of the higher level requirements, such as using the Python programming language, developing encryption to ensure data integrity and security, and reducing redundancy of stored data.

Our ability to predict the actual effectiveness of such a system is limited by a lack in physical implementation. We cannot yield definite benefits for lawyers, patent examiners, judges, jurors, and inventors who would use the system. In addition, we cannot predict how long and the cost to create and implement the ideal universal system. Despite such limitations, we can safely assume that a similar transfer towards a reliance on big data evidenced in healthcare will appear in the patent industry. With more advanced resources, time, and technology, future research can enhance these findings.

Scientists can build a prototype big data system to test with industry practitioners. They can further test the effectiveness of technological aids to jurors and judges presiding over courtroom cases. Studies about the speed at which lawyers produce stronger cases can be accomplished. With increased technological resources, researchers can analyze how far in advance an early warning system will predict the future of a patent case. Finally, researchers can study how many more patent applications an examiner can thoroughly study or grant with the aid of big data. As one can see, there is more to study about this subject matter. We succeeded in presenting the foundation for more research, but because of resource limitations, we could not further investigate the ideas listed above.

Predictive analytics reaps numerous benefits for its users. Lawyers, patent examiners, inventors, judges, and jurors can experience these benefits with the implementation of a universal big data system. Developing this system will resolve discrepancies in the patent industry, from an overabundance of patent applications in the USPTO to settling infringement cases outside of the courtroom and to eradicating patent trolls altogether from the playing field. Conflict between existing patents and future inventions can be avoided, costs and time would be reduced in the event that litigation *is* unavoidable, and the future and legal ramifications of smartphone and general

technology would be more predictable. The costs are far outweighed by the benefits, as witnessed in the patent industry's parallel: healthcare. It is only a matter of time before patent practitioners accept and pursue a reliance on big data analytics' early warning, preventative, and predictive capabilities.

Bibliography

EMC. (2015). *Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data*. EMC Education Services, I-Xvii. doi:10.1002/9781119183686.fmatter

Gupta, R., Gupta, H., Nambiar, U., & Mohania, M. (2010). Efficiently querying archived data using Hadoop. *Proceedings of the 19th ACM International Conference on Information and Knowledge Management - CIKM '10*, 1301-1304. Retrieved from <http://dl.acm.org.liucat.lib.liu.edu/citation.cfm?id=1871606>

Jee, K., & Kim, G. (2013, January 30). Potentiality of Big Data in the Medical Sector: Focus on How to Reshape the Healthcare System. Retrieved from <http://synapse.koreamed.org/search.php?where=aview&id=10.4258/hir.2013.19.2.79&code=1088HIR&vmode=FULL>

European Commission. (2014). *Medical Intelligence in Europe*. Retrieved March 28, 2016, from http://ec.europa.eu/health/preparedness_response/generic_preparedness/planning/medical_intelligence_en.htm

Vanarse, S., Dr. (2014). *Big Data Breathes Life Into Next-Gen Pharma and R&D*. WIPRO: Applying Thought. Retrieved from <https://www.wipro.com/documents/big-data-breathes-life-into-next-gen-pharma-RD.pdf>.

Groves, P., Kayyali, B., Knott, D., & Van Kuiken, S. (2013). The 'big data' Revolution in Health Care. Center for US Health System Reform Business Technology Office. Retrieved from http://www.pharmatalents.es/assets/files/Big_Data_Revolution.pdf

Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: Promise and potential. *Health Inf Sci Syst Health Information Science and Systems*, 2(1), 3. Retrieved from <http://hissjournal.biomedcentral.com/articles/10.1186/2047-2501-2-3>

Yuanjun, L. (2014). IT Transforms Medical Services. *China Today*, 63(8), 28-30. Retrieved from <http://0-web.a.ebscohost.com.liucat.lib.liu.edu/ehost/detail/detail?sid=d3ec439e-8cb6-4b58-9732-1cc9438032fd@sessionmgr4002&vid=0&hid=4107&bdata=JnNpdGU9ZWWhvc3QtbGl2ZSZzY29wZT1zaXRl&preview=false#AN=97666302&db=a9h>

Pasquale, F. (2013). Grand Bargains for Big Data: The Emerging Law of Health and Information. *Maryland Law Review*, 72(3), 682-772. Retrieved from <http://0-web.a.ebscohost.com.liucat.lib.liu.edu/ehost/detail/detail?sid=b17fdb9-3d22-4930-a829->

[709416b021df@sessionmgr4005&vid=0&hid=4107&bdata=JnNpdGU9ZWhvc3QtbGl2ZSZzY29wZT1zaXRl&preview=false#AN=88318050&db=a9h](http://www.cnet.com/news/apple-v-samsung-patent-trial-recap-how-it-all-turned-out-faq/)

Tibken, S. (2014, May 7). Apple v. Samsung patent trial recap: How it all turned out. Retrieved March 28, 2016, from <http://www.cnet.com/news/apple-v-samsung-patent-trial-recap-how-it-all-turned-out-faq/>

Vascellaro, J. E. (2012, August 25). Apple Wins Big in Patent Case. Retrieved March 28, 2016, from <http://www.wsj.com/articles/SB10000872396390444358404577609810658082898>

Vascellaro, J. E. (2012, April 12). Smartphone Patents: The Never-Ending War. Retrieved March 28, 2016, from <http://www.wsj.com/articles/SB10001424052702303624004577339931560320116>

Feldman, R., Ewing, T., & Jeruss, S. (2013). That AIA 500 Expanded: The Effects of Patent Monetization Entities. *UCLA Journal of Law & Technology*, 17(2), 1-107. Retrieved from <http://0-web.a.ebscohost.com.liucat.lib.liu.edu/ehost/detail/detail?sid=a7ab65c3-7e51-44b7-8a52-8e26fb72d190@sessionmgr4004&vid=0&hid=4107&bdata=JnNpdGU9ZWhvc3QtbGl2ZSZzY29wZT1zaXRl&preview=false#AN=95640737&db=a9h>

Cusumano, M. A. (2013). The Apple-Samsung Lawsuits. *Communications of the ACM*, 56(1), 28-31. Retrieved from <http://0-dl.acm.org.liucat.lib.liu.edu/citation.cfm?id=2398366>

Inmon, W. (2015, August 19). Textual data marches on - Anticipating and Preventing Lawsuits. Retrieved March 28, 2016, from <http://analyticsindiamag.com/textual-data-marches-on-anticipating-and-preventing-lawsuits/>

Katz, D. M. (2013). Quantitative Legal Prediction. *Emory Law Journal*, 62(4), 1-58. Retrieved from http://law.emory.edu/elj/_documents/volumes/62/4/contents/katz.pdf

DeCesaris, M. (2015, July 20). United States: Using Big Data In Gathering Expert Testimony. Retrieved from <http://www.mondaq.com/unitedstates/x/413716/trials+appeals+compensation/Using+Big+Data+In+Gathering+Expert+Testimony>

Pavolotsky, J. (2012, November). Demystifying Big Data. Retrieved March 28, 2016, from http://www.americanbar.org/publications/blt/2012/11/02_pavolotsky.html

Gaffney, A. (2013, May 31). FDA Looks to Big Data to Assist Off-Label Marketing Investigations. Retrieved March 28, 2016, from <http://www.raps.org/regulatoryDetail.aspx?id=8601>

Surdeanu, M., Nallapati, R., Gregory, G., Walker, J., & Manning, C. D. (2011). Risk analysis for intellectual property litigation. *Proceedings of the 13th International Conference on Artificial Intelligence and Law*, 116-120. Retrieved from <http://0-dl.acm.org.liucat.lib.liu.edu/citation.cfm?id=2018375>

Magalhães, J., Quoniam, L., Ferreira, V., Ferreira, P., & Boechat, N. (2013). Research on Pharmaceuticals Patents in Times of Big Data: A Contribution of the Web 2.0 for Medicinal Chemistry. Retrieved from <http://www.esciencecentral.org/journals/research-on-pharmaceuticals-patents-in-times-of-big-data-a-contribution-of-the-web-for-medicinal-chemistry-ipr.1000105.php?aid=18535>

Martin, J. (2013, October 28). Big Data for Patents: Deriving Novel Insights from Publicly Available Data. Retrieved March 28, 2016, from <http://insights.wired.com/profiles/blogs/big-data-for-patents-deriving-novel-insights-from-publicly#axzz3pLzu0Xhb>

Dysart, J. (2013, April 1). How lawyers are mining the information mother lode for pricing, practice tips and predictions. Retrieved March 28, 2016, from http://www.abajournal.com/magazine/article/the_dawn_of_big_data/

Litan, R. E. (2015, April 23). Big data is key to disrupting the U.S. patent industry. Retrieved March 28, 2016, from <http://www.brookings.edu/research/opinions/2015/04/22-big-data-key-innovation-litan>

Healthit.gov, What is HIE? (2014, May 12). Retrieved February 07, 2016, from <https://www.healthit.gov/providers-professionals/health-information-exchange/what-hie>

Healthit.gov, Health IT Quick Stats. (2016, January). Retrieved February 07, 2016, from <http://dashboard.healthit.gov/quickstats/quickstats.php>

Marr, B. (2015, April 21). Big Data: Saving 13,000 Lives A Year By Predicting Earthquakes. Retrieved February 12, 2016, from <http://www.forbes.com/sites/bernardmarr/2015/04/21/big-data-saving-13000-lives-a-year-by-predicting-earthquakes/2/#5b4954bb23de>

Macri, G. (2015, October 01). Patent Trolls are Already Abusing the Apple v. Samsung Ruling. Retrieved February 20, 2016, from <http://www.insidesources.com/patent-trolls-are-already-abusing-the-apple-v-samsung-ruling/>

Howard, B. (2015, July 14). 2015 First Half Patent Case Filing Trends. Retrieved February 20, 2016, from <https://lexmachina.com/2015-first-half-patent-case-filing-trends/>

Nazer, D., & Ranieri, V. (2014, July 09). Why Do Patent Trolls Go to Texas? It's Not for the BBQ. Retrieved February 20, 2016, from <https://www.eff.org/deeplinks/2014/07/why-do-patent-trolls-go-texas-its-not-bbq>

Redondo, S. (2014, November 22). What is Semantic Search and Why Should I Care? Retrieved February 21, 2016, from <https://www.searchenginejournal.com/seo-101-semantic-search-care/119760/>

Bajwa, S. (2014). Apple v. Samsung: Is it Time to Change our Patent Trial System? *Global Business & Development Law Journal*, 27(1), 4th ser., 1-31. Retrieved from <http://digitalcommons.mcgeorge.edu/cgi/viewcontent.cgi?article=1003&context=globe>

Burney, B. (2015, December 30). A Closer Look at Liquid Lit Manager. Retrieved from <http://www.legaltechnews.com/id=1202745982131/A-Closer-Look-at-Liquid-Lit-Manager?slreturn=20160124144642>

Cornell University Law School. (2015). Patent. Retrieved March 15, 2016, from <https://www.law.cornell.edu/wex/patent>

USPTO. (2014, October). General Information Concerning Patents. Retrieved March 15, 2016, from <http://www.uspto.gov/patents-getting-started/general-information-concerning-patents>

Reisinger, D. (2015, December 4). Samsung Will Pay Apple for Damages - But Wants The Cash Back. Retrieved from <http://fortune.com/2015/12/04/samsung-pay-apple-patents/>

Tomasco, S. (2009, June 2). IBM and EU Partners Create a Better Way to Fight AIDS Virus. Retrieved March 22, 2016, from <https://www-03.ibm.com/press/us/en/pressrelease/27627.wss>

IDC. (2015, November 9). New IDC Forecast Sees Worldwide Big Data Technology and Services Market Growing. Retrieved March 31, 2016, from <http://www.idc.com/getdoc.jsp?containerId=prUS40560115>

Industrial Engineer. (2016). The Front Line. *Industrial Engineer*, 48(4). Retrieved from <http://www.iienet2.org/industrialengineer/Details.aspx?id=41301>

Adamson, D. (2015). Big Data in Healthcare Made Simple. Retrieved April 16, 2016, from <https://www.healthcatalyst.com/big-data-in-healthcare-made-simple>

Byrne, M. (2015, July 14). Meet MUMPS, the Archaic Health-Care Programming Language That Predicted Big Data. Retrieved April 16, 2016, from <http://motherboard.vice.com/read/meet-mumps-the-archaic-health-care-programming-language-that-predicted-big-data-2>

Jeevan, M. (2015, September 04). How I chose the right programming language for Data Science. Retrieved April 18, 2016, from <http://bigdata-madesimple.com/how-i-chose-the-right-programming-language-for-data-science/>

Asay, M. (2015, July 24). In data science, the R language is swallowing Python. Retrieved April 19, 2016, from <http://www.infoworld.com/article/2951779/application-development/in-data-science-the-r-language-is-swallowing-python.html>

Vesset, D., & Morris, H. D. (2011). The Business Value of Predictive Analytics. *IDC: Analyze the Future*, 1-12. Retrieved from nxdimension.net/wp-content/uploads/2013/04/ibm-spss-predictive-analytics-business-value-whitepaper.pdf.

Rathenberg, J. (2016, April 20). Personal Communication, Email Interview

Pyramid. (2016). Pyramid Introduction. Retrieved April 25, 2016, from <http://docs.pylonsproject.org/projects/pyramid/en/latest/narr/introduction.html>

PostgreSQL. (2016). About PostgreSQL. Retrieved April 25, 2016, from <http://www.postgresql.org/about/>

Patel, K., Sundararajan, S., Marangos, C., & Zimmers, E. W., Jr. (2016, April). Discovering Hidden Connections with Big Data. *Industrial Engineer*, 48(4). Retrieved from <http://ie.binus.ac.id/2016/06/09/discovering-hidden-connection-with-big-data/>

Batrinca, B., & Treleaven, P. C. (2015, February). Social media analytics: A survey of techniques, tools and platforms. *AI & Society*, 30(1), 89-116. Retrieved from <http://link.springer.com/article/10.1007/s00146-014-0549-4#/page-1>

Idreos, S., Groffen, F., Nes, N., Manegold, S., & Mullender, S. (2012). MonetDB: Two Decades of Research in Column-oriented Database Architectures. *IEEE*, 34, 40-45. Retrieved from <http://dare.uva.nl/record/1/380005>

Nicolaou, A. (2014, May 15). The 9 Best Languages For Crunching Data. Retrieved July 13, 2016, from <http://www.fastcompany.com/3030716/the-9-best-languages-for-crunching-data>

Mullin, J. (2016, January 05). Trolls made 2015 one of the biggest years ever for patent lawsuits. Retrieved July 30, 2016, from <http://arstechnica.com/tech-policy/2016/01/despite-law-changes-2015-saw-a-heap-of-patent-troll-lawsuits/>

Kerstetter, J. (2012, April 5). How much is that patent lawsuit going to cost you? Retrieved July 30, 2016, from <http://www.cnet.com/news/how-much-is-that-patent-lawsuit-going-to-cost-you/>

Nicas, J., & Chin, J. (2016, July 20). Stronger Chinese Patent Laws Also Help U.S. Companies. Retrieved July 30, 2016, from <http://www.wsj.com/articles/stronger-chinese-patent-laws-also-help-u-s-companies-1468994404>

McDonald, C. (2016, June 07). How Big Data is Reducing Costs and Improving Outcomes in Health Care | MapR. Retrieved September 05, 2016, from <https://www.mapr.com/blog/reduce-costs-and-improve-health-care-with-big-data>

SAS. (2016). Big data analytics: What it is and why it matters. Retrieved September 20, 2016, from http://www.sas.com/en_us/insights/analytics/big-data-analytics.html

McDonald, C. (2016, June 07). How Big Data is Reducing Costs and Improving Outcomes in Health Care. Retrieved November 13, 2016, from <https://www.mapr.com/blog/reduce-costs-and-improve-health-care-with-big-data>

Ryan, M. (n.d.). Critical Steps for a Successful Data Conversion. Retrieved November 18, 2016, from <http://www.legalfiles.com/Portals/0/whitepages/Critical-Steps-for-Successful-Conversion.pdf>

Additional Sources

Kaplan, B. (2014). How Should Health Data be Used? Yale Institution for Social and Policy Studies, 1-39. Retrieved from <http://bioethics.yale.edu/sites/default/files/files/ISPS14-025.pdf>

Forest Rim Technology Inc. (2016). Solution 3: Litigation Early Warning System. Retrieved March 28, 2016, from <http://www.forestrimtech.com/business-corporate-information/solution-3-litigation-early-warning-system>

Liquid Litigation Management Inc. (2016). About Liquid Litigation Management. Retrieved March 28, 2016, from http://liquidlitigation.com/about_us

Biblowitz, A. (2012). Big Data - Mitigating Risks and Finding Benefits. *The Metropolitan Corporate Counsel*, 20(12), 1-2. Retrieved from http://llm.liquidlitigation.com/rs/liquidlitigationmanagement/images/BigData-MitigatingRisks-FindingBenefits_LLM.pdf

Doherty, S. (2014, September 29). Lex Machina Out With Legal Analytics For Big Pharma. Retrieved March 28, 2016, from <http://xseedcap.com/news-article/lex-machina-legal-analytics-big-pharma/>

Lex Machina. (2014). Fish & Richardson Delivers Strategic IP Business Counsel with Lex Machina. Retrieved from <https://lexmachina.com/wp-content/uploads/2014/12/FR-CaseStudy.pdf>

Smartphone market revenue worldwide 2013-2015 | Statistic. (2016). Retrieved February 08, 2016, from <http://www.statista.com/statistics/412145/global-smartphone-sales-value-global-region/>

Cooley, D. C., Kacedon, B. D., & Paul, J. C. (2012, November 12). Joinder of Multiple Defendants Rejected Under the America Invents Act When Plaintiff Alleged Infringement by Defendants Involved in Different Transactions With Different Products. Retrieved February 17, 2016, from <http://www.finnegan.com/resources/articles/articlesdetail.aspx?news=4945aad6-4b75-4d2b-9801-e810eb33b0e8>

Collie, R. (2010, September 15). Interview with JamieMac of Predixion Software. Retrieved March 22, 2016, from <http://www.powerpivotpro.com/2010/09/interview-with-jamiemac-of-predixion-software/>

Kim, I., Kremer, G., & Cornell, B. (2016, April). Examining Consumer Emotions. *Industrial Engineer*, 48(4). Retrieved from <http://www.iienet2.org/IndustrialEngineer/details.aspx?id=41256>