The Relationship between Product Description and Trademark Extension Strategy

中文摘要

本研究旨在運用文字探勘方法，探討企業在財報中的產品敘述與商標延伸之間的關係。商標延伸意旨企業在拓展新市場時，向主管單位申請延伸既有商標的保護範圍。本研究採用文字相似度分析，衡量企業在年度財報（10-K Form）中針對各產品的描述於每年的變化程度，並且使用此數值來預測商標在未來被延伸的可能性。文獻指出企業的商標策略對其品牌價值，至企業價值都有重大的影響。商標延伸相較建立新品牌不僅在成本、風險都較小，也能繼承客戶對既有品牌的信賴，有較高的成功機會。因此，本研究的貢獻在於提供關於商標策略的預測性資訊。

本研究的假說為產品敘述的變化與商標延伸的策略有關。此假說的意涵在於，當企業在年報中提供的對商品敘述不同於往年的敘述時，很可能預示著其現有的商標/產品線未來即將延伸至新的產業。

本研究著眼於美國的「工業電腦與商用電腦設備」領域，並且運用 1993 – 2019 年由美國證券交易委員（SEC）提供的年度財報。本研究的結果指出，產品敘述的相似程度與商標是否延伸為負相關。

在未來，本研究預期能結合更多在商標文件中的文字資訊，並能開展以公開資訊（財報）與文字探勘為出發點的商標研究，期望能在資料中找出更多影響企業商標策略與決策的重要因素，以提供企業與投資人關於企業未來的商標策略的相關線索與洞見。

關鍵字：商標延伸、商標策略、文字探勘、文字相似度分析

Abstract

This study investigates the relationship between corporate product description and corporate trademark strategy on extension. In particular, I apply a *text-similarity* method to capture the year-to-year change of description of the products matched to registered trademarks for a firm, and investigate whether the similarity of product description explains the trademark extension strategy. The literature indicate that the trademark extension is an important corporate strategy, which ultimately affect a firm's economic value through the price premium of products.

I argue that the *dissimilarity* of product description from this year to the previous year is associated with a greater incidence of trademark extension. The underlying notion is that, when a firm provides description for the same product from this year to another year and such year-to-year description is not alike, it indicates that the firm shows more incentives to expand its current business line to others.

My empirical study examines firms operated in the industry of "Industrial and Commercial Machinery and Computer Equipment" in the U.S. The sample period spans from 1993 to 2019. Indeed, I find that the similarity of product description is negatively associated with a corporate trademark extension; however, the finding is not significantly significant.

In future study, I plan to incorporate more information in the trademark documents (for example, the textual description of the intended use for goods or services) to study the corporate trademark/product strategy, and expand the study to different industries to validate the generalization of my finding. I believe my ongoing research can provide new insights for corporates or investors who have interests in analyzing corporate trademark extension strategy.

Keywords: Product description**,** Trademark extension, Annual report, Textual analysis

Contents

1. Introduction

The *product language* refers to the combinations of words that firms use to describe their products. Importantly, the product language is highly associated with a firm's strategies. For example, Hoberg and Phillips (2018a) document that the product language links to a corporate synergy strategy. My study is to contribute the literature on product language and corporate business strategies. In particular, I ask whether the product description in corporate annual reports relates to trademark extension strategy.

The trademark extension denotes a registered trademark is subsequently registered by the same owner and in different classes, and thus the original trademark owner keeps the right to extend their brands to the new areas, which refers to the brand extension in a marketing practice that uses an established brand name in one category to introduce product in totally different categories (Choi, 1998).

The trademark extension has been documented as one of important business strategies to increase firm value. For example, trademark extensions generate a separate, internally coherent group of marks that jointly protect the underlying brand and preserve its distinctiveness (Nasirov, 2020). In addition, it has been shown that the introduction costs for the extension of pre-existing brands are lower compared with products under a new brand (Tauber, 1988), and thus less risky.

My research question is to examine the relationship between the similarity of product description and the trademark extension strategy. The underlying notion is that, when a firm provides description for the same product from this year to another year and such year-to-year description is not alike, it indicates that the firm shows more incentives to expand its current business line to others.

To answer my research question, I design a new empirical method. First, I create my own measure to capturing the similarity of year-to-year product description,

denoted as *Similarity Score*. The procedure of the measure is briefly introduced here, and the details can refer to Section 3.2. I use "trademark names" to locate a corporate "product" appeared in its annual reports. Specifically, I require "trademark name" and "product name" to be exactly matched. Then, for a trademark-product matched product, I collect all of the sentences containing the matched name in a 10-K document. I then use these sentences to construct a similarity matrix. The *Similarity Score* is the cosine similarity between product descriptions in year t and product description in year t–1. That is, how is the description for a product appeared in this year similar to the product in the previous year.

Second, to identify whether a trademark is an extended trademark, I create a binary variable that equals to one when a registered trademark that was registered before and not newly register as a new trademark with the identical trademark name but in different classes, otherwise zero. I denote the binary variable as *IsExtension*.

My sample covers all public firms in the USA and in the industry of "Industrial and Commercial Machinery and Computer Equipment" (industry classification of 2-digit SIC code of 35). The sample period is from 1993 to 2019 and contains 158 unique firms that have 404 unique trademarks.

I summarize my findings as follows. First, I find that number of trademarks matched to product names appeared in annual reports increased dramatically in the past decade, from 2010 to 2019, indicating that firms increasing file for trademark registration for their products. Second, the similarity score of extended trademarks is 0.79 and the mean similarity score of not extended trademark is 0.81. The finding provides the preliminary evidence, indicating that a lower similarity score (i.e., larger year-to-year change of product description) leads to the higher likelihood of observing a trademark extension. Third, I run regressions that regress the incidence of trademark

extension on the similarity score of year-to-year product description. The regression results show that the trademark extension is negatively correlated with the similarity score; However, the coefficient of determination is not high. At the end of this study, I provide some discussions on this weak evidence and some improvements for my future research.

2. Literature Review and Hypothesis Development

2.1 Trademark extension

The trademark extension denotes a registered trademark is subsequently registered by the same owner and in different classes, and thus the original trademark owner keeps the right to extend their brands to the new areas, which refers to the brand extension in a marketing practice that uses an established brand name in one category to introduce product in totally different categories (Choi, 1998). Brand extension established brand names to enter new product markets or reinforce existing market positions (Keller and Aaker, 1992). Therefore, trademark extensions generate a separate, internally coherent group of marks that jointly protect the underlying brand and preserve its distinctiveness (Nasirov, 2020). In addition, it has been shown that the introduction costs for the extension of pre-existing brands are lower compared with products under a new brand (Tauber, 1988), and thus less risky. And brand extensions are more likely to succeed than creating a new brand (Lane and Jacobson, 1995; Block, Fisch, and Sandner, 2014; Thoma, 2019).

1. Sattler et. al. (2010) study the extent to which consumers are willing to pay a price premium for extended products, and the impact of potential success drivers on consumers' attitudes toward the extension and the extension price premium. Cohen, D

et. al. (2014) investigates the effect of trademark extension on firm's market valuation. And they conclude that investors can approximate the extension's future success easier based on the strength and history of the parent brand.

Overall, the above mentioned studies indicate that the trademark extension is an important corporate strategy, which ultimately affect a firm's economic value through the price premium of products.

2.2 Corporate product description in annual reports

There are several studies that use textual descriptions in annual reports to examine corporate strategic behaviors. For example, Li et. al. (2013) develop the measure of competition based on management's disclosures in their 10-K filing. Hoberg and Phillips (2010) use the product descriptions in 10-K filings to examine the extent to which whether firms exploit product market synergies through asset complementarities in mergers and acquisitions.

In particular, for the *product description* in annual reports, Hoberg and Phillips (2018a) analyze the words that firms use to describe their products in the annual reports. Their studies focus on multiple-industry firm operations and find that firms operate across industries with higher product language overlap. They conduct several corporate strategic studies of using the product descriptions and construct an online Hoberg and Philips Data Library for researchers who are interested in quantified information from textual content of annual reports. For example, there are text-based network industry classifications (TNIC) data, industry concentration and total similarity data, and product market fluidity data.

Hoberg and Phillips's (2018a) clearly indicate that the product description in annual reports is important and useful information to explain or even forecast corporate business strategies.

2.3 The relation between product description and trademark extension

My study relates to Hoberg and Phillips's (2018a), in which I also use the product description in corporate annual reports. However, differently from their study, I create my own text-based measure of capturing whether a firm changes its description on a product from this year to another year. This newly proposed measure is based on the similarity score of product description for a given product. Importantly, I premise that this measure is highly associated with corporate business strategies on the trademark extension. The underlying notion is that, when a firm provides different descriptions on the same product from this year to another year, it could indicate that the firm has incentives to expand its current business line of the product to other and different business lines. When a firm has made such expansion, it is likely the firm register new trademarks, in which the *trademark name* is the same as the trademark name registered before, but the registered class is different, which this is what the literature refers to the trademark extension.

Hypothesis 1: The dissimilarity of product description from this year to the previous year is associated with a greater incidence of trademark extension.

3. Data and Variable Construction

3.1 Data

My study is to examine the relationship between disclosed information of product description and a corporate trademark extension strategy. I study public firms in the

USA and in the industry of “Industrial and Commercial Machinery and Computer Equipment” (industry classification of 2-digit SIC code of 35). The sample period is from 1993 to 2019.

The disclosed information of product descriptions is based on the corporate released annual reports. The 10-K form is required by the Securities and Exchange Commission (SEC) for publicly traded firms in the USA. The 10-K form is also known as the annual report provided by firms. I collect annual reports based on 10-K filings retrieved from EDGAR. The contents of these 10-K documents are parsed using PERL (a text-processing programming language), identified by CIK code, and classified by fiscal year.

To capture the trademark extension, I collect trademark documents from the U.S. Patent and Trademark Office (USPTO). The procedure is as follows. I use python to interact with the TSDR Data API provided by USPTO, to retrieve structured trademark documents that can be easily parsed by my program (in a data format called JSON). All the trademark documents are identified by application serial numbers. Figure 1 shows an example of the preview of the structured data responded by the TSDR Data API provided by USPTO.

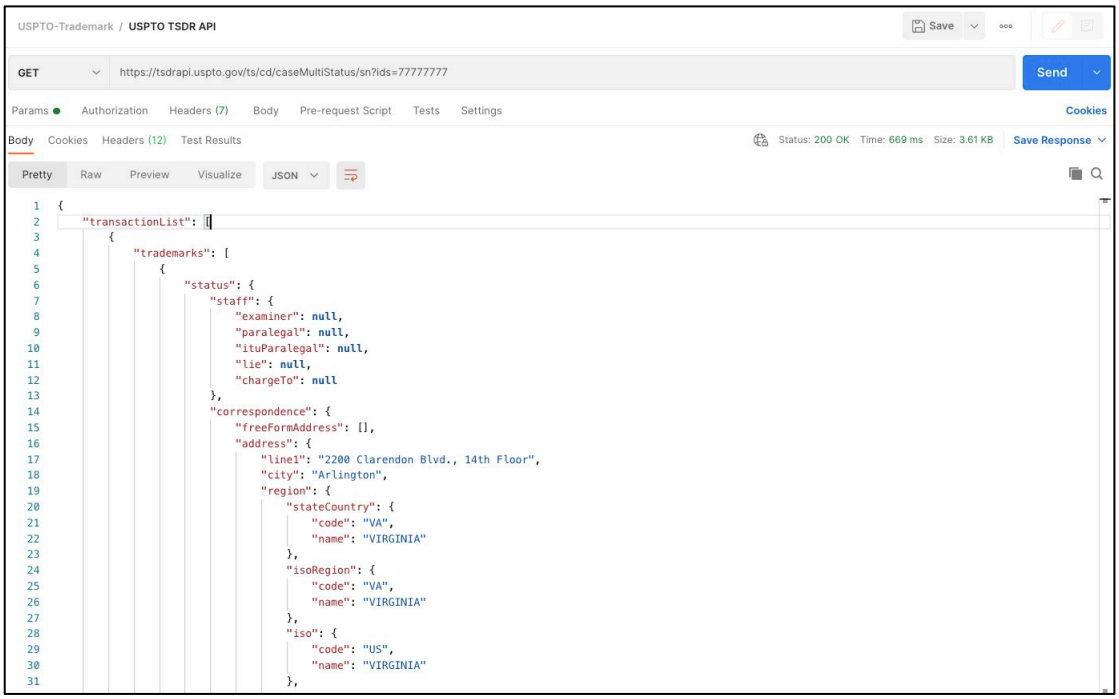
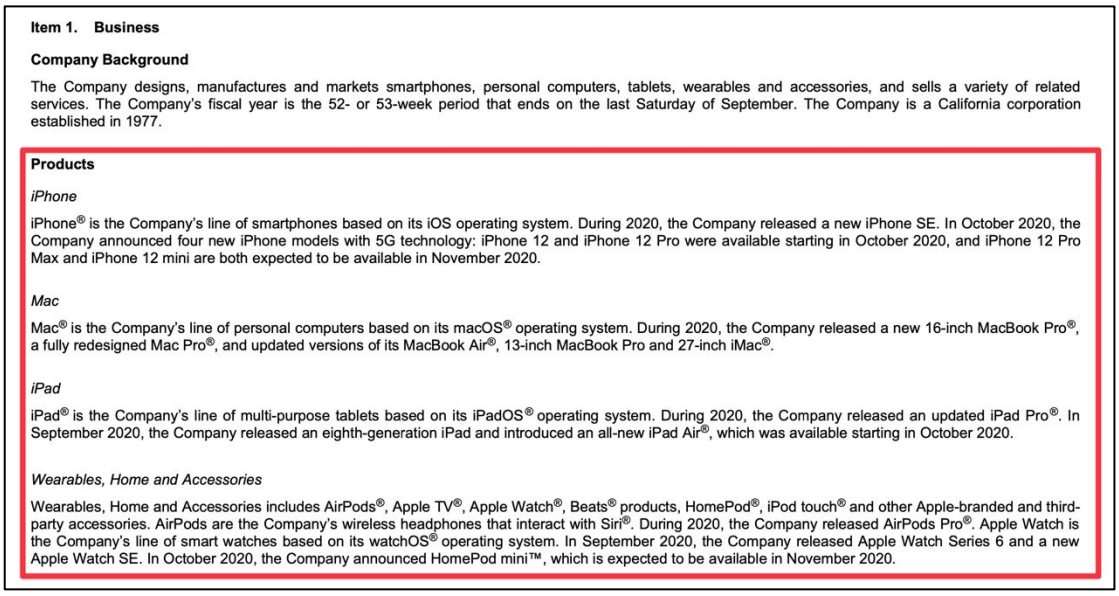


Figure 1. A preview of the structured data responded by the TSDR Data API

3.2 Variable construction

3.2.1 Locate the products that exactly match to trademark names

For the 10-K forms, in particular, the "business description section" of a 10-K form is mandated by SEC regulation, and it is required that firms need to describe the significant products they offer to their customers. Figure 2 shows an example of product description appeared in the 10-K form for Apple Inc, which describes the key products, such as iPhone, Mac, iPad etc.



Item 1. Business

Company Background

The Company designs, manufactures and markets smartphones, personal computers, tablets, wearables and accessories, and sells a variety of related services. The Company's fiscal year is the 52- or 53-week period that ends on the last Saturday of September. The Company is a California corporation established in 1977.

Products

iPhone

iPhone® is the Company's line of smartphones based on its iOS operating system. During 2020, the Company released a new iPhone SE. In October 2020, the Company announced four new iPhone models with 5G technology: iPhone 12 and iPhone 12 Pro were available starting in October 2020, and iPhone 12 Pro Max and iPhone 12 mini are both expected to be available in November 2020.

Mac

Mac® is the Company's line of personal computers based on its macOS® operating system. During 2020, the Company released a new 16-inch MacBook Pro®, a fully redesigned Mac Pro®, and updated versions of its MacBook Air®, 13-inch MacBook Pro and 27-inch iMac®.

iPad

iPad® is the Company's line of multi-purpose tablets based on its iPadOS® operating system. During 2020, the Company released an updated iPad Pro®. In September 2020, the Company released an eighth-generation iPad and introduced an all-new iPad Air®, which was available starting in October 2020.

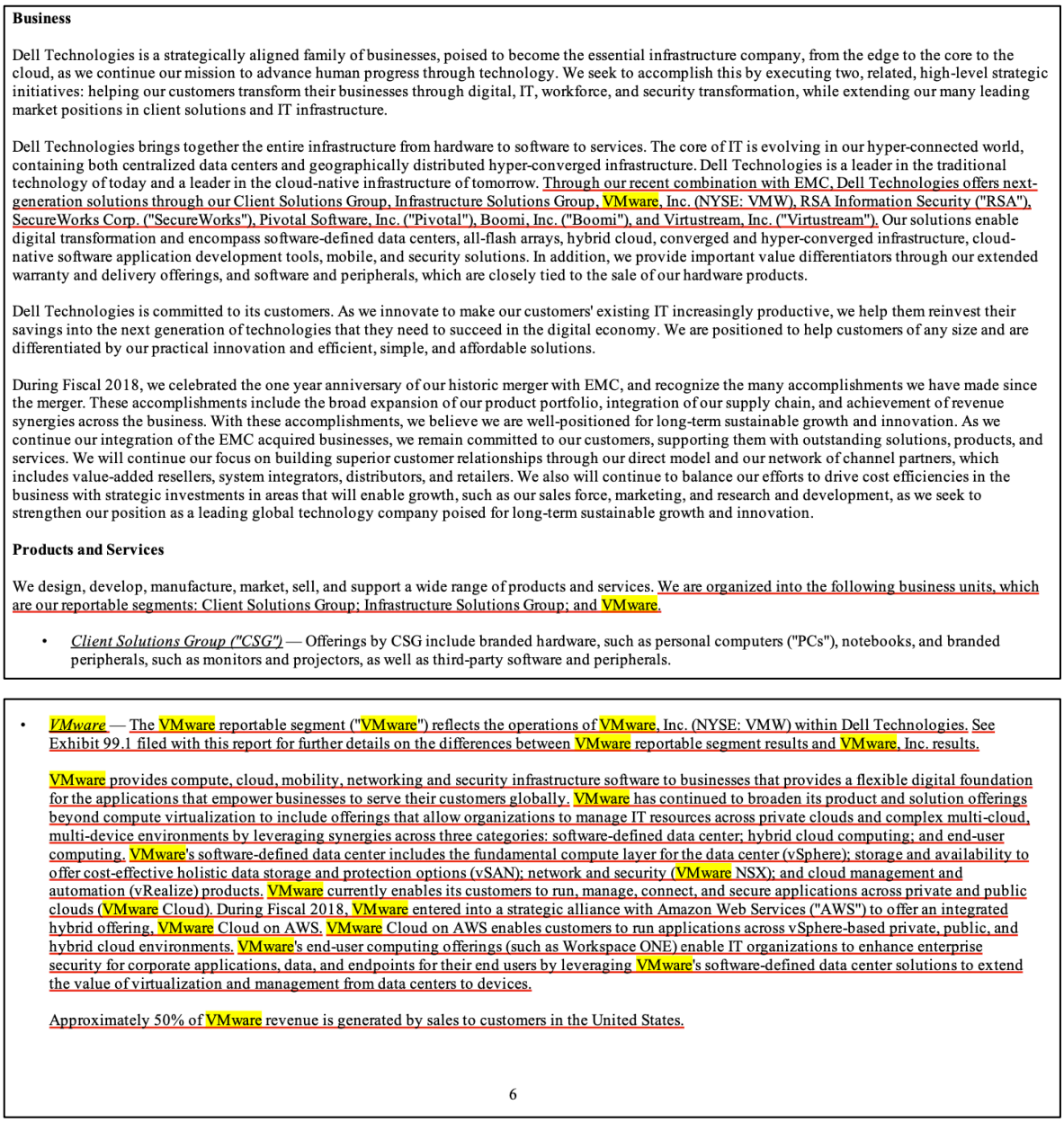
Wearables, Home and Accessories

Wearables, Home and Accessories includes AirPods®, Apple TV®, Apple Watch®, Beats® products, HomePod®, iPod touch® and other Apple-branded and third-party accessories. AirPods are the Company's wireless headphones that interact with Siri®. During 2020, the Company released AirPods Pro®. Apple Watch is the Company's line of smart watches based on its watchOS® operating system. In September 2020, the Company released Apple Watch Series 6 and a new Apple Watch SE. In October 2020, the Company announced HomePod mini™, which is expected to be available in November 2020.

Figure 2. An example of product descriptions of the 10-K form for Apple Inc.

Obviously, at a company-wise level, there are many products and many trademarks. My main research question is to examine the relationship between the year-to-year change of product description and corporate trademark extension strategy. For the starting point, I use "trademark names" to locate a corporate "product". Specifically, I require them to be exactly matched. Then, I call them trademark-matched products.

For a trademark-product matched product, I collect all of the sentences containing the matched name in a 10-K document. I then use these sentences to construct a similarity matrix, which contains the score of "how this trademark in this year is similar to the trademark in the previous year". Figure 3 shows an example of the product description of "VMware" provided by Dell Inc.



Business

Dell Technologies is a strategically aligned family of businesses, poised to become the essential infrastructure company, from the edge to the core to the cloud, as we continue our mission to advance human progress through technology. We seek to accomplish this by executing two, related, high-level strategic initiatives: helping our customers transform their businesses through digital, IT, workforce, and security transformation, while extending our many leading market positions in client solutions and IT infrastructure.

Dell Technologies brings together the entire infrastructure from hardware to software to services. The core of IT is evolving in our hyper-connected world, containing both centralized data centers and geographically distributed hyper-converged infrastructure. Dell Technologies is a leader in the traditional technology of today and a leader in the cloud-native infrastructure of tomorrow. Through our recent combination with EMC, Dell Technologies offers next-generation solutions through our Client Solutions Group, Infrastructure Solutions Group, VMware, Inc. (NYSE: VMW), RSA Information Security ("RSA"), SecureWorks Corp. ("SecureWorks"), Pivotal Software, Inc. ("Pivotal"), Boomi, Inc. ("Boomi"), and Virtustream, Inc. ("Virtustream"). Our solutions enable digital transformation and encompass software-defined data centers, all-flash arrays, hybrid cloud, converged and hyper-converged infrastructure, cloud-native software application development tools, mobile, and security solutions. In addition, we provide important value differentiators through our extended warranty and delivery offerings, and software and peripherals, which are closely tied to the sale of our hardware products.

Dell Technologies is committed to its customers. As we innovate to make our customers' existing IT increasingly productive, we help them reinvest their savings into the next generation of technologies that they need to succeed in the digital economy. We are positioned to help customers of any size and are differentiated by our practical innovation and efficient, simple, and affordable solutions.

During Fiscal 2018, we celebrated the one year anniversary of our historic merger with EMC, and recognize the many accomplishments we have made since the merger. These accomplishments include the broad expansion of our product portfolio, integration of our supply chain, and achievement of revenue synergies across the business. With these accomplishments, we believe we are well-positioned for long-term sustainable growth and innovation. As we continue our integration of the EMC acquired businesses, we remain committed to our customers, supporting them with outstanding solutions, products, and services. We will continue our focus on building superior customer relationships through our direct model and our network of channel partners, which includes value-added resellers, system integrators, distributors, and retailers. We also will continue to balance our efforts to drive cost efficiencies in the business with strategic investments in areas that will enable growth, such as our sales force, marketing, and research and development, as we seek to strengthen our position as a leading global technology company poised for long-term sustainable growth and innovation.

Products and Services

We design, develop, manufacture, market, sell, and support a wide range of products and services. We are organized into the following business units, which are our reportable segments: Client Solutions Group; Infrastructure Solutions Group; and VMware.

- *Client Solutions Group ("CSG")* — Offerings by CSG include branded hardware, such as personal computers ("PCs"), notebooks, and branded peripherals, such as monitors and projectors, as well as third-party software and peripherals.

- *VMware* — The VMware reportable segment ("VMware") reflects the operations of VMware, Inc. (NYSE: VMW) within Dell Technologies. See Exhibit 99.1 filed with this report for further details on the differences between VMware reportable segment results and VMware, Inc. results.

  VMware provides compute, cloud, mobility, networking and security infrastructure software to businesses that provides a flexible digital foundation for the applications that empower businesses to serve their customers globally. VMware has continued to broaden its product and solution offerings beyond compute virtualization to include offerings that allow organizations to manage IT resources across private clouds and complex multi-cloud, multi-device environments by leveraging synergies across three categories: software-defined data center; hybrid cloud computing; and end-user computing. VMware's software-defined data center includes the fundamental compute layer for the data center (vSphere); storage and availability to offer cost-effective holistic data storage and protection options (vSAN); network and security (VMware NSX); and cloud management and automation (vRealize) products. VMware currently enables its customers to run, manage, connect, and secure applications across private and public clouds (VMware Cloud). During Fiscal 2018, VMware entered into a strategic alliance with Amazon Web Services ("AWS") to offer an integrated hybrid offering, VMware Cloud on AWS. VMware Cloud on AWS enables customers to run applications across vSphere-based private, public, and hybrid cloud environments. VMware's end-user computing offerings (such as Workspace ONE) enable IT organizations to enhance enterprise security for corporate applications, data, and endpoints for their end users by leveraging VMware's software-defined data center solutions to extend the value of virtualization and management from data centers to devices.

  Approximately 50% of VMware revenue is generated by sales to customers in the United States.

6

Figure 3. An example of product description: all the sentences contain the product of "VMware" from Dell Inc.

3.2.2 Similarity of product description

This section illustrates the procedure of my measure of capturing the similarity of year-to-year product description. I call the measure as the *Similarity Score*.

The *Similarity Score* is the cosine similarity between product descriptions in year *t* and product description in year *t*–1. That is, how is the description for a product in this year similar to the product in the previous year. First, we covert all the words in a document into a vector, and each of the document can be represented by a vector. Figure 4 shows an example, in which each of the document is a vector in a 3-dimentional space, with each axis representing the word counts of the words appear in the document.

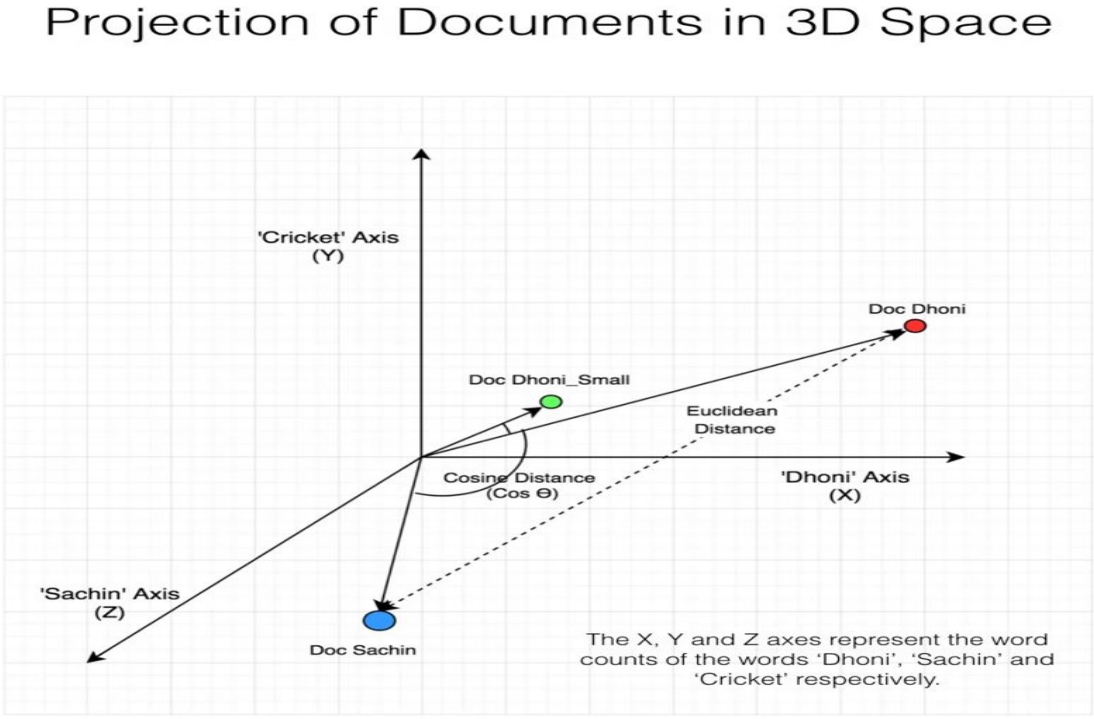


Figure 4. The concept of "word to vector"

(Image from: https://www.machinelearningplus.com/nlp/cosine-similarity/)

Then we take the cosine similarity measure which capture the similarity of documents by the cosine value of the two vectors in high-dimensional space. In the following equation, $\vec{A}$ *and* $\vec{B}$ are two documents represented by two vectors, and the cosine value of these vectors capture the similarity of two documents.

$$\cos\theta = \frac{\vec{A} \cdot \vec{B}}{\|\vec{A}\| \, \|\vec{B}\|} = \frac{\sum_1^n A_i B_i}{\sqrt{\sum_1^n {A_i}^2} \sqrt{\sum_1^n {B_i}^2}}$$

In my setting, $\vec{A}$ is the document of product descriptions in year t and $\vec{B}$ is the document of product descriptions in year t-1, then the cosine value of them represent "how similar is the product described in this year and the product described in last year.

The similarity score is a value between 0 and 1. Similarity score of 0 means that two documents are totally different, and similarity score of 1 means that two documents are totally identical. I denote the variable of *Similarity Score* as the variable that measures the year-to-year similarity of product description in the annual reports. The similarity measure is refer to Huang (2008) and Huang et al (2011).

3.2.3 Indicator of trademark extension

Trademark extension refers to the action that a firm extends the coverage of a trademark because it has new products/services under an existing trademark. For example, Amazon was famous for its e-commerce businesses, and thus they register the trademark "Amazon" in the category "IC 042. US 100 101. G & S", which represent "computerized online ordering service featuring the wholesale and retail distribution of books". However, when Amazon decided to enter the financial market in 2020, they extend the trademark "Amazon" in the category "IC 036. US 100 101 102. G & S", which is related to "financial and monetary services". In my methodology, I measure trademark extension programmatically with three steps, shows as follows:

- ✓ **Step 1:** Each trademark has one or more U.S Class(es) → a set of U.S Classes
- ✓ **Step 2:** Trademarks with the same word mark may appear in different years (due to renewal) with different sets of U.S. Classes. (due to extension)
- ✓ **Step 3:** Trademark extension occurs when:

  For 2 same trademarks TM(t) and TM(t-1),

  $\text{Set}\{ \text{USClasses}(\text{TM(t)}) \} - \text{Set}\{ \text{USClasses}(\text{TM(t-1)}) \} \neq \emptyset$

In the below, I use the trademark **"AMAZON"** as an example to illustrate each step of the procedure of identifying the trademark extension.

Step 1: Figure 5 indicates that, the trademark "AMAZON" has a set of U.S Classes of {100, 101} in 1995. (US Serial Number: 2078496)

Step 2: Figure 6 indicates that, there is another trademark with the same word mark "AMAZON" with a set of U.S classes {100, 101, 102} (US Serial Number: 90296565) in 2020.

Step 3: Figure 7 indicates that the set of U.S Classes of "AMAZON" in 2020 is {100, 101, 102}, and the set in 1995 is {100, 101}. Because {100, 101, 102} – {100, 101} = {102} ≠ ∅, we know they extend their trademark to U.S Class 102, which is class "Insurance and financial".

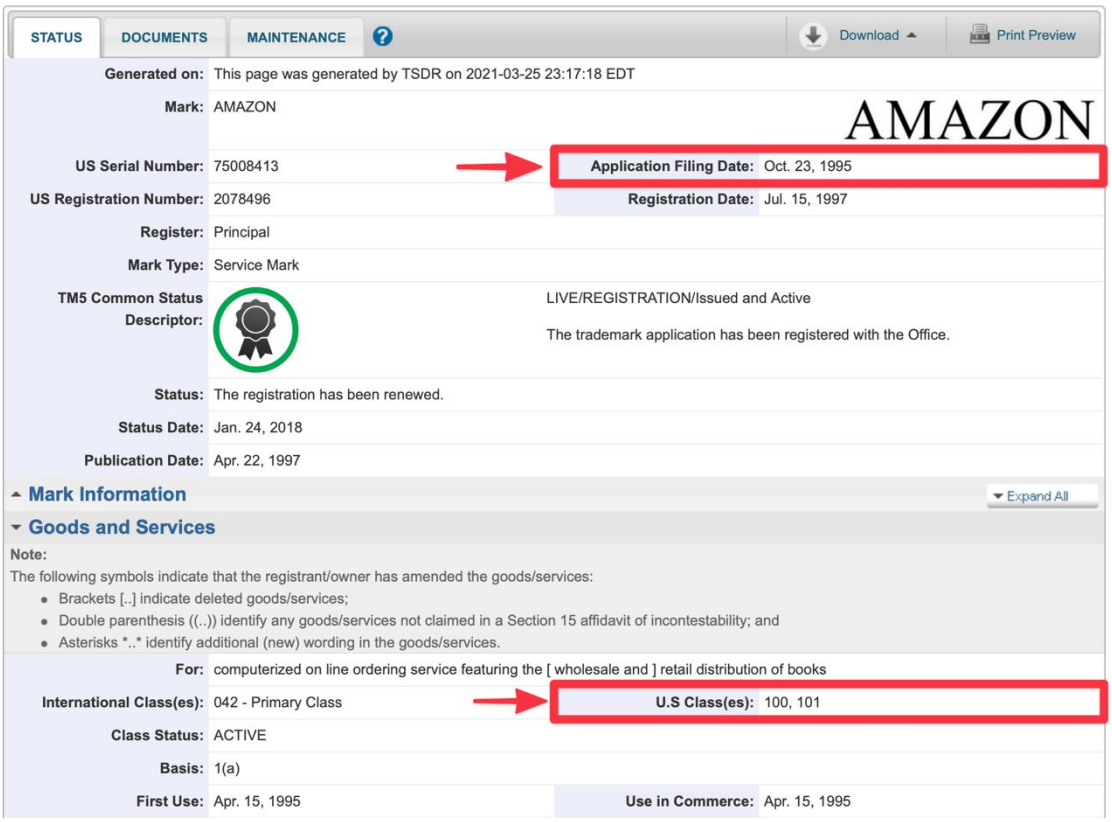


Figure 5. For this example, the trademark "AMAZON" has a set of U.S Classes of {100, 101} in 1997. (US Serial Number: 2078496)

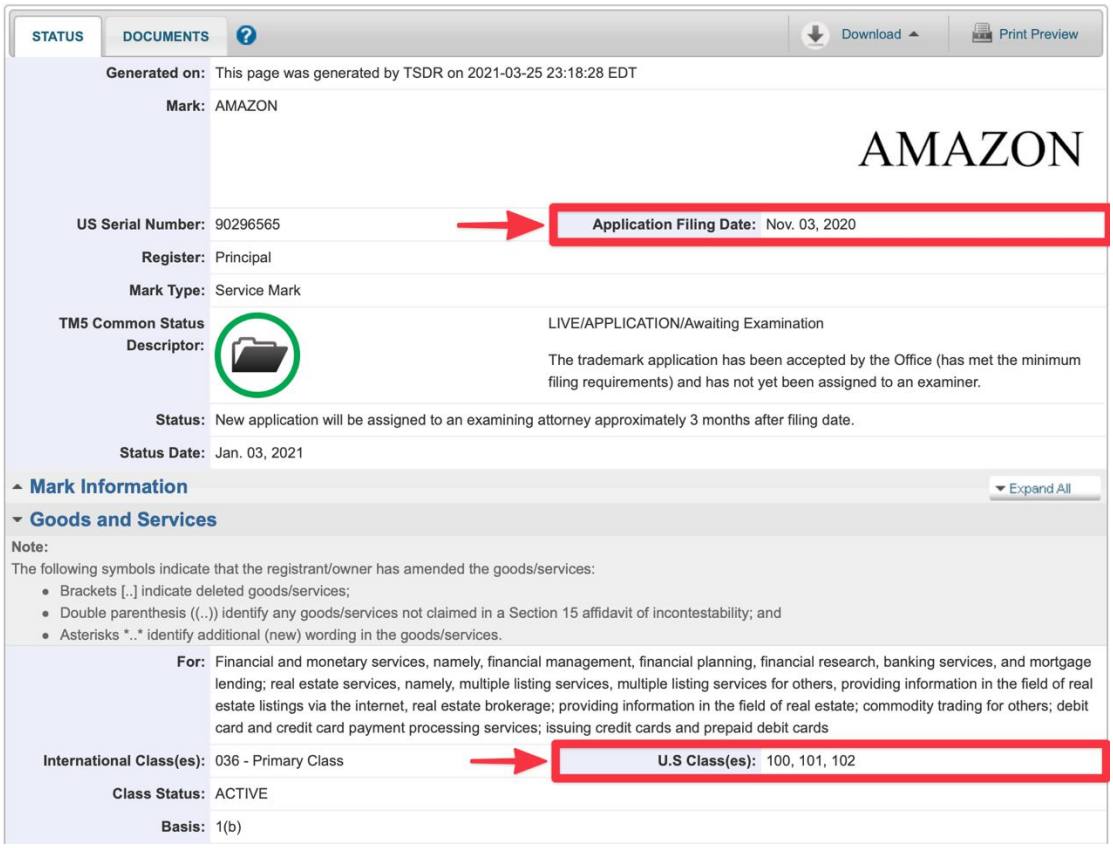


Figure 6. There is another trademark with the same word mark "AMAZON" with a set of U.S classes {100, 101, 102}. (US Serial Number: 90296565)

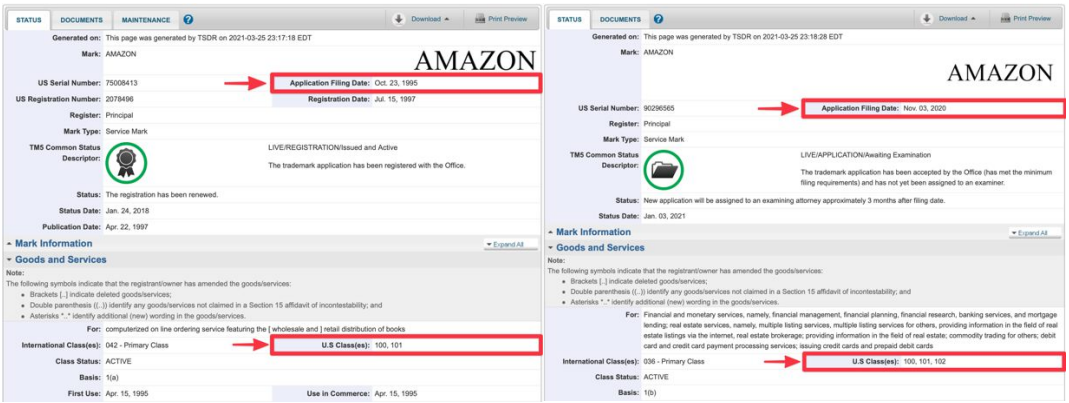


Figure 7. The set of U.S Classes of "AMAZON" in 2020 is {100, 101, 102}, and the set in 1995 is {100, 101}. Because {100, 101, 102} – {100, 101} = {102} ≠ ∅, we know they extend their trademark to U.S Class 102, which is class "Insurance and financial".

(Please refer to Appendix A for more information about U.S Trademark Classes) That is, there is a trademark extension occurred in 2020.

Specifically, I create the variable of "*IsExtension*" as the binary variable that equals to 1 when a trademark *i* is an extended trademark and 0 otherwise. The definition of trademark extension is when a registered trademark that was registered before and not newly register as a new trademark with the identical trademark name but in different classes.

4. Empirical Results

4.1 Summary statistics

My sample covers all public firms in the USA and in the industry of "Industrial and Commercial Machinery and Computer Equipment" (industry classification of 2-digit SIC code of 35). The sample period is from 1993 to 2019 and contains 158 unique firms that have 404 unique trademarks. Figure 8 shows the breakdown analysis for the number of trademarks by year, which indicates that number of trademarks matched to product names appeared in annual reports increased dramatically in the past decade, from 2010 to 2019.

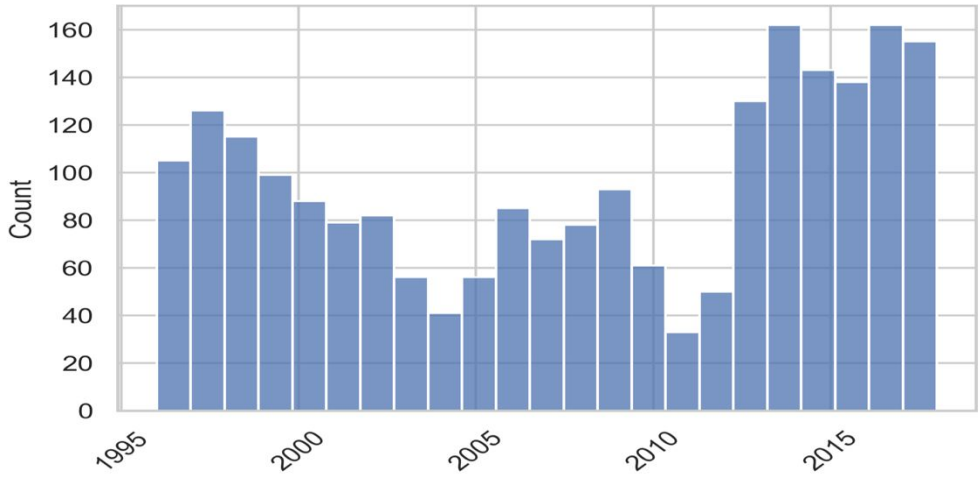


Figure 8. Number of trademarks by year from 1995 to 2019.

Table 1 provides the summary statistics. The mean value of the variable of *IsExtension* indicates that my sample has only 66 trademarks (i.e., 2,209 × 0.03=66) that are classified as trademark extension (named, extended trademark) and 2,143 number of trademarks that are NOT classified as trademark extension (named, not extended trademark). The similarity score is about 0.817.

Table 1. Summary statistics

Table 1 reports summary statistics for the data used to analyze the relationship between year-to-year product similarity and trademark extensions. The sample period is of 1993–2019.

Variable	No. of Obs.	Mean	Std. Dev	Min	Q1	Median	Q3	Max
Similarity Score	2,209	0.817	0.193	0.005	0.758	0.893	0.947	1
IsExtension	2,209	0.030	0.170	0	0	0	0	1

4.2 Preliminary results

Table 2 shows the distribution of two groups of trademarks: (1) extended trademark, and (2) not extended trademark. The mean similarity score of extended trademarks is 0.79 and the mean similarity score of not extended trademark is 0.81. The finding provides the preliminary evidence, indicating that a lower similarity score (i.e., larger year-to-year change of product description) leads to the higher likelihood of observing a trademark extension.

Table 2. Similarity Score of Year-to-Year Product Description

This table presents the distribution of the Similarity Score for the group of extended trademark and the group of not extended trademark, respectively.

Type of Trademarks	Extended Trademark	Not Extended Trademark
No. of Observations	66	2,143
Mean	0.793	0.818
Std. Dev	0.219	0.192
Min	0.035	0.005
Q1	0.769	0.758
Median	0.864	0.894
Q3	0.931	0.948
Max	0.989	1

Figure 9 further displays the Box plot of the similarity score for the group of the not extended trademarks and for the group of extended trademarks. I find that the *Similarity Score* is slightly higher for the not extended trademarks. This figure further corroborates the finding shown in Table 2.

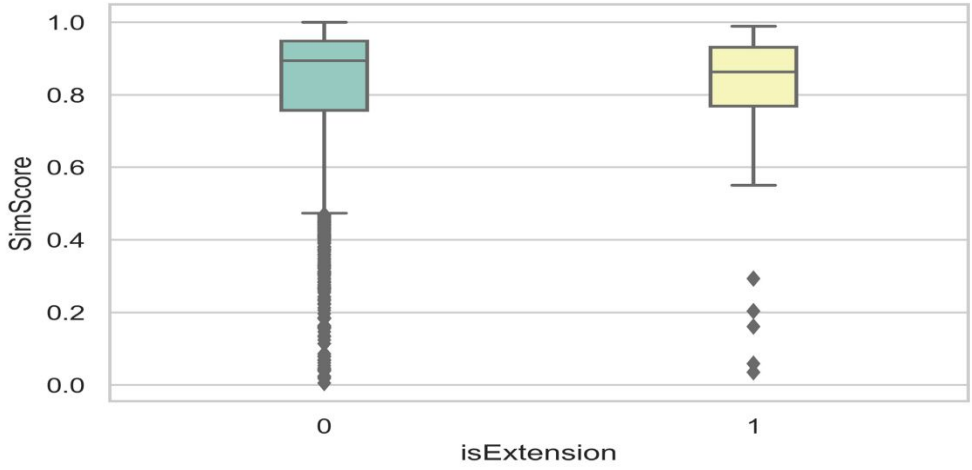


Figure 9. Box plot of the similarity score

4.3 Univariate regression analysis

I proposed to conduct a regression analysis on the data using following equation:

$$IsExtension_{i,j,t} = \beta \times SimScore_{i,t} + +\lambda_t + \phi_j + \varepsilon_{i,t}, \qquad (1)$$

where the dependent variable is *IsExtension* and it is the binary variable that equals to 1 when a trademark *i* owned by firm *j* is an extended trademark and 0 otherwise. In addition, *j* and *t* denote the j^{th} firm that owns trademark *i* in year *t*. ϕ_i and λ_{jt} are firm-fixed effect and year-fixed effects. The definition of trademark extension is when a registered trademark that was registered before and not newly register as a new trademark with the identical trademark name but in different classes. The interested independent variable is *Similarity Score* and it is the variable that measures the year-to-year similarity of product Description in the annual reports. My hypothesis argues that the *less* similar the year-to-year product description, the more likely that a corporate use trademark extension strategy. Therefore, we expect the coefficient of *Similarity Score*, β, should be negative and statistically significant.

I run four model specifications with alternative fixed effects. Model (1) does not include any fixed effects, Model (2) include firm fixed effects, Model (3) include year fixed effects, and Model (4) includes both firm and year fixed effects. Across all models, the trademark extension is negatively correlated with year-to-year product description similarity score as we expected. However, the coefficient of determination is not high, and in the following section I will provide some possible causes and discussion to address the topic.

Table 3. Univariate regression analysis

This table presents the univariate regression results.

	Dependent variable: ***IsExtension***			
	Model (1)	Model (2)	Model (3)	Model (4)
Similarity Score	–0.020	–0.013	–0.015	–0.013
	(–0.61)	(–0.46)	(–0.54)	(–0.47)
Constant	0.046*	0.041*	0.049*	0.061**
	(1.66)	(1.70)	(1.80)	(2.10)
Standard Errors clustered at	Firm	Firm	Firm	Firm
Firm Fixed Effect	No	Yes	No	Yes
Year Fixed Effect	No	No	Yes	Yes
Number of observations	2,209	2,209	2,209	2,209
R^2	0.001	0.001	0.016	0.015

5. Conclusion, Discussion, and Future Research

5.1 Conclusion

My study creates the novel measure of capturing a year-to-year product similarity score and the novel indicator of capturing a product-matched-trademark extension. In particular, the product description in my study refers to the sentences of containing the trademark name in corporate annual reports. I examine the relationship between two variables for 158 unique firms that have 404 unique trademarks that appeared from 1993 to 2019. I provide empirical evidence and show that a trademark is more likely to be extended in a distinct business category when the product description on that trademark share less similarity with the product description on the same trademark appeared in the previous year.

5.2 Discussion

Overall, I find that a trademark is more likely to be extended in a distinct business category when the product description on that trademark share less similarity with the product description on the same trademark appeared in the previous year. That is, by studying a corporate product description, it allows investors to forecast a corporate trademark strategy of expanding to different business lines.

I do, however, acknowledge that my study can be improved in several ways. The most important one is about the procedure of identifying a trademark that is extended. My current sample only contains 3% of extended trademarks among all selected trademarks. The major reason could be that, I take the narrow definition of trademark extension, which it requires the "mark word" of a trademark has to be exactly matched with the mark word of any existing trademark. However, a broader definition of trademark extension may improve the study. That is, we consider the whole "trademark family" instead of the "exact-match-only" trademark.

I illustrate the rationale by using an example for the trademark of "Amazon." The Amazon company registered the trademark "Amazon" in the category "IC 042. US 100 101. G & S", which represent "computerized online ordering service featuring the wholesale and retail distribution of books", and in 2008, they register a trademark "amazon web service" in the class "IC 042. US 100 101. G & S" which represent "Application service provider". The case of "amazon web service" can be broadly defined as a trademark extension of "amazon" because they belong to the same trademark family. However, how to identify whether two trademarks belong to the same trademark family programmatically, or how to set the similarity threshold for two trademarks to be grouped into the same trademark family are the issues that need to be

addressed. Figure 10 and Figure 11 shows the records of "Amazon" trademark in USPTO database.

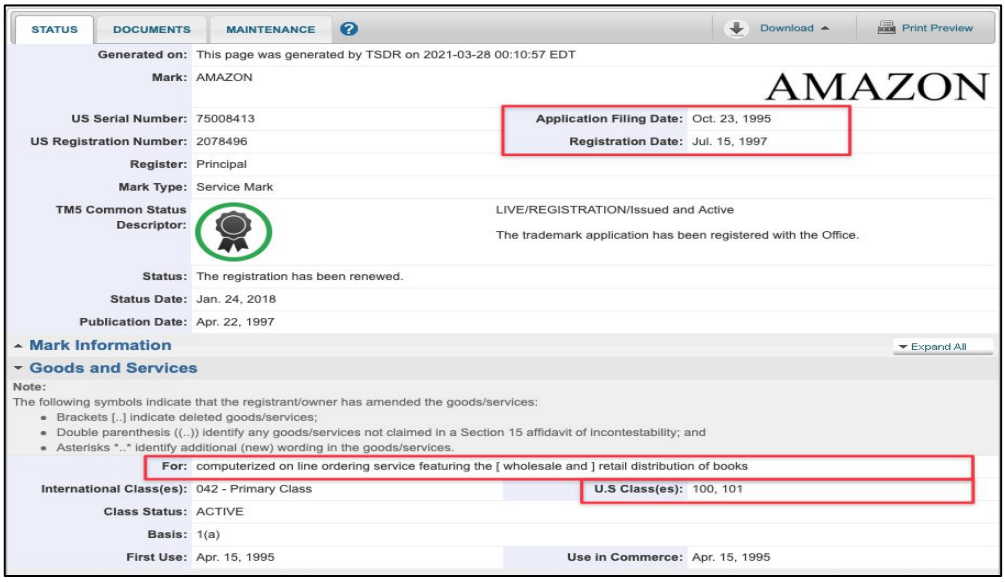


Figure 10. Descriptions and U.S Classes of Amazon as the online bookstore.

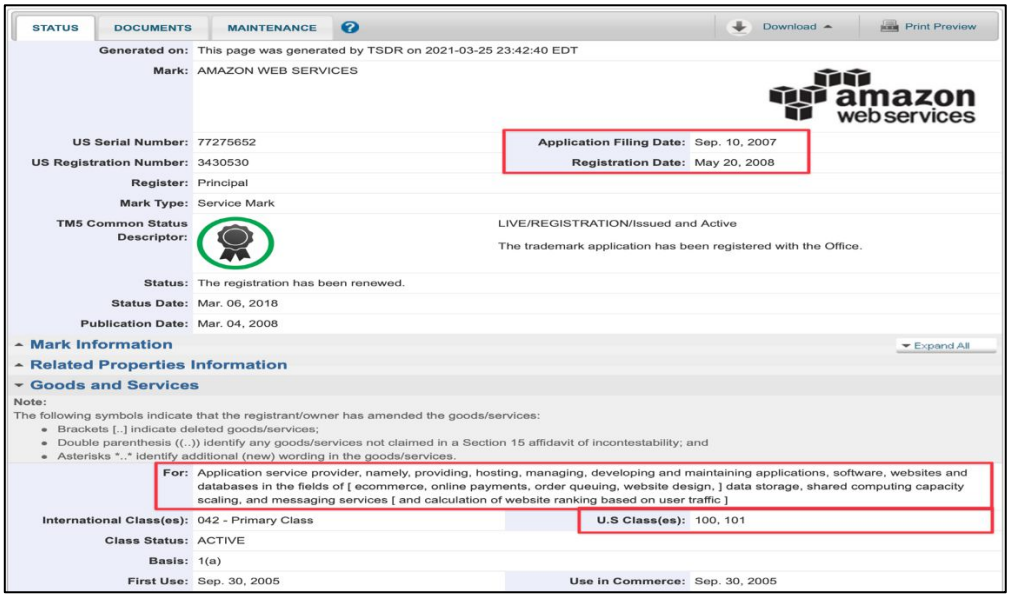


Figure 11. Descriptions and U.S Classes of Amazon as the web services provider.

Another example is about the trademark of "Tesla." In 2018, Tesla want to publish their own tequila and register for the trademark "Teslaquila". The trademark registration was failed because the trademark is similar to the word "Tequila".

Therefore, they eventually change their product name to “Tesla Tequila” and register the trademark “Tesla” under a new class for “Distilled agave liquor; distilled blue agave liquor” and the application is accepted in 2020. Figure 12 shows the record of Tesla trademark in USPTO database.

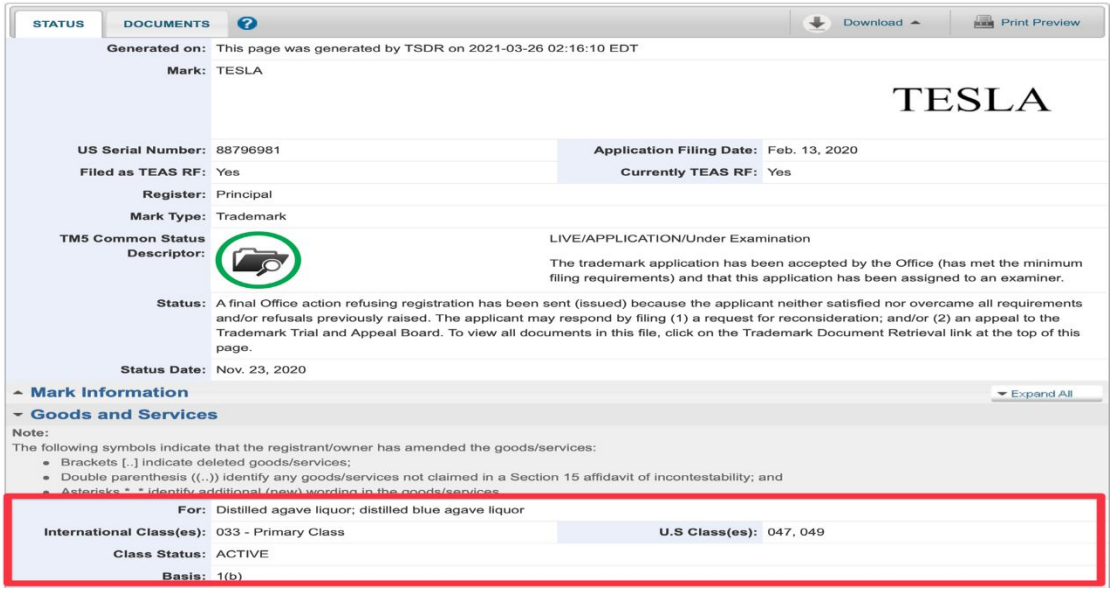


Figure 12. Example of trademark extension of tesla

5.3 Future research

For my future studies, I intend to incorporate more information in the trademark documents to study the corporate trademark/product strategy at finer granularity, and expand the study to other industries to examine if my finding for firms located in the industry of Industrial and Commercial Machinery and Computer Equipment can be generalized to other industries. In addition, I plan to link these trademarks with more public data (for example, the market value of the trademark) to expand the study on the economic values of brand extension (Sattler et. al. 2010, Block et. al. 2014). I expect to find factors that affect a firm’s choice on its trademark strategy and provide valuable

insights for corporates and investors who have interests in analyzing corporate trademark strategy.

References

(1) Block, J. H., Fisch, C., & Sandner, P. G. (2014). Trademark families: Characteristics and market values. *Journal of Brand Management*, 21, 150−170.

(2) Cohen, D. (1986). Trademark strategy. *Journal of Marketing*, 50, 61−74.

(3) Choi, J. P. (1998). Brand extension as informational leverage. *The Review of Economic Studies*, 65, 655−669.

(4) Hoberg, G., & Phillips, G. (2010). Product market synergies and competition in mergers and acquisitions: A text-based analysis. *The Review of Financial Studies*, 23, 3773−3811.

(5) Hoberg, G., & Phillips, G. (2016). Text-based network industries and endogenous product differentiation. *Journal of Political Economy*, 124, 1423−1465.

(6) Hoberg, G., & Phillips, G. (2018a). Conglomerate industry choice and product language. *Management Science*, 64, 3735−3755.

(7) Huang, A. (2008). Similarity measures for text document clustering. *In Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008), Christchurch, New Zealand* (Vol. 4, pp. 9−56).

(8) Huang, C. H., Yin, J., & Hou, F. (2011). A text similarity measurement combining word semantic information with TF-IDF method. *Jisuanji Xuebao(Chinese Journal of Computers)*, 34, 856−864.

(9) Keller, K. L., & Aaker, D. A. (1992). The effects of sequential introduction of brand extensions. *Journal of Marketing Research*, 29, 35−50.

(10) Lane, V., & Jacobson, R. (1995). Stock market reactions to brand extension announcements: The effects of brand attitude and familiarity. *Journal of Marketing*, 59, 63–77.

(11) Li, F., Lundholm, R., & Minnis, M. (2013). A measure of competition based on 10-K filings. *Journal of Accounting Research*, 51, 399−436.

(12) Nasirov, S. 2020. Trademark value indicators: Evidence from the trademark protection lifecycle in the US pharmaceutical industry. *Research Policy*, 49, 103929.

(13) Sattler, H., Völckner, F., Riediger, C., & Ringle, C. M. (2010). The impact of brand extension success drivers on brand extension price premiums. *International Journal of Research in Marketing*, 27, 319−328.

(14) Tauber, E. M. (1988). Brand leverage: strategy for growth in a cost-control world.

(15) Thoma, G. 2019. Composite value index of trademark indicators, *World Patent Information*, 56, 64–75.

Appendix A. List of U.S Trademark Classes

Class 1	Raw or partly prepared materials
Class 2	Receptacles
Class 3	Baggage, animal equipments, portfolios, and pocket books
Class 4	Abrasives and polishing materials
Class 5	Adhesives
Class 6	Chemicals and chemical compositions
Class 7	Cordage
Class 8	Smokers' articles, not including tobacco products
Class 9	Explosives, firearms, equipments, and projectiles
Class 10	Fertilizers
Class 11	Inks and inking materials
Class 12	Construction materials
Class 13	Hardware and plumbing and steamfitting supplies
Class 14	Metals and metal castings and forgings
Class 15	Oils and greases
Class 16	Protective and decorative coatings
Class 17	Tobacco products
Class 18	Medicines and pharmaceutical preparations
Class 19	Vehicles
Class 20	Linoleum and oiled cloth
Class 21	Electrical apparatus, machines, and supplies
Class 22	Games, toys, and sporting goods
Class 23	Cutlery, machinery, and tools, and parts thereof

Class 24	Laundry appliances and machines
Class 25	Locks and safes
Class 26	Measuring and scientific appliances
Class 27	Horological instruments
Class 28	Jewelry and precious-metal ware
Class 29	Brooms, brushes, and dusters
Class 30	Crockery, earthenware, and porcelain
Class 31	Filters and refrigerators
Class 32	Furniture and upholstery
Class 33	Glassware
Class 34	Heating, lighting, and ventilating apparatus
Class 35	Belting, hose, machinery packing, and nonmetallic tires
Class 36	Musical instruments and supplies
Class 37	Paper and stationery
Class 38	Prints and publications
Class 39	Clothing
Class 40	Fancy goods, furnishings, and notions
Class 41	Canes, parasols, and umbrellas
Class 42	Knitted, netted, and textile fabrics, and substitutes thereof
Class 43	Thread and yarn
Class 44	Dental, medical, and surgical appliances
Class 45	Soft drinks and carbonated waters
Class 46	Foods and ingredients of foods
Class 47	Wines

Class 48	Malt beverages and liquors
Class 49	Distilled alcoholic liquors
Class 50	Merchandise not otherwise classified
Class 51	Cosmetics and toilet preparations
Class 52	Detergents and soaps
Class 100	Miscellaneous
Class 101	Advertising and business
Class 102	Insurance and financial
Class 103	Construction and repair
Class 104	Communication
Class 105	Transportation and storage
Class 106	Material treatment
Class 107	Education and entertainment
Class 200	Collective membership
Class 201	Goods
Class 202	Services
Class A	Goods
Class B	Services