Retail Innovation Trends: a Topic Modelling Analysis of Corporate Patents

Author: Alexandru Nestor
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands

ABSTRACT,

With the changes that can be seen developing in various industries and markets in the past years, retail has been no exception. With studies pointing to managers' inability of making decisions, this paper aims to disambiguaite this seemingly technologically intensive industry. This research uses the powerful Latend Dirichlet Allocation semantic analysis algorithm, in conjunction with technological and R&D management frameworks in order to identify a set of trends and technologies that are currently defining the fast paced retail markets. The resulting trends are then put into context in order to create a comprehensive image of the industry, which is analysed using a strategic positioning model, finally creating actionable inputs and considerations for industry players.

Graduation Committee members: Dr. Igors Skute, Dr. Maximilian Goethner.

Keywords
Innovation, R&D, Strategy, Topic Modelling, Unsupervised, Retail

1. INTRODUCTION

As recent years have presented themselves with high levels of economic and societal turmoil (Council of foreign relations, n.d.; United Nations, 2021; Person, 2021), organizations that may have already been operating in a highly competitive environment now have to deal with the VUCA world (Volatile, Uncertain, Complex and Ambiguous) (Buder, 2021). This has accentuated existing pressures on companies, such as those specific to classical economic theories, but have also created new ones, such as increasing stakeholder pressures, Co2 and other externalities costs, scarcity of resources and human capital, etc. In a scarcity-based environment, even a bad decision can mean high costs for companies. There are several organizational perspectives that can be used to support, debate and elaborate ways of dealing with the aforementioned, such as Mahoney's resource-based-view (1992) and Henry Chesbrough's outside-in approach (2003), which could take this research stream in many directions. However, the focus throughout this paper is mainly on improving decision makers' capacities of technology and R&D portfolio management using algorithmic analysis through text clustering. This will be facilitated by Shankar's et al (2021) classification of retail technologies.

A significant motivator for this paper is that decision makers are now more important than ever, including, if not even more so, those from retail companies, which are highly impactful due to their consistent and extended contact with consumers. PWC (2014) has conducted a survey which states that only 41% of retail executives feel fully prepared to make impactful decisions, while 54% feel somewhat prepared and the rest 5% are moving from being unprepared to prepared. This means that there is a barrier and an inability when it comes to making big decisions for retail companies, decisions, which retail executives would already have to be doing highly complex, global environments (Reinartz et al 2011). Another highly impacting factor is that the situation relating innovation within retail companies, up until recently, has been that many of them would outsource their R&D and position themselves as adopters of innovation, rather than producers (Pantano, 2014). Although this is not necessarily the case for business to consumer (B2C) companies that originally started out as producers, the situation is becoming more dynamic. As such, whether we are to consider the internet, mobile technologies, the exponential increase of computing power or even artificial intelligence and robotics, it is certain that emergent technologies are creating rapid shifts within all markets, disrupting the retail industry and forcing participants to rethink their strategies (Grewal et al., 2021). In fact, we are seeing a drastic change throughout the entire sector, as researchers are trying to study this phenomenon and understand it better (Grewal et al., 2021). Moreover, the large R&D expenditures by big companies are causing a growing divide between themselves and smaller companies (HBR, 2021). The resulting situation and environment serve to portrait the increasing impact that novel technologies have on society, with the overarching research and development portfolio management implications.

As nowadays, most of the information and data has become digitized, so has the use and need of accurate and performant digital research techniques and data analysis tools become more prevalent (Silva et al., 2021, Kherwa & Bansal, 2018). That is why there is an increasing need for performant fore sighters to use this data to create value for decision makers and help them broaden their views and adopt new perspectives (Buder, 2021). In this sense, the explosion of data in the past decade is crucial, as, due to this development, with the right methods, decision makers can know more and directly translate even larger amounts of knowledge into actionable information (McAfee, 2012). One such database can be patent and trademark offices pertaining to various national or supra national administrations, such as that in the United States of America, where large amounts of patents related to the retail industries can be found. Patenting constitutes the main procedures through which companies protect their business concepts and technologies (Tang et al 2012), with some even calling it at the heart of a nation's policies regarding technological innovation (Mansfield, 1986, Kim et al, 2013). The number of patent filings have been growing significantly (Kwon et al, 2022), which has in turn led businesses and governments alike to consider patent analysis as a valid source of information (Kim et al, 2015). More specifically, the patents can help decision makers conjure a comprehensive perspective of the industry's technological outlook, developments, tendencies which can ultimately be used as an input for strategic decisions regarding R&D management, such as resource allocation, technological pathing, prioritisation and so on.

This paper draws upon several theories and theoretical constructs in order to achieve technological trend identification within the retail sector. With the use of the Latent Dirichlet Allocation algorithm for word clustering, retail technology frameworks and innovation portfolio management frameworks, which are discussed in detail in further sections, the following research question is proposed for the study:

What are the current and emerging technological development patterns in the retailing sector with the highest economic and societal impact?

2. THEORETICAL FRAMEWORK

2.1 General Overview of the Importance of Emerging Tech Development and R&D

Throughout the past years, given the increase in the amount of data available, there has been an interest in the development and usage of semantic analysis techniques. Such studies are directed at analyzing technological trends in retail, of which there are none that address specifically the set of the previously described companies, nor the retail sector generally, except for very few, although not focused on technological trends. Despite this fact, it is relevant to acknowledge the method's convergence with other sectors as a sign of validity for proposing the application of algorithmic analysis on this study's subject niche. Algorithmic tools are used, throughout the academic world to analyze various topics, such as technologies and programming languages, with significant results pertaining to topics such as website "Design/CSS", "Mobile App Development", "Object oriented Programming", "UI Development" and "Security" (Johri & Bansal, 2018). Other studies include the field of logistics, with themes such as "Vehicle communication", "Risk forecasting", "Shipping service", "Sensing", "Database system", "Process optimization" (Choi & Song, 2018) and so on. One retail related research has been conducted by Gurrola-Perez et al (2022) , and does not refer to anything related to R&D, technology or innovation. The closest related research, converging with both the innovation and retail spheres, as it analyses business trends in mobile commerce, is conducted by Saritas et al (2021). This research identifies 10 topic clusters with sub-terms for each, which are ranked from one to five based on their importance, using a niche text analysis technique. Except for the latter, all the papers are using latent Dirichlet allocation (LDA).

In the context of technological portfolio management, terms such as R&D and innovation are bound to appear often. Defining them is therefore a fitting beginning of a theoretical framework.

Given the complex nature of nowadays' retailers, the global scope of their operations as well as the point which technology has reached, the reader is encouraged to consider innovation holistically. The rise of buzzwords such as "deep tech", "industry 4.0" "artificial intelligence", "quantum computing" and so on, only serves to further support that innovations are often vague, or even invisible to the naked eye. According to Tidd and Bessant (2018), innovation is a multi-level concept, going from invention, design, management, implementation and commercialization. With innovative behaviors, organizations can generate new ideas, create new products and services as well as improve the already existing ones in order to better meet customer needs. Innovation has the aim of creating firm-specific competitive advantage (Tidd and Bessant, 2018). R&D has a smaller reach, in terms of meaning, however they both serve the same purpose throughout the paper and can be used interchangeably.

"Research and experimental development (R&D) comprise creative and systematic work undertaken in order to increase the stock of knowledge – including knowledge of humankind, culture and society – and to devise new applications of available knowledge" (OECD, 2015, p.44).

Retail is a term that is generally associated with B2C sales. As market driven, companies that perform direct-to-consumer sales are having to either continuously innovate or adopt innovations as market orientation has been found to be, among others, a key driver for innovation in these specific companies (Pantano, 2014). In consequence, retailing, a highly competitive and fast-paced market, has been seeing remarkable changes recently. Past studies within the field of retail innovation have identified a range of complex and novel technologies that have the potential to disrupt the entire sector (Shankar et al, 2021), and thus improve the adopters' position in the market. The latest general technological trends show that there is a huge increase in innovation driving external factors, corporate social responsibility being one of them. CSR is an important external motivator for technological advancement. Increasing pressures from NGOs, regulators and various stakeholders are forcing decision makers to think more and more about adhering to United Nations' (n.d) Sustainable Development Goals (SDG) and to take into account externalities and aspects that have not been taken into consideration so far, which may raise costs, create inefficiencies and therefore increase the need for an increased operational efficiency and efficacy. Besides that, the internal factors weigh in at least as strong. According to the resource-based view (Mahoney,1992), firms often compete for limited resources in the environments they find themselves in, which are often stretched quite thinly (MacMillan & McGrath, 2002). Technology can be considered a resource in itself or it can increase access to other resources, finally increasing the competitive advantage. Shankar et al (2021) states this in the research conducted about types of retail technology. They argue that technology, specifically if it is developed in house, can be a great source of competitive advantage despite the high costs. The same authors classify these technologies as commoditizing or value adding, dimensions which can describe a succesful or unsuccesful position in the market. From an internal perspective, companies thus have to use innovations strategically in order to avoid opportunity costs (Palmer & Raftery,1999, Mahoney, 1992). This type of approach on research and development requires having a selective focus as it is bound to increase the chances of success (Mahoney,1992) and therefore reduces risks.

Just as almost everything in complex technological and organizational systems is nuanced, so is R&D and technology portfolio management. Mueller D.'s (1969) technological development cycle model argues that there are multiple ways in which companies can take advantage of technology's potential to generate return on investment. Spread over four cycles, the model presents an overview in relation to required R&D and payoff uncertainty, as well as industry output. However, this depends on their relative entrance position into the development cycle. To minimize payoff uncertainty and reduce potential losses and costs, industry players can look beyond their own operations, adopt an outside- in approach and enter technological development or even adoption at one of the early stages, such as the imitation stage, where the technological uncertainty is significantly lower. This significantly reduces the level of potential risk and thus ensures a higher chance of product, or service for that matter, to market deliverability. In the context of technology adoption and development, many other researchers debate the complexities of technological portfolio management, including Shankar et al (2021), which states that retailers' decisions on adoption of technology comprises of a set of issues, related to the previously discussed concepts and factors. Timing is the first one mentioned by the authors, followed by the monetary component, technology path, as well as sequence of action and execution. Moreover, the same research states that decision makers can also strive towards the bypassing of adoption barriers, which can offer a competitive edge.

All of these arguments suggest that a holistic and strategical approach based to decision making, based on thorough analyses of externally residing competencies and capacities is bound to increase the strategic situation and profitability, and topic modelling is a tool that can contribute towards that.

2.2 Technological Trends and Frameworks in Retail

There have been many researchers to systematically address the emergence and importance of innovative technology and their afferent particularities and contexts, specifically for large retailing companies, such as Grewal et al (2017), which comes up with a relevant framework of five key topic areas and their roles. The same author later, in 2021, provides a unique perspective with the introduction of a 6-piece model called "The Strategic Wheel of Retailing in the New Technology Era", each piece being a point of strategical importance for the introduction of technology. Although the author makes significant contributions to this research stream and introduces an overview of studies which identify technological advancements in the field, the methods only include customer interviews, surveys, databases, chatbot conversations and experiments.

A study of a crucial importance has been performed by Mostaghel et al (2022), which, based on a bibliometric analysis, analyses past innovations on the business model on the fond of digitalization, finding that the keyword "Innovation" being at the centre of the occurrence and co- occurrence maps. Other papers point out tangent factors, such as innovation drivers (Pantano, 2014) and the impact of technology on retail (Shankar et al, 2021). Bick et al (2022) identifies that a large section of the retail sector has initiated a technological transformation and are currently undergoing a major change. Two key dimensions are identified in this sense: technological architecture, with omnichannel integration, datafication and modernization, as well as operating model which should be product-led, automated and talent driven. These, among others, have been major contributions to the research stream and have brought forth, along with the theoretical contributions, a set of technological constructs or trends that are changing the industry (See Table 1. in Appendix section). They create a comprehensive idea of what the industry outlook and trends are, in terms of specific innovations as well as their purpose, level of impact, novelty, relevance and so on. This signals that this is a complex and heavily diverse environment that could bring about unexpected results. Given the LDA algorithm's need for topic name input, a certain degree of knowledge about specific trends and technologies, such as the aforementioned studies, is useful, as it would help assign correct and specific names for each cluster. These papers can therefore be used for the identification of specific technologies or technology systems and integrations.

Although such an extensive literature review is relevant so as to create a significant information base to successfully and correctly label the topics, throughout this paper only one of them is used extensively in order to facilitate a deeper understanding of specific technologies and their implications. Shankar et al (2021) provides in their research paper a comprehensive classification of retail related technologies. They comprise of newness (radical versus incremental), by IT relatedness (IT based vs non-IT based), by outcome (commoditizing e.g., price comparison, mobile apps, etc. versus value-adding), by nature of change (facilitating versus disruptive), and by stakeholder (employee, customer or supplier facing).

2.3 Text Mining Application in R&D Management

Unlike manual alternatives for text clustering, which are often subjective and inefficient, topic models are generative models which can be used to permeate large collections of unstructured documents. They work based on the assumption that each document is a mixture of topics and that there is a specific probability distribution for each word (Lee & Kang, 2017). Topic modeling, an exploratory text analysis technique, has been used by researchers as early as the late 1990's, with the Latent Dirichlet Allocation (LDA) being developed in 2003 (Giri, 2022). This means that LDA together with other types of topic modelling techniques have already been applied to a wide range of research in various domains. Unsupervised topic modelling, as opposed to the supervised topic modelling, is a versatile method that is used to make sense of unstructured text data without any previous information about it (Giri, 2022). This solves one of the major shortcomings of supervised learning, which is the requirement for labeled data. As previously mentioned, LDA has been successfully used in the identification of technological trends. Considering that this information lies outside of an organization, it is useful to corroborate this with internal R&D efforts in order to maximize organizational performance. Huang et al (2023) states that LDA is particularly efficient with semi-structured documents, a trait specific to patents, and comes with high accuracy, which finally helps conduct this research more effectively. Given that the retail sector has considerable patent databases, the Latent Dirichlet Allocation algorithm will be used. As for patents as a source of information in the context of R&D management, Ernst (1998) identifies that, as early as 1998, there have already been many voices pointing out the potential of patents in strategic portfolio management. He finishes his study by concluding that the usage of patent information continuously and strategically should become an essential part of planning activities for companies.

3. RESEARCH DESIGN

3.1 Data Collection

Patent submissions are a suitable source of information regarding any innovation and technology trends, including R&D efforts (Kim et al, 2015). The nature of those makes them an accessible and quantitative source can fulfill the requirements for a topic modelling analysis, with the LDA algorithm. Amazon, Walmart, 7-eleven, Target and Kroger are the US based retail giants that have been chosen for this. A closer look into these companies shows their impact over the market, with 942,25 billion USD sales in USA alone, in 2022 (NRF, n.d.). The data has been collected from the US patent registry, in the form of patent titles. Due to the platform's technical limitations, it was impossible to extract the text bodies, therefore just the titles have been used. The criteria for choosing the company was size, due to the fact that, as mentioned earlier, these companies have larger R&D budgets and are more likely to foster tangible results for the analysis, as well as the fact that the bigger the company the bigger the capacity to explain the overall situation in the retail industry.

Based on available data, 29641 patents have been granted in USA for the analysed companies, Amazon, Target, Walmart, 7-Eleven and Kroger. These companies make up 35.36% of the top 100 total yearly retail sales for 2022(NRF, n.d.). Out of the total patent number, one third has been extracted, namely 10048, out of which 3219 are valid for analysis, making up 10.86% of the total patent grant number. This is statistically significant and, by inference, it has the capacity of explaining real industry trends.

3.2 Text Preprocessing

The human input on this data is highly necessary in order to achieve the goal of the research. The first text alteration will be the conversion to lowercase, in order to have a consistent format. Words that do not add significant meaning to the corpus, such as articles, prepositions and pronouns, etc will be removed in order to facilitate the analysis, through a stop word removal function. They are usually not measured as keywords in text mining applications (Vijayrani,2015). The terms will then be tokenized, which is usually used to break up a sequence of strings into smaller pieces, which are, in the current context, words, and remove punctuation marks. Next, a lemmatization algorithm will be applied on the corpus. This function enables the searcher to not worry about different word variants and derivations, as they are included together (Korenius et al, 2004). It is also used to split compound words. All of these are used to filter out "noise" and to help conduct better analysis.

3.3 Latent Dirichlet Allocation

When compared to other topic modelling algorithms, LDA comes with several advantages, from case to case. Bastani et al (2018) identifies a set of algorithms used for similar purposes and provides an explanation for a comparison between four of them. The comparison starts by explaining the complexity and relevance of tf-idf algorithm, pointing out that it is not useful in the presence of a large number of unique words. LSA(Latent semantic analysis) is outlined by the same research as being somewhat effective in achieving the desired results, but lacks the capacity to fit the model into data in order to represent documents with multiple topics. When the pLSA (Probabilistic Latent

Semantic Analysis) iteration is introduced into discussion, the same author points out to another serious shortcoming, called overfitting, and thus it cannot be used for this research. Finally LDA is a solution to fix all of the shortcommings and fit the scope of topic modelling of patents. LDA assumes that the documents are represented as mixtures on top of emergent topics, in which topics are characterized by a word frequency distribution and their relevance to the topic (Blei, 2003). It is a generative probabilistic modeling approach that reveals hidden semantic structures from texts. The model defines $w_{d,n}$,$\forall n = 1, ...,N$; (black circle in fig. 1) $\forall d = 1, ..., D$ as a word, where D is a collection of M documents denoted by $D=\{w_1,w_2,\ldots,w_m\}$. Topics are β_k, $\forall k = 1, ..., K$, with the topic distribution per document θ_d, $\forall d = 1, ..., D$. (Bastani et al, 2018). The η and α are hyperparameters for prior distributions of β_k and θ_d. The variable $z_{d,n}$, with $\forall n= 1, ...,N$; and $\forall\ d = 1, ..., D$; is the topic assignment per word. Blei et al(2003) describes k, the dimensionality of the topic variable z, as an assumption for the number of topics. This is a crucial aspect of this research, as it is necessary to consider prior technology and innovation related research on the retail industry in order to establish this variable. The parameter α is a topic smoothing parameter, while η is a term smoothing parameter (Bastani et al, 2018). Attributing 0.1 to both parameters will ensure that the results are semantically significant. (Blei et al, 2003). As a result, the fewer the topics, the broader. A combination of theoretical inputs and algorithmic reiterations will be used to establish these dimensions.

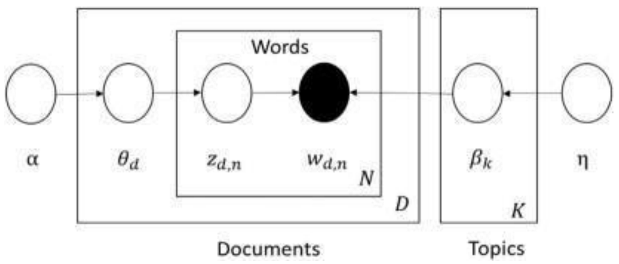

Figure 1. Latent Dirichlet Allocation visual scheme (Blei, 2003).

3.4 Post Processing

Chuang et al(2012) introduce a tool to rank and visualise the LDA term distributions, based on two measures: distinctiveness and saliency, based on which Sievert et al's (2014) LDAvis is constructed. The latter argue that visualising the clusters solely based on their probability under a topic is suboptimal. These are two key components that make a valuable addition to the LDA algorithm in the combination of two visualisation systems, thus making the analysis clearer and more effective to interpret. Based on this, an interface is created.

The terms within each topic are ranked based on a relevance coefficient (Sievert et al, 2014). The authors propose a ranking coefficient that considers the frequency of the term under the topic, along with the exclusivity of a term within the topic, relating to its appearance in relation to other topics' exclusion. Relevance is defined as the following: $r(w,k\mid\lambda) = \lambda \log(\phi kw) + (1-\lambda)\log(\phi kw/pw)$, where ϕkw is P(w|k) which is defined as $w \in \{1, ..., V\}$ for topic $k \in \{1, ..K\}$, where V is the number of total words in a vocabulary, and p_w is the marginal probability of term w in the document P(w) , and λ is the weight given to the probability of w under topic k relative to its lift (on the logarithmic scale). The lift is defined as the ratio of the term's probability in a topic to its marginal probability in the corpus. Finally, this coefficient will enable the viewing of the words in a topic in a descending order, with the most important words on top. This is most useful to analyse the topics on an independent basis. A more abstract metaphor for the relevance metric, that can help the reader better visualise this, is a magnifying glass. The closer the coefficient is to the maximum value, which is 1, the more far away the perspective is and the higher the possible correlation or similarities between the topics, due to the removal of the p_w coefficient. If λ has the directly opposite extreme value of 0, it emphasises heavily on the topic "individuality". The results are visualised with Sievert et al's (2014) "LDAvis" method, which presents itself as a global topic interface, with a matrix containing the topic bubbles, spread over four quadrants, as well as the frequency per topic and for the overall frequency of the corpus on the right side of the interface. The red bars represent topic specific frequencies for each term while the grey bars represent the corpus specific frequencies of each term. (fig. 2, Appendix)

4. RESULTS

4.1 Topic Interpretation

Applying the algorithm on the patent corpus is predicted to result, due to the nature of the database, in a set of words spread over a relatively restricted background, related to the paper's topic. Regardless, human input is still required. Setting the number of topics is done through trial and error. This implies the success/failure of an application of the method, based on a thorough analysis of the results and a set of reiterations. Subsequently, number of 20 topics has been chosen. The results come in the form of cross-document word clusters, with a probabilistic distribution over topics. The topic's validity will be analyzed on the basis of the composing words. Theoretical input will be necessary here in order to sort through the ambiguity of the clusters to finally attribute a topic title or an overarching theme to the topic. This is based on a broad aggregate of previous technology and innovation research on the field of retail. Based on the outputs of the analysis, the results will then be associated with specific keywords or innovation classes (Table 1. Appendix). As specified earlier, the most useful theoretical contribution for this is Shankar's et al (2021) classification of retail technologies, which points out five dimensions specifically useful in removing the ambiguity of the cluster. The resulting analysis can be used as a reference point for technologies present in the retail markets. The key point of this analysis is the λ metric that was presented in 3.4. Established through trial and error, it will offer the capability of analysing vague topics or topics with a high similarity to others, based on a lower lambda which showcases the topics' uniqueness, on a case by case basis.

4.2 Thematic structure of Topics

4.2.1 Digital inventory video management system

The first topic to be analysed is Topic Nr. 5. This is the first as it is the most isolated term, therefore having the most unique characteristics which is also easier to interpret. Because of that, a high value of λ = 0.85 has been chosen. This does not filter out any commonalities it might have with other topic as that is not needed, due to its inherent uniqueness. The most prevalent words are "display", "fixture", "device", "inventory", "system", "video", "interface" and "audio". This can be an in-store video-based inventory management system, for employees. The word "interface" clearly denotes the IT characteristics of it. Video recording has been around for a while, therefore this is an incremental innovation, through the linkage of this with other technologies. It is a facilitative technology, as it improves on already existing elements and infrastructure.

4.2.2 Digital Inventory filtering and sorting system

Topic 12 seems to be unique as well. Because of that, a high relevance coefficient of 0.6 can be used to take into consideration a higher proportion of its characteristics. Keywords such as "chair", "multi", "load", "element", "electrowetting", "motion", "recognition", "message", "stand", "stocking", "balancing", "reflective", "layer", "support", "navigation", "automatic" and "interface" are present. This can be another in-store type of technology, possibly employee facing due to the combination of "stocking", "support", "interface". Similar, and possibly even complementary to the previous topic, this is possibly a commoditizing and facilitating technology that improves on already existing processes, incrementally.

4.2.3 Mobile inventory querying device

Topic 18 (λ = 0.6), with the same coefficient due to a similar position, has words such as "detection", "device", "automatic", "holder", "medium", "assembly", "monitoring", "query", "object". It is similar in this way to the previous two technologies. This can be regarded as a warehouse related technology, made to improve on logistical and operational processes such as assembly and organization of the warehouse, digitally ("automated", "device"). This seems to be IT related, and with characteristics close to the previous two topics, which could be attributed to their similar positioning on the top side of the topic matrix. This innovation is employee facing, with the scope of reducing costs and making existing systems more efficient.

4.2.4 Automated checkout system

On the opposite side of the matrix, Topic 8 presents itself with the following terms: "shopping", "cart", "store", "checkout", "verification", "determination", "physical", "system", "method", "automated". This is yet another in store technology, but customer facing ("shopping", "cart", "store"). It can be related to an automated system which scans products from the customer's shopping cart. This topic has been analysed with a λ of 0.6. This is a disruptive, IT based technology which replaces cashiers and has the capacity of changing the retail sector. The radical change is the automation of the final stage of in-store purchases, which completely changes the customers' experience, as well as drastically reducing costs for the organization

4.2.5 Order and shipment processing system

Topic 4 is surrounded by Topics 8, 1, 9 and to some extent, 20. This is why a smaller λ of 0.24 seems to be showing better results. The terms "apparatus", "order", "container", "hanger", "sortation", "fulfillment", "providing, "warehouse", "recommendation", "process" are visible. This is a logistic based solution which could be specifically related to the processing of large orders ("container") from warehouses. It can be a physical sortation machine, or a software solution. It is a supplier or employee-facing technology, that may or may be related to already existing technologies, as an incremental improvement, or may be portraying a stand-alone automated sortation system, which would be a radical and disruptive technology, due to the high cost of operating warehouses (i.e. the replacement of warehouse workers).

4.2.6 Ecommerce sortation system

The topic 1(λ = 0.5) terms are "method", "system", "management", "product", "identifying", "user", "mobile", "device", "delivery", "search". This could be yet another customer-oriented IT based technology for product search and delivery. This is common nowadays due to the rise of e-commerce ("mobile", "device") and upon its introduction could have been considered a radical and disruptive technology.

4.2.7 Smart Doorbell System

Topic 9(λ = 0.25) is best characterised by "video", "time", "real", "stream", "mountable", "identification", "imaging", "continuous", "doorbell". This combination of terms can best describe a product rather than an interation or integration, which does not necessarily qualify on all dimensions from Shankar et al's (2021) framework, as it is not directly related to retail as a process, except for the fact that it is IT based and value adding.

4.2.8 Shelf Display Units

Topic 20 is a topic that could refer to shelf display units, which are new in-store technologies that are directed at consumers and are meant to act as an on-shelf marketing component ("shelf", "unit", "display"), but can also refer to in- store robot advisors ("conversational"), with a possible specificity to pharmacies ("pharmacy" being an almost exclusive term to this topic). This is a value adding technology, due to the extra information that it can offer about the products, integrated on current systems, and widely used technologies therefore facilitating and incremental. The relevance metric used for this is 0.42.

4.2.9 Marketplace AI system

Topic 17 is clearly referring to one of the AI systems that have been emerging in the recent past. "Learning", "machine", "model" point at this. "Inference", "population", "training" denote that this IT based model is trained with user generated data, which therefore can have a user-oriented component and service ("generating", "retail") through the implementation of recommendations and item feeds. Terms such as "seller" point out that this is also used to the retailer's advantage (e.g., portfolio and product positioning, pricing, business intel etc.) on online marketplaces. This is a radical technology which has the potential to revolutionise the way retail works. It can be argued that it is radical and new, which can be seen in the recent AI recommendation systems, which bring value to all of the parties involved by collecting user data and creating a custom experience for each customer. The characteristics of this topic can be seen under a relevance metric of 0.4.

4.2.10 Voice assistant speech recognition integration

In topic 10, the terms "processing", "speech", "position", "scalable", "system", "tracking", "table" are bringing about a vague description of a speech recognition item. This can be possibly related to already existing voice assistant products. Not many things can be said about this topic, as it is relatively close to other topics and increasing the relevance metric from a value of 0.29 to, for example, 0.57, would offer some more insights, but those would still be vague and with a low topic specificity proportion (e.g., "system", "data", "content"). As such, it is safe to consider this as being an incremental software innovation, as software is often rather abstract, which makes use of already existing systems and resources in order to either provide value or improve the operational function of a product or service.

4.2.11 Autonomous vehicle

Topic 3 comes with keywords such as "data", "network", "based", "dynamic", "route", "vehicle", "automated", "aerial", "traffic". The essence of topic 3 is best captured by a λ value of 0.53. This is an IT based technology related to autonomous vehicles, conventional or even airborne, which uses networks and data to manage its routes, planning and capacities. This can be considered highly disruptive technology, radical by nature, which has the potential of changing the way retail is being done, through the replacement and automatisation of cost and labor intensive processes, in a more efficient way. It could relate to anything such as a product, a service or even a process augmentation. It seems to be value adding, and could be anything from customer facing to employee facing and supplier facing.

All of the above topics have been relatively straight forward and did not require a large number of trial and error reiterations and comparisons. Although based on a somewhat subjective assessment, this is an effective and efficient way to generate useful results in a qualitative and fast manner. The next 9 topics that are about to be discussed are another story and could require a higher number of relevance alterations, due to their proximity to eachother, on the matrix. This proximity denotes a high degree of similarity in between the topics. This is why smaller λ values will be used, to effectively capture their essence.

4.2.12 Voice recognition security system:

Topic 2, with terms such as “device”, “network”, “voice” and “security” is rather self-explanatory. The distribution of the terms is rather accentuated under the 0.2 relevance metric, highlighting the unique characteristics of the topic. They could relate to a voice recognition security system, which can be either a product in the retailers’ portfolio or relate to an in-house developed system that a company might use for their own security. IT based, it can be argued that it commoditizes or facilitates already existing systems and processes incrementally.

4.2.13 Virtual management system

Topic 14 can be best observed under a λ of 0.38, with the following prevailing terms: “provider”, “image”, “camera”, “central”, “docket”, “substrate”, “execution”. This, by itself, does not say much. However, zooming out to a λ of 0.68 will give out more terms such as “tracking”, “virtual”, “management”, “service”, “network”. Although not highly specific, as the proportion of topic frequency to overall frequency is quite low, it shows that this could be a service- related technology, based on a system or network of cameras that are used for accurately tracking, computing, controlling and managing (“compute”, “management”, “cloud” etc.). This, as many others, is an IT based technology that combines a set of technologies in order to provide value and make processes more efficient, incrementally. It can be regarded as a facilitative type of technology. Based on the fact that it is rather vague, it will be assumed that this is a software related integration or improvement.

4.2.14 Cloud Inventory Management system

With a λ of 0.41, topic 15 gives out terms such as “storage”, “database”, “remote”, “access”, “service”, “server”. This could be a planning and inventorying web based application that can allow employees to manage and monitor storage remotely. This is a radical innovation, as it enables the absence of management in warehouses and enables higher flexibility of work. The keyword “client” gives out that this may be integrated in further systems or even be a customer facing innovation. Although not much more can be concluded about the keywords, it can be assumed that this, when paired with other previously mentioned technologies, can create a system that can greatly change industry specific processes.

4.2.15 Real-time geolocation system

Topic 16 is the most centrally positioned topic bubble and, with a moderate relevance metric of 0.4, it describes a sensory based real time analysis technology with terms such as “adaptive”, “metric”, “sensor”, “mapping”, “reporting”. This is a highly abstract technological integration that can not be further described, other than the fact that it is IT based. Emphasizing on term exclusivity will reveal words such as “adaptive”, “coordinate”, “personal”, which takes towards a geolocation type of technology. This can be an employee facing technology, or even an in-store analysis software system.

4.2.16 Sensor routing system

Topic 11 is the closest to the previously mentioned one. A lambda relevance value of 0.25 does not reveal much technological insights, with “natural”, “routing”, “language”, “update” and “mode” dominating the distribution. A slight increase to 0.35 reveals some interesting keywords, such as “determining, “sensor”, “structure” and “array”, which, compared could be a technology similar to the previously mentioned topic 16’s geolocation innovation, but sensor based.

The remainder of the topics, 7, 19, 13 and 6 present the highest concentration of topics out of the entire analysis. The only insight that can be extracted from topic 7 is that it is a footwear related technology. Any other variations do not have any technological related output. Topic 19 (UI experience innovation) is clearly a web-based integration. A moderate relevance score determines a clear and straight forward output “page”, “purchase”, “feedback”, “experience”, “multi”, “session”, “scanning”, “tenant”. This is an IT, web-based integration with possible links to online retailing or e-commerce, which has become more prevalent in the past years. This is a technology that possibly improves user experience on the website and can therefore be considered an incremental, commoditizing, facilitative and shopper-facing innovation.

A quick visual scan of topics 13 and 6 denotes that they are heavily related to each other. Based on a quick analysis of topic 6(web data integration) under a 0.25 relevance metric, one could conclude that based on terms such as “rule”, “authentication”, “lifecycle”, this is yet another web integration, just like topic 19, and is assumed to have the same characteristics (software, incremental, commoditizing, facilitative and shopper facing) due to the lack of further information. The same can be said about cluster number 13(web data analysis algorithm), based on keywords such as “environment”, “platform”, “protocol”, “connection”, “private”, “asynchronous”.

5. DISCUSSION

This study has provided the reader with a comprehensive overview of technological and innovation trends from within the retail industry. A large number of patent titles have been used in order to create an overarching picture of these trends. An output of 20 topics has been generated, each pertaining to different technologies and their types. As it can be seen in Figure 2 (Appendix), the overall distribution of the topics is relatively skewed, or clustered. Therefore, it was predicted that the resulting analysis of technologies would be just as skewed. More precisely, it was predicted that many of the topic themes would be very similar, if not almost identical. This would have created a biased perception of the technological outlook of the retail industry that would not provide much value to foresighters. However, with the use of the “magnifying glass” coefficient, it was possible to succesfully distinguish between them, which has resulted in diverse visible technological solutions that covers many operative areas and aspects of the retailing process. The distribution of the results within the scope of the retail technological classification system has been somewhat normal, with few objections. The presence of both incremental and radical technologies can be observed. The same can be stated about the distribution of the results in terms of the outcome dimension, as well as the stakeholder dimension. However, the stakeholder dimension has shown a low presence of supplier-facing technologies, which could indicate a gap to be explored by market players (which could foster some advantages). The results also showcase an almost exclusive presence of IT based solutions, with many of them being purely IT based sofware related innovations. The presence of a few highly disruptive technologies is to be noted, such as artificial intelligence and autonomous vehicles.

5.1 Theoretical Implications

In this section, the significance of the findings will be analysed in relation to existing research from within the academic community. It can be safely stated that the usage of Shankar et al's(2021) framework has been useful for providing a basis for analysing the results. Most of the results, with few exceptions, have been successfully explained or characterized using the researchers' technological classification framework, using all but one of the six framework characteristics, namely the domain span and source of origin dimension, which needs further context in order to be used. This dimension can be used in a possible in-depth analysis of a single company, in combination with more information, which this is not the case for this research. Like the technology model, so has the aggregate of studies presented in Table 1.(Appendix) been useful, where various technologies are named punctually, as they have provided a thorough knowledge base for the successful identification of technology with the LDA analysis (e.g., ecommerce, data, analytics, driverless vehicles, automatic checkout, inventory management, cashierless store, inventory robots etc.). Moreover, the resulting technologies and classes could be used to create further descriptions and patterns to describe the industry. As an example, it can be observed that software technologies surface quite often and are most likely to be described vaguely. Although it is not easy for the untrained eye to identify them, it serves to confirm that the retail industry has become highly technological.

Sievert & Shirley's (2014) visualisation system has proven exceptionally useful in order to achieve the scope of this goal. As an integration of two previous researches, the LDAvis analysis is spearheaded by the relevance metric. The formulas integrating the relevance coefficient are making this technique particularly powerful, as they allow for the individual analysis of similar and often overlapping topics, which, given their absence, would have created a biased analysis. Consequently, although many researchers have been using LDA as a means to spot technological trends or analyse the retail industry (Gurrola-Perez et al, 2022), the combination of the two is apparently completely missing from the academic discourse. Saritas et al's(2021) paper uses another, lesser known semantic analysis technique, but only uses trends in mobile commerce, a tangential, yet different sector. This serves to bring forth the uniqueness of this research. On top of that, the available researchers do little to relate their research to existing literature about various innovation trends and frameworks, such as Shankar et al (2021), or any of the other listed literature pieces that refer to technological innovations used by retailers, in order to create actionable inputs for real-world cases. As opposed to other technological forecast attempts, such as Jun et al (2012), which states that some of the weaknesses of such techniques include translation to economic viewpoints, which this research tackles, to some extent, using Mueller's (1969) technology development stages model. Another shortcomming of the same study includes the need for a high degree of domain experience, which is also discussed and somewhat dealt with in this paper (Jun et al, 2012; Chen et al, 2017)

To conclude, the combination of these elements have enabled a successful contribution to the academic discourse through the explorative identification of real trends and have shown a successful implementation of cross disciplinary elements.

5.2 Practical Implications

This section will be used to discuss about how this study contributes knowledge practically, to decision makers from various organizations as an intelligence resource in real-world situations. As mentioned previously, a semantic analysis on patent databases can provide actionable information if used in the right context. In an environment where decision making capabilities are seriously lacking (PWC, 2014), this study aims to alleviate that problem.

One particularly useful R&D framework is Mueller's (1969) technological development cycle which provides context for managers. A thorough understanding and usage ofthis framework (fig 3. Appendix) can aid organizations achieve a successful strategic positioning in the market and its afferent emerging technological solutions. Benefits include the removal of entry barriers, mitigations of risks and uncertainties. Overall, the methodology can be used in a multitude of ways, including a year-by-year analysis, based on patent submissions dates, or to conduct large scale analysis of the industry, for specific geographical areas and even on a company-by-company basis.

Using a high relevance coefficient can offer an overall picture of the industry, with frequent cross-topic, corpus wide keywords such as "system", "interface", "video", "device", "network", "service", "data", "electronic" and "user". Based on this, a decision maker can infer that the overall technological environment is service or process based (customer and employee facing), that many of the solutions are software related and that the rise of IoT (Mostaghel et al, 2022) technologies is empirically observed. Due to the high overall frequency of the terms pointing to the technologies, it can be assumed that they subside in the later stages of Mueller's (1969) development stage, namely the competition stage and even in the standardization stage. The particularities of these later stages is the removal entry barriers through the presence of extensive data and ecosystems around them, as well as the low degree of risks and uncertainties. Subsequently, using a small relevance coefficient offers a more detailed view into more specific, less prevalent technologies that do not yet have a big impact on the market, with few clusters even hinting at recent products launched by retail giants, such as Amazon's Alexa (Peeters, n.d.) (see topic 10) and Blink or Ring smart doorbell systems (Tuohy, 2021) (topic 9), cloud computing and web services (AWS, n.d.) (topic 15), or Walmart's self-checkout machines (Keith, 2021) (topic 8). Given the relatively low frequency of the composing words of these topics, it is safe to assume that these are emerging innovations, either in the innovation stage or the initiation stage. This is supported by data evidence, as the named topics have either been successfully analyzed under a relatively low λ (topic 10:0.29, topic 9:0.25, topic 15:0.41) or have an isolated position on the matrix (topic 8, fig. 2, Appendix). These innovations are uncertain, with higher associated risks, which would prompt management to take a more cautious and possibly explorative approach, in order to mitigate potential losses. On the other hand, if the presence of capital in the organizations is sufficient, the managers could strive towards a more resource intensive approach, given a high-level relative strategic importance of the resulting capacities and technologies Mahoney (1992) which would entail higher risks but also higher rewards.

Considering a situation in which a company would have to create or manage a new or dynamic R&D portfolio, whether due undertaking a horizontal shift of its activity, strategic expansion or even just market capitalisation growth, based on the findings of this study, a manager would, as an example, attribute a higher amount of resources to technologies and technology classes which are visible in the corpus on a global level, either in specific topics, with a higher frequency or a higher lambda (e.g., topic 5, λ = 0.85, digital inventory video management system; IoT technologies). These solutions can be considered a safer option with lower risks and uncertainties.

5.3 Limitations and Further Research

This study has faced a certain degree of limitations, particularly due to the poor and inconsistent manner in which patents are stored and organized in the public database. A significant setback for this study has been the incapacity to extract the patent text bodies due to the platform's inability of giving out a full output of patents. Therefore, the patents' titles have been used, which are often in the form of numbers, thus making them unfit for analysis and significantly lowering the quantity of resulting valid patents.

Despite the fact that the visualisation system has enabled the study of the results efficiently and effectively, it is necessary to acknowledge the afferent drawbacks. Subjectivity is one important drawback to be associated with it. During the analysis of the topics, a certain degree of subjectivity has been undertaken, especially given the introduction of a new variable, the relevance coefficient, which creates variation and only serves to increase this subjectivity, despite its overall positive effects on the study. Conclusively, the analysis is highly impacted by the forecaster's previous knowledge or biases towards industry specific technologies and classifications, which could wrongfully twist the results in many ways.

Throughout this study, only patents filed in the USA have been used. Although the United States have highly mature and developed markets, focusing exclusively on this area does not reflect the complete geographical spread of the entire retail industry. An important competitor in this case is the European Union that, with an accessible and extensive patent database and fast paced consumer markets, can provide extensive information that would further help to complete the overall picture about the retail industry. Another significant but emerging market that bound to provide valuable information on the matter is the Chinese one. Although not that easily accessible, due to various reasons, including language barriers, China has long been known as a production powerhouse, and is due to produce important results in the case of a LDA patent analysis, due to the recent economic developments that the country has been undergoing. Consequently, this can serve as a starting point for future studies, more specifically, on the two mentioned markets. Other possibilities for further works include the discussion or improvement of the interpretation frameworks in order to further remove some of the subjectivity of the interpretation, either quantitatively or qualitatively, such as elaborating more on how Shankar et al's (2021) framework can serve to improve decision making for managers rather than serve as a tool solely to remove ambiguity from the results, or even creating a standardised decision process using either the already mentioned constructs or by bringing new frameworks and theories.

5.4 Conclusion

To conclude, this paper provides a successful contribution to the trends pertaining to the retail industry particularly for R&D portfolio management. This is achieved through the combination of the powerful and versatile LDA algorithm (Blei, 2003), in combination with the LDAvis (Sievert & Shirley, 2014) visualisation system, thus resulting in a effective and efficient way of analysing technology and innovation trends of retailers. Coupled with the framework presented by Shankar et al (2021), which successfully solves one of the drawbacks of LDA and placed into real world context with Mueller's (1969) technological development cycle, this results into actionable information that can be used by various industry actors such as R&D portfolio managers, product managers, government officials etc. to analyse the technological development and outlook within the retail sector and take strategic decisions such as resource allocation, technological development paths, product and portfolio positioning,etc.

6. REFERENCES

1. Sievert, C., & Shirley, K. (2014, June). LDAvis: A method for visualizing and interpreting topics. In Proceedings of the workshop on interactive language learning, visualization, and interfaces (pp. 63-70).

2. Tang, J., Wang, B., Yang, Y., Hu, P., Zhao, Y., Yan, X., ... & Usadi, A. K. (2012, August). Patentminer: topic-driven patent analysis and mining. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1366-1374).

3. Mansfield, E. (1986). Patents and innovation: an empirical study. Management science, 32(2), 173- 181.

4. Kwon, K., Jun, S., Lee, Y. J., Choi, S., & Lee, C. (2022). Logistics technology forecasting framework using patent analysis for technology roadmap. Sustainability, 14(9), 5430.

5. Kim, G. J., Park, S. S., & Jang, D. S. (2015). Technology forecasting using topic-based patent analysis.

6. Council on Foreign Relations. (n.d.). Global monetary policy tracker. Council on Foreign Relations. Retrieved April 16, 2023, from https://www.cfr.org/global/global-monetary-policy-tracker/p37726

7. Buder, F. (2021). The Value of Foresight in a VUCA world - Results from a Survey of Organizational Foresight Capacity.

https://doi.org/10.13140/RG.2.2.34439.52646

8. Decision making in retail- PWC. (n.d.). Retrieved April 16, 2023, from https://www.pwc.com/gx/en/issues/data-and-analytics/big-decisions survey/industry/assets/retail.pdf

9. Reinartz, W., Dellaert, B., Krafft, M., Kumar, V., & Varadarajan, R. (2011). Retailing innovations in a globalizing retail market environment. Journal of Retailing, 87, S53-S66.

10. Mahoney, J. T., & Pandian, J. R. (1992). The resource-based view within the conversation of strategic management. Strategic management journal, 13(5), 363-380

11. Grewal, D., Gauri, D. K., Das, G., Agarwal, J., & Spence, M. T. (2021). Retailing and Emergent Technologies. Journal of Business Research, 134, 198–202. https://doi.org/10.1016/j.jbusres.2021.05.004

12. The gap between large and small companies is growing. why? Harvard Business Review. (2021, January 20). Retrieved April 17, 2023, from https://hbr.org/2019/08/the-gap-between-large-and- small-companies-is-growing why?registration=success

13. Silva, C. C., Galster, M., & Gilson, F. (2021). Topic modeling in software engineering research. Empirical Software Engineering, 26(6). https://doi.org/10.1007/s10664-021-10026-0

14. Kherwa, P., & Bansal, P. (2018). Topic modeling: A comprehensive review. ICST Transactions on Scalable Information Systems, 159623. https://doi.org/10.4108/eai.13-7-2018.159623

15. McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. Harvard business review, 90(10), 60-68.

16. Pantano, E. (2014). Innovation drivers in retail industry. International Journal of Information Management, 34(3), 344-350.

17. Chesbrough, H. W. (2003). Open innovation: The new imperative for creating and profiting from technology. Harvard Business Press.

18. MacMillan, I. C., & McGrath, R. G. (2002). Crafting R&D project portfolios. Research-Technology Management, 45(5), 48-59.

19. Huang, L., Hou, Z., Fang, Y., Liu, J., & Shi, T. (2023). Evolution of CCUS Technologies Using LDA Topic Model and Derwent Patent Data. Energies, 16(6), 2556.

20. Ernst, H. (1998). Patent portfolios for strategic R&D planning. Journal of engineering and technology management, 15(4), 279-308.

21. Bastani, K., Namavari, H., & Shaffer, J. (2019). Latent Dirichlet allocation (LDA) for topic modeling of the CFPB consumer complaints. Expert Systems with Applications, 127, 256-271.

22. Bick, R., Skinner, V., Keutel, M., & Bout, S. (2022, May). The tech transformation imperative in retail. www.mckinsey.com. Retrieved April 18, 2023, from https://www.mckinsey.com/~/media/mckinsey/industries/retail/our%20insights/the%20tech%20transformation%20imperativ e%20in%20retail/the-tech- transformation-imperative-in-retail.pdf

23. Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. Journal of machine Learning research, 3(Jan), 993-1022.

24. Chang, M. H., & Harrington, J. E. (1998). Organizational structure and firm innovation in a retail chain. Computational & Mathematical Organization Theory, 3, 267-288.

25. Choi, D., & Song, B. (2018). Exploring technological trends in logistics: Topic modeling-based patent analysis. Sustainability, 10(8), 2810. https://doi.org/10.3390/su10082810

26. Giri. (2022, February 4). Topic modeling with LDA explained: Applications and how it works. HDS. Retrieved April 17, 2023, from https://highdemandskills.com/topic-modeling- intuitive/?utm_content=cmp-true

27. Grewal, D., Roggeveen, A. L., & Nordfält, J. (2017). The future of retailing. Journal of Retailing, 93(1), 1–6.

https://doi.org/10.1016/j.jretai.2016.12.008

28. Johri, V., & Bansal, S. (2018). Identifying trends in technologies and programming languages using topic modeling. 2018 IEEE 12th International Conference on Semantic Computing (ICSC). https://doi.org/10.1109/icsc.2018.00078

29. Mostaghel, R., Oghazi, P., Parida, V., & Sohrabpour, V. (2022). Digitalization Driven Retail Business Model Innovation: Evaluation of past and avenues for future research trends. Journal of Business Research,146, 134–145. https://doi.org/10.1016/j.jbusres.2022.03.072

30. Mueller, D. C., & Tilton, J. E. (1969). Research and development costs as a barrier to entry. The Canadian Journal of Economics/Revue canadienne d'Economique, 2(4), 570-579.

31. OECD. (2015). Frascati Manual 2015 - OECD- Ilibrary.org. Concepts and definitions for identifying R&D. Retrieved April 18, 2023, p.44, from https://www.oecd-ilibrary.org/docserver/9789264239012-4-en.pdf?expires=1593711163&id=id&accname=guest&checksum=EFF6091D4C87D47BC0E111193155ACC 9

32. Palmer, S., & Raftery, J. (1999). Opportunity cost. Bmj, 318(7197), 1551-1552.

33. Person. (2021, July 3). What's causing the Global Supply Crunch? Supply Chain Magazine. Retrieved April 16, 2023, from https://supplychaindigital.com/procurement/whats-causing-global-supply-crunch

34. Saritas, O., Bakhtin, P., Kuzminov, I., & Khabirova,

E. (2021). Big Data augmentated business trend identification: The case of mobile commerce. Scientometrics, 126(2), 1553–1579. https://doi.org/10.1007/s11192-020-03807-9

35. Shankar, V., Kalyanam, K., Setia, P., Golmohammadi, A., Tirunillai, S., Douglass, T., ... & Waddoups, R. (2021). How technology is changing retail. Journal of Retailing, 97(1), 13-27.

36. Tidd, J., & Bessant, J. R. (2020). Managing innovation: integrating technological, market and organizational change. John Wiley & Sons.

37. United Nations. (2021, August 9). IPCC report: 'Code red' for human driven global heating, warns UN chief. Climate and Environment.

38. United Nations. (n.d.). Sustainable development goals: 17 goals to Transform Our World. United Nations. Retrieved April 18, 2023, from https://www.un.org/en/exhibits/page/sdgs-17-goals- transform world#:~:text=Sustainable%20Development%20Goals%3A%2017%20Goals%20to%20Transform%20our,empower%20all

39. Chuang, J., Manning, C. D., & Heer, J. (2012, May). Termite: Visualization techniques for assessing textual topic models. In Proceedings of the international working conference on advanced visual interfaces (pp. 74-77).

40. Top 100 retailers 2022 list. NRF. (n.d.). https://nrf.com/resources/top-retailers/top-100- retailers/top-100-retailers-2022-list

41. Value added by industry: Retail trade as a percentage of GDP. FRED. (2023, March 30). https://fred.stlouisfed.org/series/VAPGDPR

42. Council on Foreign Relations. (n.d.). Global monetary policy tracker. Council on Foreign Relations. Retrieved April 16, 2023, from https://www.cfr.org/global/global-monetary-policy-tracker/p37726

43. Yoon, J., Park, H., & Kim, K. (2013). Identifying technological competition trends for R&D planning using dynamic patent maps: SAO-based content analysis. Scientometrics, 94, 313-331.

44. Kozinets, R. V., Hemetsberger, A., & Schau, H. J. (2008). The wisdom of consumer crowds: Collective innovation in the age of networked marketing. Journal of macromarketing, 28(4), 339- 354.

45. Vijayarani, S., Ilamathi, M. J., & Nithya, M. (2015). Preprocessing techniques for text mining-an overview. International Journal of Computer Science & Communication Networks, 5(1), 7-16.

46. Korenius, T., Laurikkala, J., Järvelin, K., & Juhola,

M. (2004, November). Stemming and lemmatization in the clustering of finnish text documents. In Proceedings of the thirteenth ACM international conference on Information and knowledge management (pp. 625-633).

47. Gurrola-Perez, P., Lin, K., & Speth, B. (2022, September). Retail trading: an analysis of global trends and drivers. world-exchanges.org. https://www.world-exchanges.org/storage/app/media/WFE-Retail-Investment%20Sep%2020%202022.pdf

48. Peeters. (n.d.). What Is Alexa?. Amazon. https://developer.amazon.com/en-US/alexa

49. Tuohy, J. P. (2021, October 1). Why does Blink still exist?. The Verge. https://www.theverge.com/22704290/amazon-blink-ring-camera-doorbell-brands-smart-home-why

50. Cloud computing services - amazon web services (AWS). (n.d.). https://aws.amazon.com/

51. Keith. (2021, December 26). Walmart is looking to remove all cashiers from stores, going fully self checkout. Positively Osceola. https://www.positivelyosceola.com/walmart-is-looking-to-remove-all-cashiers-from-stores-going- fully-self-checkout/

52. Wolpert, S., & Roth, A. (2020). Development of a classification framework for technology based retail services: A retailers' perspective. The International Review of Retail, Distribution and Consumer Research, 30(5), 498-537.

53. Varadarajan, R., Srinivasan, R., Vadakkepatt, G. G., Yadav, M. S., Pavlou, P. A., Krishnamurthy, S., & Krause, T. (2010). Interactive technologies and retailing strategy: A review, conceptual framework and future research directions. Journal of Interactive Marketing, 24(2), 96-110.

54. Jun, S., Park, S. S., & Jang, D. S. (2012). Technology forecasting using matrix map and patent clustering. Industrial Management & Data Systems.

55. Chen, H., Zhang, G., Zhu, D., & Lu, J. (2017). Topic-based technological forecasting based on patent data: A case study of Australian patents from 2000 to 2014. Technological Forecasting and Social Change, 119, 39-52.

7. APPENDIX

Authors	Title	Context	Keywords for Technologies and Innovation Themes
Shankar et al, 2021	How Technology is Changing Retail	Classification of technologies by stakeholder, it relatedness, source, novelty, nature and outcome	AI, GPS, IoT, VR, AR, MR(mixed reality), 5G, ecosystem, mobile devices, chatbots, smart mirrors, payment technologies, blockchain
Mostaghel et al, 2022	Digitalization driven retail business model innovation: Evaluation of past and avenues for future research trends	Aggregate dimensions of research bibliography: business model innovation, value creation, delivery and capture	IoT, VR, robots, network, ecosystem, data, digital service, digital platform, customer data, knowledge collection,
Pantano, 2014	Innovation drivers in retail industry	Technology by innovation difusion	mobile apps, self-service, digital signage, ubiquitous stores, ubiquitous computing
Grewal et al,2021	Retailing and emergent technologies	Strategic Wheel of Retailing in New Technology Era	Chatbot, Data, online
Grewal et al, 2017	The Future of Retailing	Organizing framework for retailer innovation interest areas	Big data, Analytics, robots, drones, driverless vehicles, internet of things, machine to machine commerce, virtual reality, augumented reality, virtual shows,
Bick et al, 2022	The tech transformation imperative in retail	Architectural versus operating model	omnichannel integration, ecommerce, cloud, automation, microservice, SDLC, CI/CD, data,
Reinartz et al, 2011	Interactive Technologies and Retailing Strategy: A Review, Conceptual Framework and Future Research Directions	Review of research in interactive technologies	Image interactivity tech (IIT), RFID, IVR (interactive voice response)
Wolpert &Roth, 2020	Development of a classification framework for technology based retail services: a retailers' perspective	Technology based retail services literature review	Automatic checkout, Inventory management, Cashierless store, Checkout tunnel, Customer price discrimination, Digital receipt, automatic pricing, personal shopping assistant, information terminal, intelligent scale interactive kiosk system, inventory robots, mobile coupons, loyalty program, smart cart, vending machine,

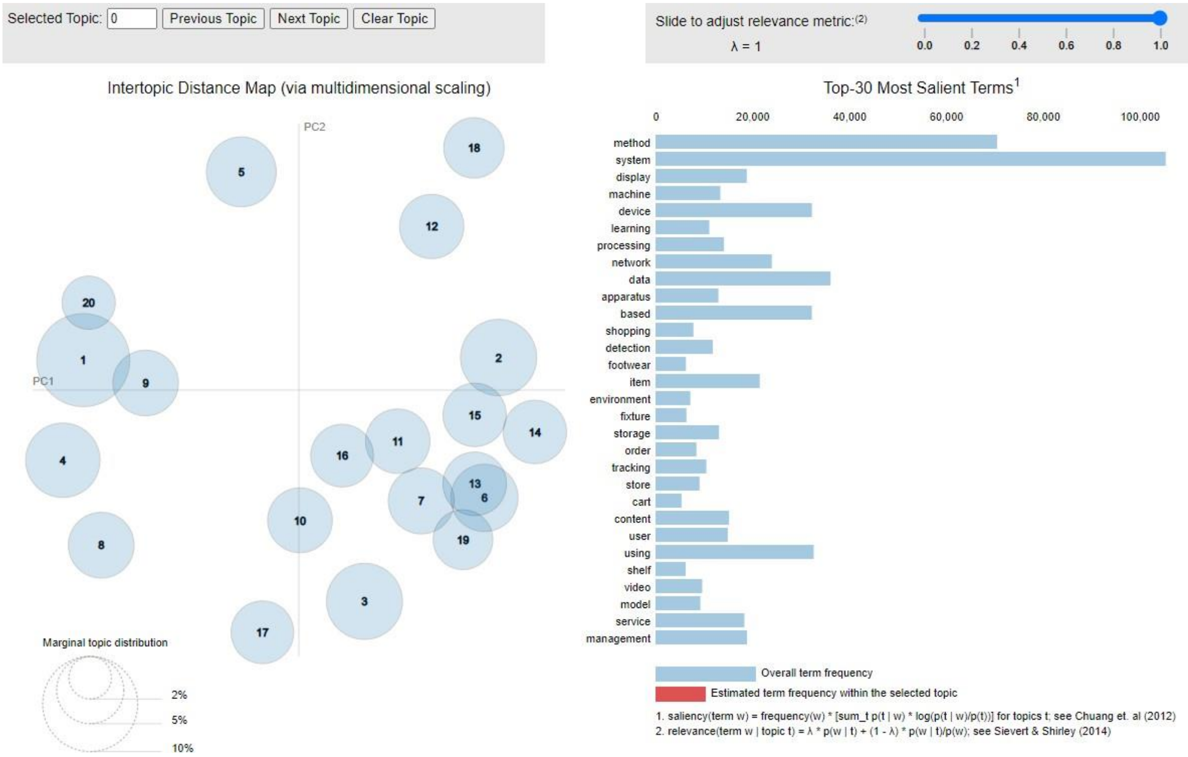

Figure 2. LDAvis Interface

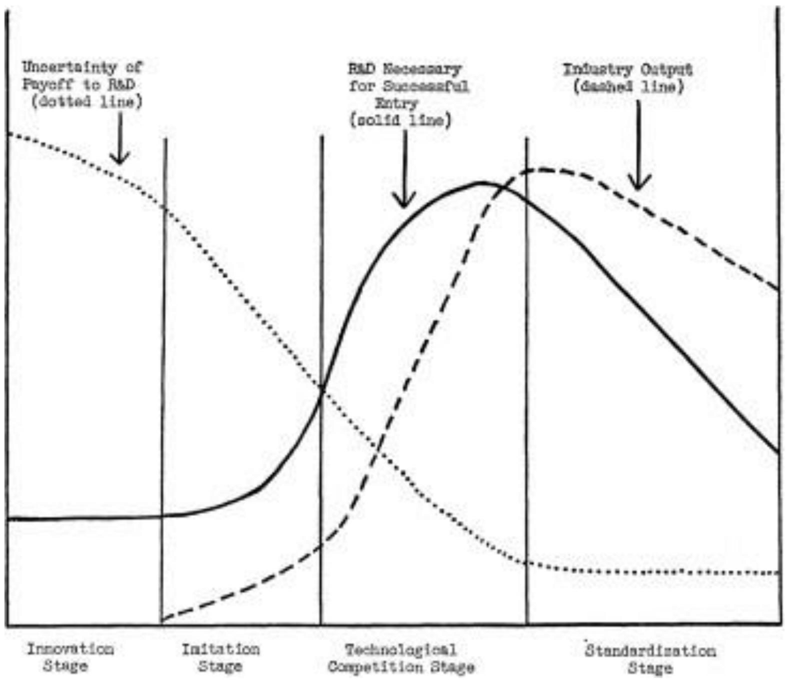

Figure 3. Mueller D.'s (1969) technological development cycle