IMAGINE TOMORROW, CHALLENGE TODAY: CONSTRUCTING A TECHNOLOGICAL ROADMAP FOR SPORTS INNOVATION THROUGH AUTOMATED PATENT ANALYSIS

ir. X.J. ter Wengel

07-08-2022

UNIVERSITY OF TWENTE.

Imagine Tomorrow, Challenge Today: Constructing A Technological Roadmap For Sports Innovation Through Automated Patent Analysis

MASTER THESIS

of

Xanthe ter Wengel

to obtain

the degree of Master of Science in Business Administration

at the University of Twente,

to be publicly defended on Thursday, the 11th of August 2022

Faculty of Behavioural, Management and Social Sciences (BMS), University of Twente

This research project was commissioned by Demcon advanced mechatronics, which is part of the Demcon Group, Enschede, the Netherlands.

Author

ir. Xanthe Jannie ter Wengel
1803220

Examination committee

dr. Matthias de Visser (1st supervisor)
dr. Agata Leszkiewicz (2nd supervisor)
Emma van Geel, MSc (external supervisor)

About Demcon

This paper was commissioned by Demcon, which is a Dutch engineering firm. Founded in 1993 by Dennis Schipper and Peter Rutgers at the University of Twente, Demcon has grown from a startup into an international organization with more than 1,000 employees. Demcon is active in a wide array of markets, including aerospace, defense & security, life sciences & health, and smart industry (Kuitert, 2016). Recently, Demcon acquired Johan Sports, which develops and produces sensors to monitor human movement of teams. This acquisition is part of a bigger project, in which Demcon wants to expand its business to the sports technology innovation market. This paper explores the technology developments in sports innovation and constructs a technological roadmap for Demcon for further expansion in the sports technology market.

Abstract
The sports technology market has increasingly expanded over the past years, mainly under influence of Industry 4.0. This research employed automated patent analysis with a conceptual competence analysis to identify business opportunities for Demcon. A Latent Dirichlet Allocation model was applied to patents extracted from Google Patents to identify technology development trends in sports. Three trends in sports technology were found: video analysis, sports equipment, and sensor technology. All of these trends could be promising opportunities for Demcon. It is advised that Demcon first focuses on gaining more insight into these specific markets and identifying its players. Further research could focus on the extension of the framework by including the needs of end-users or enabling the framework to filter relevant data sources. Practical implications for Demcon consist of contacting sports organizations or smaller companies for contract R&D. This study has contributed to the application of automated patent analysis for identifying business opportunities.

Keywords: sports technology, patent analysis, technological roadmap, market research

Executive summary

Sports technology is the field of innovations that are applied to sports. The sports technology market has increasingly expanded over the years. This is mainly caused by the rapid development of technology. Examples of such technologies include smartwatches, goal-line technology, and competitive gaming.

Recently, Demcon acquired Johan Sports as a first step toward entering the sports technology market. Johan Sport develops sensors for monitoring human movement in a team setting. Demcon wants to further expand its business in the sports technology market.

However, it is unclear what the sports technology market looks like and what opportunities there are for Demcon in this market, based on its current assets and competencies.

Therefore, current technology development trends in sports technology were researched and related to the assets of Demcon to construct a technological roadmap. It was found that current sports technology trends are focused on video analysis, sports equipment, and sensor technology. All of these trends could be promising opportunities for Demcon. It is advised that the next step for Demcon is to gain further understanding of these markets and identify players in this market. This could be done by interviews with experts, coaches, and athletes, in addition to desk research. Additionally, sports organizations and smaller companies could be contacted to provide contract R&D.

Table of contents

List of figures and tables

Introduction

Sports innovations are crucial for keeping people active and for continued participation at the top level in international sports (Bregman, 2019). However, going from scientific knowledge to innovations is often one bridge too far (Passarelli, Cariola, & Vecellio, 2018). Investments from the corporate world can play a key role in turning knowledge into innovation, which would benefit both parties by stimulating research and creating profitable innovations (Steiber & Alänge, 2020). The acquisition of Johan Sports by Demcon, a Dutch engineering firm, is an example of such an investment. Johan Sports develops and produces sensors for monitoring human movements in team sports, such as football and hockey. This acquisition has enabled Demcon to acquire technical assets which are valuable in sports technology, such as sensor technology. Demcon wants to further grow its business in the sports technology innovation market, as this market has increasingly expanded over the past years. This expansion is caused by the onset of Industry 4.0 and the application of associated technologies such as artificial intelligence and virtual reality to sports (Le, Le, Tromp, & Nguyen, 2018; Passos et al., 2021). However, it is unclear what opportunities exist for Demcon in sports technology innovation and what a business plan in sports technology could look like. The most promising option for Demcon in the sports technology market would be to focus on sports technologies where its existing technological assets have a high superiority. This would enable Demcon to ensure efficient use of its assets, which in turn would enhance successful business diversification. Thus, deeper insight into the sports technology market and how Demcons' assets can be aligned with sports technologies is needed.

Market research can be applied in various ways to gain insight into a (new) market and at all stages of the product or service life cycle (Hague, Harrision, Cupman, & Truman, 2016). In this case, market research is applied to map the technological landscape of sports technology innovations, which can capture players in the market, new developments, and development trends (Paap, 2020). This provides insight into the building blocks of this industry and can guide Demcon through a technology roadmap in strategic decision-making (S. Lee, 2013; Song, Yoon, Ko, & Han, 2016). On top of that, the identification and exploitation of new technology opportunities have led to numerous successful business endeavors, illustrating how crucial this process is for companies (Newbert, Walsh, Kirchhoff, & Chavez, 2006). Earlier methods for (technological) landscape mapping depended on traditional market research methods, such as interviews, surveys, and the Delphi method (Brink, 2001; Cruz & Boehe, 2008). However, these types of methods can be time-consuming and are prone to subjective bias (Y. Lee, Kim, Song, Park, & Shin, 2014; Shibata, Kajikawa, Takeda, & Matsushima, 2008; Xu, Dong, & Feng, 2020). Therefore, accompanied by the onset of Industry 4.0, market research has started with the integration of big data analytics in research.

For identifying the technological landscape with big data, patents have been found to be an up-to-date and reliable data source. Patents are an important source of technological intelligence and can thus be used to gain a strategic advantage (S. Lee, 2013; Yeap, Loo, & Pang, 2003). Moreover, patent analysis is a promising tool for decision support, technology monitoring, and the identification of technology trends (Choi, Yoon, Kim, Lee, & Kim, 2011; J. Gerken, Moehrle, & Walter, 2010; Moehrle & Geritz, 2007; Yoon & Kim, 2011). Increasingly, text mining is used to analyze patents to increase the analysis scope and richness of information (C.

Lee & Lee, 2019; Paap, 2020). Machine learning and text mining can be employed to enable the autoclassification of patents in an efficient and effective manner (Chan, Ng, Ang, Ng, & Kor, 2021; Forestal, Liu, Pi, & Li, 2021; Kong, Zhou, Liu, & Xue, 2017). However, the application of automated patent analysis with technological roadmapping is limited.

Hence, this article proposes a framework for identifying opportunities in the technological landscape of a new market for firms, based on the autoclassification of patent analysis with text mining. This framework is then applied to the sports technology innovation market for Demcon. First, patent data is collected from the European Patent Office and the United States Patent and Trademark Office databases for sports technology innovations. Second, a Latent Dirichlet Allocation (LDA) model is applied to the extracted titles and abstracts from these papers to identify multinominal distributions of words in these patents, which are used to identify technological development trends in sports technology innovation. Third, Demcons' core competencies and technical assets are identified with the core competence chart by Edgar & Lockwood (2012). Finally, a technological roadmap is constructed for Demcon, which captures trends in sports technology innovations and provides potential business opportunities for Demcon based on its competencies and technical assets.

The research question of this study is defined as 'What opportunities exist for Demcon in the sports technology market?' The following sub-questions are defined: (*RQ1*) What are current technology development trends in sports technology? (*RQ2*) What are the core competencies and technical assets of Demcon? (RQ3) How do Demcons' assets match with the found sports technology development trends? This paper builds on the article by S. Lee (2013) on the application of patent analysis in the roadmapping process and on the autoclassification of patents to identify technology development trends, using text mining (Forestal et al., 2021). From an academic perspective, this paper proposes a new method for identifying technology opportunities from patent analysis. It is the first study to test the applicability of patent analysis for strategic decision-making in a business-to-business context by taking core competencies and technical assets into account. This ensures efficient use of assets for business diversification. From a practical perspective, this article provides guidance for future marketers in employing text mining and automated patent analysis for exploring new markets. It also fosters knowledge expansion in the field of sports technology innovations.

The rest of this article is organized as follows. The following section reviews the literature. The methods section discusses the data collection, processing, and trend extraction. Subsequently, the results are shown. The paper ends with a discussion and conclusion. Here, the aforementioned research questions are answered, and practical advice is provided for Demcon for future strategic decision-making.

Literature review

Moving forward: innovations in sport

The interest in sports innovations is growing, as sports organizations view innovation as a way to gain a competitive advantage (Ringuet-Riot, Hahn, & James, 2013). For athletes, this advantage is realized by enhanced performance (Nlandu, 2012). The onset of Industry 4.0 has accelerated the knowledge expansion in new technologies, which can also be applied to sports innovations. In this study, sports technology innovation is defined as innovations related to sports that implement new or improved technology (Ratten, 2017). Innovations in this context can be services, products, or technologies. Based on literature research via Scopus using 'sport AND technology' as a search query, four categories of sports technology innovations could be distinguished, which are described below and shown in figure 1. When an interesting category was found, further information was obtained by looking at references mentioned in the text and by searching articles in Scopus, with the search term based on the technology mentioned.

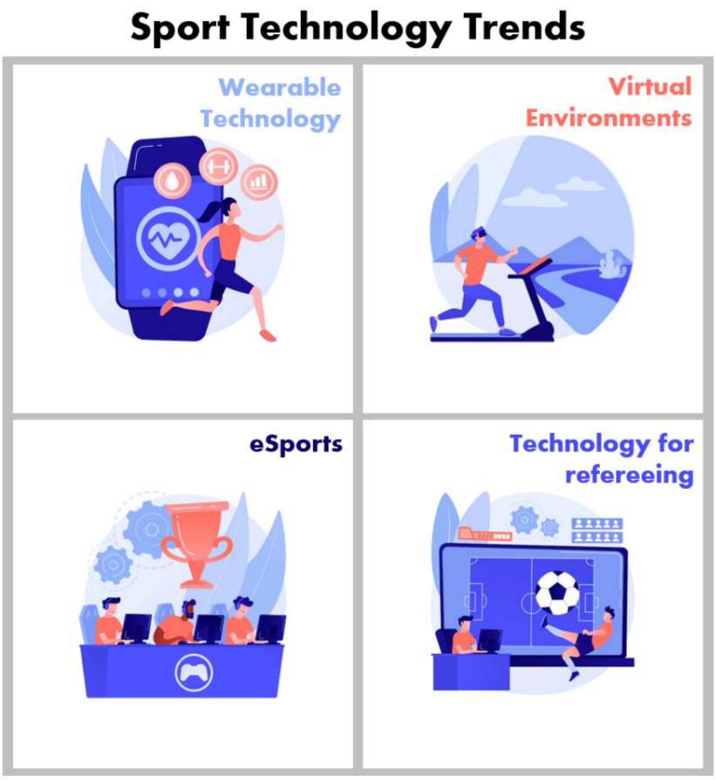

Figure 1. Sport technology trends. Found trends in sports technology based on literature research. Four trends were found: wearable technology, virtual environments, eSports, and technology for refereeing.

Wearable technology

Wearable technology is used to gather data from athletes. It is one of the biggest trends in sports technology (Bal & Dureja, 2012; Neumann et al., 2018; Robertshaw, 2021; Winand & Fergusson, 2016). Wearable technology is defined as all 'devices worn on or close to the body to monitor, analyze, transmit or receive data' (Düking et al., 2018, p. 178) to support the user (Adesida, Papi, & Mcgregor, 2019; Düking et al., 2018; Purwar & Daim, 2021). Wearable technology is used by athletes for monitoring their training and tailoring it to their needs and goals (Gabbett et al., 2017; Janssen et al., 2020; Mayne, Bleakley, & Matthews, 2021; Neumann et al., 2018; Westmattelmann, Daniel Grotenhermen, Stoffers, & Schewe, 2021). Moreover, the data provided by wearables is more and more viewed as crucial by athletes for improving performance (Barbosa, 2018; Ofoghi, Zeleznikow, MacMahon, & Raab, 2013). Wearable technology can thus be used to enhance performance.

Examples of wearable technology include smartwatches (wristwatch), smart rings (worn on finger), insoles, and shoe-embedded sensors (into or onto a shoe) (Purwar & Daim, 2021; Repanovici, 2018; Saidani, Haddad, Mezghani, & Bouallegue, 2018; Siepmann & Kowalczuk, 2021). However, wearables can be placed on any part of the body or can even be integrated into clothing, such as socks, gloves, or shirts (Mendes, Vieira, Pires, & Stevan, 2016; Scataglini, Moorhead, & Feletti, 2020). Each device can be equipped with multiple sensors to measure various body parameters. Parameters measured by sensors are, for example, heart rate, (relative) position, and mechanical pressure (mostly used for insoles in gait analysis) (Adesida et al., 2019; Aroganam, Manivannan, & Harrison, 2019; Saidani et al., 2018; Thompson, 2022). Furthermore, most devices also give feedback on the load of the training which is performed. The general aim of providing this feedback is to show athletes whether they might be at risk for injury, based on the response of their body to training. This training load can be calculated by measuring the rate of perceived exertion, duration of the training, and heart rate (Rao, Seshadri, & Hsu, 2021; Seshadri et al., 2021). Wearable technology has thus become an essential tool for most athletes in their training and performance.

Virtual environments

Virtual environments enable athletes to replicate races or matches in a virtual world. According to Le Noury, Polman, Maloney, & Gorman (2022), virtual environments can be viewed as a spectrum. One end of the spectrum is a completely virtual world, while the other end of the spectrum is a seamless merge of the virtual environment with the real world. Technologies along this spectrum are thus defined by varying degrees of virtual implementation. Examples of such technologies include virtual reality, augmented reality, and mixed reality. Virtual reality is found at the end of the spectrum, in which the user interacts with a completely virtual world, using virtual reality glasses. Using animated scenes or animated footage of the real world, users can interact with objects, but only virtual interactions are possible. Augmented reality is found at the center of the spectrum. In augmented reality, objects in the real world are enhanced or added by the computer. A recent example of an augmented reality application is Pokémon Go, in which users have to travel to different locations in the real world to catch virtual characters (Khamzina, Parab, An, Bullard, & Grigsby-Toussaint, 2020). Mixed reality can be found at the other end of the spectrum. In mixed reality, users can physically interact with virtual objects for a seamless merge of the virtual environment and the real world. Virtual reality thus refers to varying degrees of virtual implementation in the real world.

Virtual environments can have various purposes in sports. Athletes can use virtual environments to train the application of their skills in the real world (Michalski, Szpak, & Loetscher, 2019). Such skills could be for example perceptual-cognitive skills (such as decision-making skills in football or basketball) and motor skills (such as technique in darts, rowing, tennis, and skiing) (Fortes et al., 2021; Michalski et al., 2019; Pagé, Bernier, & Trempe, 2019; Rauter et al., 2013; Tirp, Steingröver, Wattie, Baker, & Schorer, 2015). It is also possible to use virtual environments for virtual races (Neumann et al., 2018). This became especially popular during the COVID-pandemic when real-life racing was not possible (Westmattelmann, Daniel Grotenhermen et al., 2021). Using Zwift, which is an online training platform for running and cycling, users were still able to compete against each other (Robertshaw, 2021). One could argue that these types of racing could be viewed as eSports, which is discussed in the next paragraph. However, these types of races do not use traditional gaming devices, such as a keyboard, a computer mouse, or a controller, but use stationary bikes or treadmills instead (Wong & Meng-Lewis, 2022). In other words, it is the practice of a traditional sport in a virtual environment. As such, virtual races as in Zwift are viewed as a subcategory to virtual environments. Virtual environments are thus not restricted to use in training by athletes but can also be used for racing.

eSports

Digitalization has challenged and reshaped the traditional sports practice, which has led to the introduction of eSports. eSports can be defined as competitive video gaming (Hallmann & Giel, 2018; Jenny, Manning, Keiper, & Olrich, 2017). It is one of the biggest multi-sport games in the world and is rapidly growing (Hamari & Sjöblom, 2017). An ongoing debate is whether eSports should be viewed as a sport or as a form of leisure time (DiFrancisco-Donoghue, Werner, Douris, & Zwibel, 2020; Zimmer, Haupt, Heidenreich, & Schmidt, 2022). Often, it is argued that eSports should not be viewed as a form of sports, based on its lack of physical activity (Hallmann & Giel, 2018; Jenny et al., 2017). Using only physical criteria for classifying sports, however, is iffy for several reasons. First, it raises the question of what even constitutes physical activity. By definition, all human actions are physical in nature, which means that movements made in chess or gaming are also a form of physical activity (Van Hilvoorde & Pot, 2016). Based on this notion, it is thus impossible to distinguish between a sport such as running or any form of eSports. Second, professional eSports players do exhibit signs after games that indicate physical exertion (Kane, 2017; R. Li, 2017; Rodriguez et al., 2016). Finally, the definition of sport is generally accepted to not only compromise physical activity but also skill training, competition, and tactical knowledge. All of these components are present in eSports (Hallmann & Giel, 2018; Thiel & John, 2018). As such, this study views eSports as sports technology, in which competitive gaming is its central component.

While eSports is regarded as a form of competitive gaming, it is not bound by a specific genre of games. Generally, three categories of gaming are distinguished, which can be played both individually or in teams (Kane, 2017; Thiel & John, 2018; Toth, Conroy, & Campbell, 2021): strategy games (real-time strategy, multiplayer online battle arenas), ego shooters (first-person shooters), and sport and race simulations (e.g. FIFA). eSports, while not exhibiting the traditional elements of sports, is thus generally accepted as one of the biggest trends in sports of the twenty-first century and illustrates the impact of technology on sports.

Technology for refereeing

Sports technology innovations are not limited to technologies for aiding athletes. Another application of sports technology is in objective decision-making for referees. Such technologies can help referees in making split-second decisions and potentially lead to a fairer competition

(Leveaux, 2010; Winand & Fergusson, 2016). For instance, many sports make use of a video referee, who checks video feeds and can replay important situations during a match (Spitz, Wagemans, Memmert, Mark Williams, & Helsen, 2020; Zglinski, 2022). Football also makes use of Goal Line Technology. Using a chip inside the ball, a signal is sent to the referee when the ball has crossed the line or not, indicating whether a goal was made or not (Leveaux, 2010). Cricket and tennis also use Hawk-Eye technology, which is a computer- and camera-based system that tracks the trajectory of the ball (Winand & Fergusson, 2016). Additionally, the application of artificial intelligence in refereeing is increasing (Gottschalk, Tewes, & Niestroj, 2020; Morkhat, 2018). The general aim of these technologies is to enhance objectivity in decision-making, specifically in split-second situations.

The (technological) roadmap to success

Market research can be applied to gain insight into a market. It can be applied at all stages of the product or service life cycle (Hague et al., 2016). Understanding a new market is essential for successful market entry (Wang & Lestari, 2013). Mapping the technological landscape of a new market, such as sports technology, can provide guidance to businesses for strategic planning (S. Lee, 2013; Ozcan, Homayounfard, Simms, & Wasim, 2022). A technology roadmap can consist of capturing technology development trends and aligning these with a company's competencies and technical assets when a commercial perspective is employed (Ampornphan & Tongngam, 2017; Huang, Kuo, Wang, Hsiao, & Yang, 2021; C. Lee & Lee, 2019; S. Lee, Yoon, Lee, & Park, 2009). Technology roadmaps can thus serve as a methodology to develop an integrated business plan, which may include technology, product, or service developments (S. Lee et al., 2009; H. Zhang, Daim, & Zhang, 2021). Technology roadmaps are, therefore, a valuable tool for businesses.

Mapping technology developments can aid decision-making in new product development. Technology developments can help firms in choosing innovation strategies, as they can show firms how to maintain an effective linkage between their technical assets and their objectives (Alstott, Triulzi, Yan, & Luo, 2017; Ami et al., 2020). As such, technology mapping can help firms make decisions among different alternatives, which can facilitate strategic decision-making (Arasti & Bagheri Moghaddam, 2010). Moreover, mapping technology developments can help firms in establishing a balance between the pull of the market and the push of technology (Toro-Jarrín, Ponce-Jaramillo, & Güemes-Castorena, 2016). Mapping technology developments in sports technology can thus aid Demcon in which technology to pursue, which fosters strategic decisions.

The rise of technology associated with Industry 4.0 has also affected market research practices for mapping. Traditional market research practices for technology mapping depended on qualitative research methods and small samples (Brink, 2001; Cruz & Boehe, 2008; Doz, 2011; Fossey, Harvey, McDermott, & Davidson, 2002). This consisted of holding focus groups or interviewing experts (Hennink, Hutter, & Bailey, 2020). However, these methods are time-consuming and do not necessarily lead to reliable information (Shibata et al., 2008). Therefore, accompanied by the rapid developments in data analytics, research has shifted to a combination of quantitative methods with big data (Bosch, 2016; Ducange, Pecori, & Mezzina, 2017; J. Liu, Li, Li, & Wu, 2016; Raj & Mishra, 2020). Big data in this case can be anything, from text or video to webpage log files (Jha, Dave, & Madan, 2016; Meire, Ballings, & Van den Poel, 2017; Yafooz, Bakar, Fahad, & Mithun, 2020). This has enabled marketers to generate new insights from a wide variety of sources, which has become crucial for successful decision-making (Ahn et al., 2019; Erevelles, Fukawa, & Swayne, 2015; Hung, He, & Shen, 2020; Jha et al., 2016; Raj & Mishra, 2020; Rejeb, Rejeb, & Keogh, 2020; Saura, Ribeiro-Soriano, & Palacios-Marqués, 2021). For technology roadmapping, patents have been found to be a valuable source.

Patent analysis for strategic decision-making

For mapping the technological landscape, patents have been found to be a valuable data source. Patents contain information about product and technology attributes (S. Lee, 2013). On top of that, patents are publicly available, easily accessible via databases, and commonly structured in a fixed way (Noh, Jo, & Lee, 2015). Patent analysis has been used for decision support, technology monitoring, and the identification of technological trends (Choi et al., 2011; J. Gerken et al., 2010; Moehrle & Geritz, 2007; Yoon & Kim, 2011). Moreover, patent analysis can be used to construct technological roadmaps (Ampornphan & Tongngam, 2017). Employing patent analysis for technological roadmapping has several advantages. First, it allows the mapping of technology in a systematic way and by using one of the most representative sources for technological change (S. Lee, 2013; S. Lee et al., 2009; Yoon & Kim, 2011). Second, it aids objectivity and reliability in technological mapping in comparison to experts' opinions (Y. Lee et al., 2014; Shibata et al., 2008; Xu et al., 2020). Finally, from a commercial perspective, patent analysis combined with technology mapping has the potential to foster the identification of relevant business opportunities (Feng, Liu, & Feng, 2021).

Generally, four applications of patent analysis can be distinguished: analysis based on standard industrial classification codes, analysis based on patent classes, analysis based on patent citations, and analysis based on text mining (Aharonson & Schilling, 2016). Text mining has been regarded to be the most promising method as it captures the essence of the technology being patented (Fattori, Pedrazzi, & Turra, 2003; Paap, 2020). Earlier methods of text mining in patent analysis used keyword-based approaches to classify patents (S. Lee, 2013; Y. Lee et al., 2014). However, this could lead to invalid results, as keywords could have more than a single meaning, which increases the risk of misclassification (Noh et al., 2015). Therefore, instead of keyword distribution, semantic analysis has increasingly been used to classify patents (J. M. Gerken & Moehrle, 2012; H. Zhang et al., 2021). Semantic analysis extracts Subject-Action-Object (SAO) structures, which gives a better representation of the technological contents of patents. SAO structures are extracted syntactically ordered sentences. This extraction of multiple words from the patent text with natural language processing is preferable, as it is more precise and better reflects the content of the patent, such as technological objectives and structures (Cascini, Fantechi, & Spinicci, 2004; J. M. Gerken & Moehrle, 2012; Mann, 2004).

Machine learning can be used to reveal hidden knowledge patterns from patents. It is often applied for data analytics, as machine learning techniques foster efficiency and effectiveness (Bach, Krstič, Seljan, & Turulja, 2019; C. Liu, Feng, Lin, Wu, & Guo, 2020). Machine learning can be used for various applications in data analytics, including data labeling, descriptive analytics, and predictive analytics (Nti, Quarcoo, Aning, & Fosu, 2022; Raj & Mishra, 2020).

Machine learning can be supervised or unsupervised. In supervised machine learning, labeled data is required to train the model, while unsupervised machine learning models look for patterns in the data itself (Bonaccorso, 2017). When unsupervised machine learning is combined with text mining, topic distributions in the text can be detected with natural language processing models, such as Latent Dirichlet Allocation (LDA) (Bach et al., 2019; Kang, Cai, Tan, Huang, & Liu, 2020; Moro, Cortez, & Rita, 2015; Schumaker & Chen, 2006). Earlier research has applied this successfully to identify market trends in various domains (Kim & Woo, 2022; Lau, Collier, & Baldwin, 2012; Park, Kim, Choi, & Han, 2018; Pek & Lim, 2019; Ploessl, Just, & Wehrheim, 2021; Sung & Yeo, 2019). Xu et al. (2020) applied this framework with deep learning to map the technological landscape of an emerging industry, 3D printing, to overcome the limitations of a small dataset. Zhang et al. (2021) used this framework to perform a roadmapping analysis in the field of blockchain to describe the current state of technology developments and predict future development trends in this field. Patent analysis with text mining can thus be used to describe technology development trends in a market.

Research gaps

The current research on automated patent analysis for a technological roadmap has several gaps. First, patent analysis has exclusively been used in research to describe or forecast trends. However, the application of patent analysis in businesses, specifically for the identification of business opportunities, has not yet been addressed. The use of patent analysis in businesses is complicated, as found technology development trends need to be assessed to see whether they are a fit with the business. It is, therefore, yet unclear whether patent analysis can be used for strategic decision-making and what a framework for applying patent analysis for business diversification would look like. Second, the use of automated patent analysis for trend identification has mostly been restricted to slightly smaller datasets, with less than 10,000 patents. This research addresses whether automated patent analysis can also be used by bigger datasets and whether it is favorable to use larger datasets. Finally, the application of the framework of this study in the sports technology field also aids understanding of the sports technology market and provides a general overview of the trends in this market.

The current study assesses whether unsupervised machine learning with text mining can directly be used to map the current state of technology developments from patent data to identify business opportunities. Technology development trends in sports technology are compared with core competencies and technical assets of Demcon to identify business opportunities. As such, this study combines patent analysis with a conceptual competence analysis to construct a technological roadmap, which has, to the knowledge of the author, not been done before.

Methodology

The goal of this study was to identify business opportunities for Demcon in the sports technology market. To do this, a framework was used, which can be viewed in figure 2. The framework consists of identifying current technology development trends in sports, identifying technical assets of Demcon, and matching these assets with the found technology trends to identify business opportunities.

Data collection

First, technology development trends in sports needed to be identified. Patents have been found to reflect the technological change and therefore seemed appropriate to use in this context (S. Lee, 2013; S. Lee et al., 2009; Yoon & Kim, 2011). Patents were collected from Google Patents (https://patents.google.com/), which has the advantage that multiple databases can be searched at the same time, fostering efficiency (Noruzi & Abdekhoda, 2014). Furthermore, the HTML of Google Patents is relatively easy to interpret, which makes it easier to scrape relevant information. The patents in the Google Patents database are also publicly available, which makes it more permissive for data extraction, in comparison to other web scraping sources (Boegershausen, Datta, Borah, & Stephen, 2022). Finally, Google Patents has the option to download the search results from every file in a CSV file, which not only fosters transparency but also simplifies web scraping in Python, as a simple for loop can be used to get the information from every patent.

Information extraction from each patent was done with the BeautifulSoup and requests packages in Python (see also Appendix-A). From each patent, the title and abstract were extracted, as these have been indicated to contain the most meaningful information (Liao, Le, & Zhu, 2017; Qi, Zhu, Zhai, & Ding, 2018). Patents without abstracts were excluded from the research. Patents were filtered to only include patents from the European Patent Office and the United States Patent and Trademark Office. This was done for two reasons. First, Demcon mainly operates in the Western world. Patents from the EU and US, therefore, hold the most potential for Demcon to be relevant. Second, the inclusion of just these two data sources would make sure that the computational requirements in Python Pandas would not be exceeded. The time window of the search was set from 2019 onwards. A small time window was chosen, as technology tends to develop rapidly (Milshina & Vishnevskiy, 2019; Weber, Harper, Könnölä, & Barceló, 2012). Overall, the following search query in Google Patents was used to search for patents in sports technology:

```
SEARCH QUERY = ((sport) AND (technology)) country:EP,US after:priority:20190101 language:ENGLISH
```

Second, the core competencies and technical assets of Demcon needed to be identified. These competencies and assets are needed to identify which technology development trends match with Demcon and which could thus be used most efficiently for business diversification. Information about Demcons' assets was collected from the website (https://demcon.com/), newspaper articles, and internal reports, as suggested by Edgar & Lockwood (2012). The researcher did this manually by making a list of the keywords on every page and in every document.

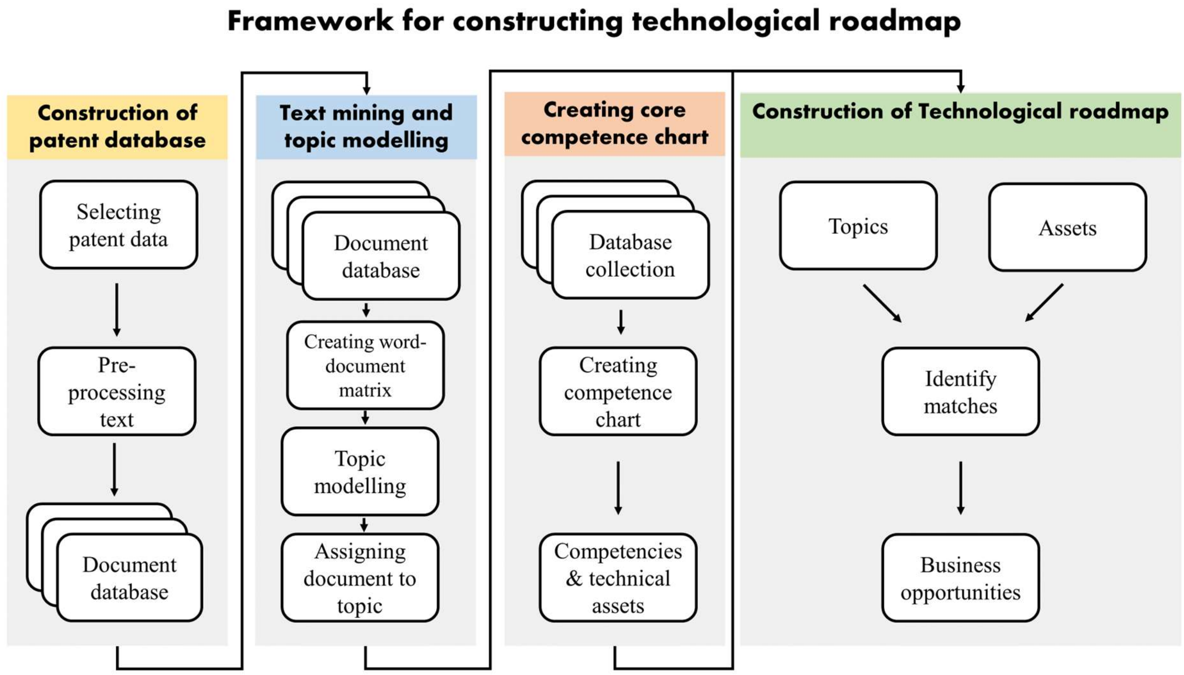

Figure. 2 Framework for constructing technological roadmap. Used framework for generating a technological roadmap from the automated patent analysis. Four phases can be distinguished: construction of the patent database, text mining and topic modeling of patent data, creation of competence chart, and construction of technological roadmap.

Data analysis

Titles and abstracts from patents for the technology development trends were first pre-processed in Python. Here, the NLTK, gensim, Spacy, and Pandas packages were used, in combination with custom scripts (see also Appendix-A). Standards methods for preprocessing text data were used, as described by Boyd-Graber, Mimno, & Newman (2014) and Vijayarani, Ilamathi, & Nithya (2015), to make the data ready for input to the model. Data pre-processing consisted of cleaning the data (removing non-English characters, lowering characters, segmenting sentences into words, removing stop words), creating bigrams, and lemmatization (reducing words to generic form). See also figure 2.

The present study used text mining to analyze patents for technology development trends. Text mining was used because earlier studies have indicated that it captures the essence of the technology most accurately (Fattori et al., 2003; Paap, 2020). When combined with machine learning, text mining also can be used to reveal knowledge patterns more easily and which would otherwise remain hidden. This is especially an advantage when working with larger datasets, as used in this study (Chhaya, Khanzode, & Sarode, 2020; Kasemsap, 2016). Topic modeling can be used to analyze text mining results by looking at topics, or groups of co-occurring words, in a text (Moro, Pires, Rita, & Cortez, 2019). Earlier studies have successfully applied topic modeling to extract market trends from various sources of data (Kim & Woo, 2022; Lau et al., 2012; Park et al., 2018; Pek & Lim, 2019; Ploessl et al., 2021; Sung & Yeo, 2019). It, therefore, seemed an appropriate solution to use in this context.

A topic model is a statistical model which can be used to summarize collections of texts and understand their latent structure (Barde & Bainwad, 2017; Boyd-Graber et al., 2014; X. Li &

Lei, 2021). Topic models can be both probabilistic and non-probabilistic, in which a probabilistic approach was preferred in this study. Probabilistic topic models calculate the probability that a document is categorized in a certain category, i.e. account for randomness, while non-probabilistic models assign documents to a category, i.e. do not account for randomness (Iwayama & Tokunaga, 1994; Kherwa & Bansal, 2020; Sun, Deng, & Han, 2012). An LDA model is the most used probabilistic model, as it is relatively easy to interpret, but model complexity is not compromised (Bach et al., 2019; Kang et al., 2020; Moro et al., 2015; Schumaker & Chen, 2006). Alternatively, a probabilistic latent semantic analysis (PLSA) model can be used. However, generally, LDA has been found to have an advantage over PLSA for classification tasks (Antons, Grünwald, Cichy, & Salge, 2020; Blei, Ng, & Edu, 2003; Lu, Mei, & Zhai, 2011). Therefore, an unsupervised LDA model was chosen for the present study.

An LDA model is an unsupervised machine learning model and was used to extract the technology development trends from the patent data. A schematic overview of the LDA model is presented in figure 3. It is built on the notion that every document of a text contains a random mixture of latent topics (Jelodar et al., 2019). Topics are then represented by words with a high distribution, i.e., a topic thus refers to a probability distribution of a set of words in the text. The LDA model from the gensim library was used in this study to extract topics from the titles and abstracts of the patents. The LDA model in this package uses the bag of words approach to distinguish words. In this case, sets of words in documents are stripped of grammar and are reduced to their generic form. A text is thus viewed as just a bag of words in which grammar is disregarded, i.e. an unordered list of words (Walkowiak, Datko, & Maciejewski, 2019; H. Zhang et al., 2021). The advantage of this is that words such as 'watching' and 'watches' are both changed into the single term 'watch' and are thus not classified as two separate entities.

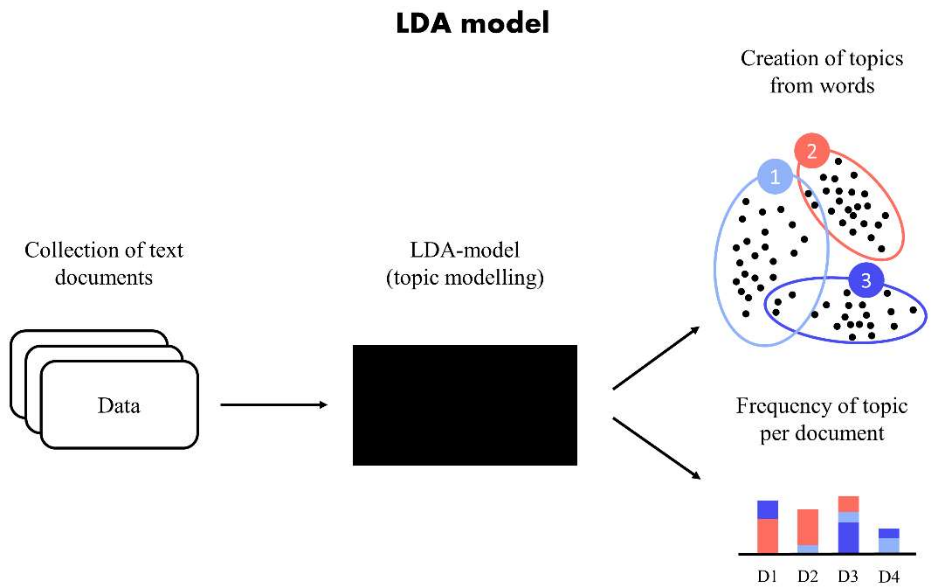

Figure 3. Schematic representation of LDA model. Adapted from Buenano-Fernandez, Gonzalez, Gil, & Lujan-Mora (2020). The model uses a collection of text documents as an input and gives topics and a frequency of a topic per document as an output.

The optimal number of topics to extract from the LDA model needs to be given beforehand. If too many topics are extracted from the text, this could lead to dispersion and longer processing times. On the contrary, if too few topics are extracted from the text, the accuracy of the model decreases, and distinguishing meaningful topics becomes harder (Evans, 2014; Hasan, Rahman, Karim, Khan, & Islam, 2021; Maier et al., 2018). Different measures exist to calculate the optimal number of topics but the most used is coherence. A coherence measure evaluates the correlation between the lower-level terms (i.e. the words that make up a topic) of a topic (Röder, Both, & Hinneburg, 2015). The general idea is that a higher coherence leads to higher accuracy and thus better performance. In this study, the LDA model was evaluated for a ranging number of topics, and the model with the highest coherence score was chosen.

The LDA model and its results were then analyzed with the LDAvis package in Python, which was developed to aid interpretability for LDA modeling (Sievert & Shirley, 2014). See also Appendix-B. One of the outputs of the LDAvis package is a horizontal bar chart for each topic, which shows the terms associated with each topic and which are ranked based on the relevance metric. Sievert & Shirley (2014) developed this relevance metric, λ (ranging from 0 to 1), to reorder the top words in a topic for aiding understanding. If a word occurs often in the documents, it will occur at the top of the topic list, but it will not necessarily contribute to the specific semantics of the topic. The smaller λ, the more topic-specific words will be at the top of the topic list words (Maier et al., 2018). Sievert & Shirley (2014) found that topics had the best interpretability when $\lambda = 0.6$ was used, so this was adopted in this study.

For determining the core competencies and technical assets of Demcon, a core competence chart was created. The framework by Edgar & Lockwood (2009) was used as guidance to construct the technical assets of Demcon. This framework constructs an overview of a business by dividing its competencies into a core phenomenon, general phenomena, product/service technologies, product/service sub-technologies, product/service classes, functional and technological skills, and integral skills. See also table 1. As such, the framework gives an overview of all activities of a business. This framework was chosen as it not only gives an extensive overview but also adopts a multi-level approach. This is especially helpful for larger companies such as Demcon, which are active in many industries and have many areas of expertise. This framework was then enhanced by discussions with a business developer and a communication specialist from Demcon. These people were chosen as they were familiar with all business aspects of Demcon on a more general scale. A small sample for the interviews was chosen, as the primary purpose of this study was to construct a framework to identify business opportunities from technology development trends in patent data. Therefore, the interviews were mainly used to make sure an accurate framework was constructed for Demcon, and less attention was given to the details of this framework. On top of that, Demcon is a dynamic company, which means that its technical assets and to a lesser extent its core competencies are subject to the rapidly changing environment in which it operates. This means that the core competence chart is also subject to change, and thus merely provides a snapshot of the current competencies and assets. It was, therefore, deemed more relevant to focus on establishing the framework than on capturing Demcons' competencies and assets as accurately as possible.

The technological development trends and the technical assets from Demcon were then used to construct a technology roadmap for Demcon in sports innovation. This consisted of comparing the found assets of Demcon with the found topics and looking for matches, i.e., whether

Demcon already had resources that could be employed for the technology development trend in question. This is thus mainly focused on assessing whether the aforementioned competencies and assets could also be applied to sports technology, instead of the current application field. For example, if Demcon would have experience in constructing machine learning algorithms for manufacturing purposes, a possible business opportunity could be the application of machine learning in sports technology. This would ensure efficient use of Demcons' assets for business diversification. The focus lies, therefore, on identifying linkages between Demcons' technology resources and the sports technology development trends. This is further elaborated on by providing practical examples of opportunities for Demcon, specifically focused on acquisitions and contract R&D.

Table 1. Related dimensions and meanings for core competence chart

Dimension	**Meaning**
Core phenomenon	Entity which people understand most thoroughly
General phenomena	Phenomena that, when combined, form the core phenomenon
Product/service technologies	Technologies that emerge directly from core phenomenon
Product/service sub-technologies	Specialized form of product/service technologies
Product/service classes	Products/services made possible by (sub)technologies
Skills	Functional: understanding of product/service classes Technical: understanding of technologies of product/services
Integral skills	Ability to do an activity based on the skills

Adapted from Edgar & Lockwood (2012). For a more in-depth analysis, please see Edgar & Lockwood (2012).

Results

This study used patent data to construct a technology roadmap for Demcon in the sports technology field. It uses the Google Patents database with the aforementioned search query. In total, this search led to 19,338 results. After the removal of patents without abstracts, 17,137 patents were left for analysis. These patents had in total 25,440 different inventors (inventor of the patent; can be more than one person) and 3,598 assignees (owners of the patent rights). Hyundai Motor Company, Ford Global Technologies, and LG electronics owned most patent rights, see also figure 4. Most patents were classified as G06F, G06Q, and H04L, which refers to data processing and data transmission, see also table 2 and Appendix-C.

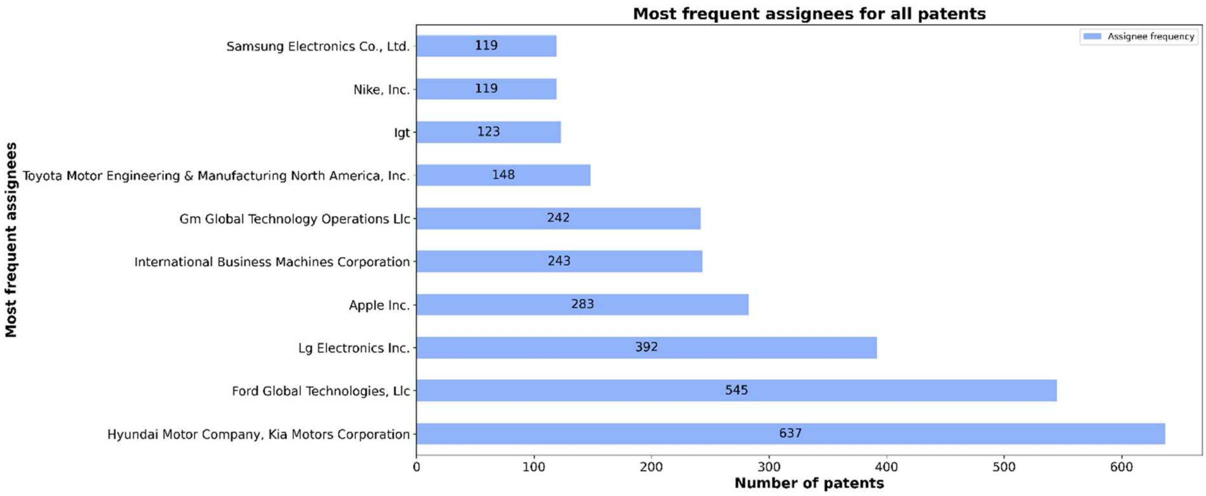

Figure 4. Most frequent assignees for all patents. Assignees hold the legal rights for a patent and are, therefore, often assigned to companies, instead of individual inventors. In total, 3,598 unique assignees hold the patent rights for the 17,137 patents used in this study, which were invented by 25,440 inventors. The above figure shows the most frequent assignees.

Topic extraction

An LDA model was then used to perform topic analysis on the titles and abstracts of each patent. The model was evaluated with a number of topics (2-15) to find the optimal number of topics (based on the highest coherence score). The optimal number of topics was found to be equal to 3 in which the coherence score was 0.52. Therefore, an LDA model with three topics was chosen. In figure 5, the most frequent terms used in all patents can be found. The term 'include' was used the most in the patent data. In figure 6, the frequency of the number of topics assigned to a document per year can be viewed. Here, each document was classified to a topic. So, if a document or patent contained multiple topics, it was assigned to the topic which had the highest frequency in the document. For all years combined, topic 1 was the most assigned topic (39.2% of tokens), followed by topic 2 (36.7% of tokens) and topic 3 (24.1% of tokens).

The LDAvis package in Python was used to order the results from the LDA model. The raw LDAvis results (with $\lambda = 0.6$), consist of an intertopic distance map and a bar chart with the top thirty most relevant terms for a topic (see Appendix-B). The top thirty most relevant terms from the bar chart were used to create table 3, which classifies the words for each topic. This table, in addition to the patents associated with a topic, was used to determine the meaning of each topic.

Table 2. Most frequent Cooperative Patents Classifications in the dataset with definition and relative count

CPC	Definition	Relative count (%)
G06F	Electrical digital data processing	17.5
G06Q	Data processing systems or methods, specially adapted for administrative, commercial, financial, managerial, supervisory, or forecasting purposes; systems or methods specially adapted for administrative, commercial, financial, managerial, supervisory, or forecasting purposes, not otherwise provided for	12.8
H04L	Transmission of digital information, e.g., telegraphic communication	10.1
H04W	Wireless communication networks	9.2
G06N	Computing arrangements based on specific computational models	9.1
H04N	Pictorial communication, e.g., television	8.8
G06V	Image or video recognition or understanding	7.6
A63B	Apparatus for physical training, gymnastics, swimming, climbing, or fencing; ball games; training equipment	6.7
G06T	Image data processing or generation, in general	6.2
A61B	Diagnosis; surgery; identification	5.8

Based on the definition by https://www.cooperativepatentclassification.org. Cooperative Patent Classifications (CPCs) were determined by taking a random sample of 1,714 patents (10%) from the total dataset. A subset was chosen, as the extraction of CPCs from all 17,137 patents would impose an extremely high demand on Google Patents, which could potentially lead to server problems for Google Patents.

For topic 1, the words ‘content’, ‘video’, ‘information’, ‘image’, and ‘receive’ indicate that this topic seems to be related to information extraction, possibly from videos (see table 3). The words ‘determine’, ‘network’, ‘identify’, ‘display’, ‘medium’, ‘application’, ‘service’, and ‘device’ seem to indicate that it focused on the collection, processing, and visualization of data. The associated patents for topic 1 (see Appendix-D) are mainly associated with data processing and collection. Therefore, topic 1 is identified as video analysis.

For topic 2, words such as ‘surface’, ‘body’, ‘layer’, ‘support, ‘material’, ‘composition’, and ‘form’ seem to indicate that the topic is related to the construction of equipment (table 3). When looking at the patents (see Appendix-D), some sports equipment can be recognized, such as the liner for a ski boot and an ice skate. Therefore, topic 2 is identified as sports equipment.

For topic 3, words such as ‘control’, ‘sensor’, ‘configure’, ‘power’, ‘light’, and ‘wireless’ seem to indicate that the topic is related to the engineering of sensors. Associated patents for this topic are focused on electrical engineering and designing system components (see Appendix-D). Therefore, topic 3 is identified as sensor technology. In table 4, the identified topics can be viewed.

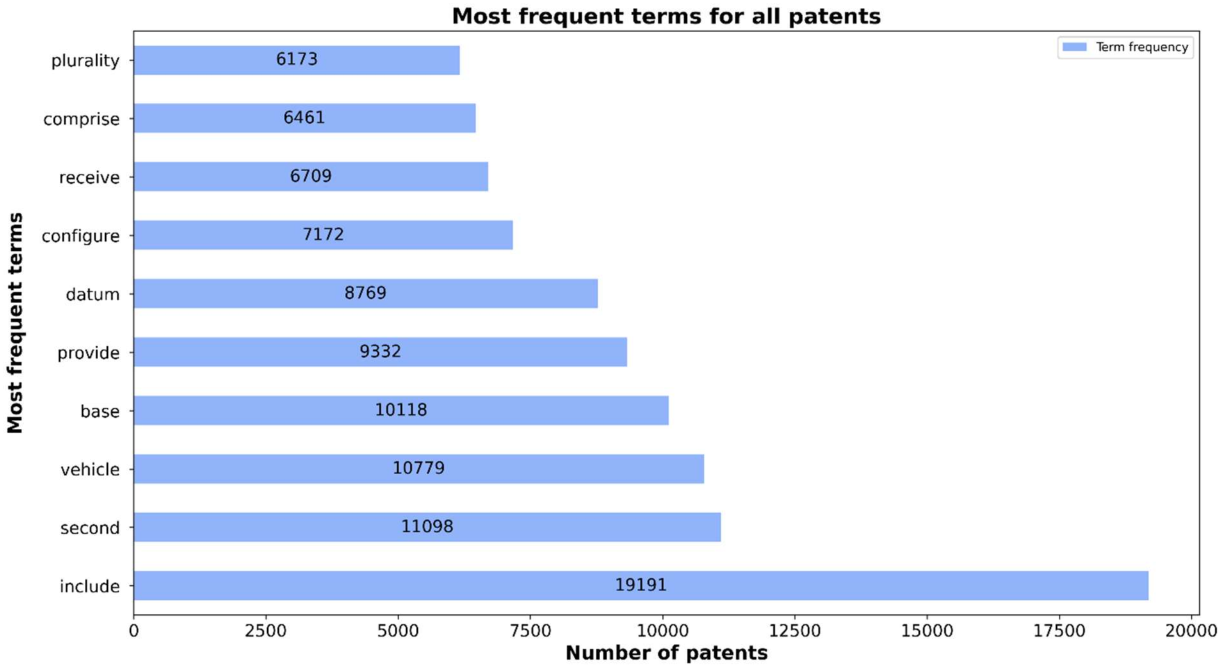

Figure 5. Most frequent terms used in all patents. The above chart is constructed based on keyword distribution in the corpus and thus not with topic modeling. In total, 31,513 unique keywords could be distinguished in the patent database. The term 'include' was observed the most in all patents.

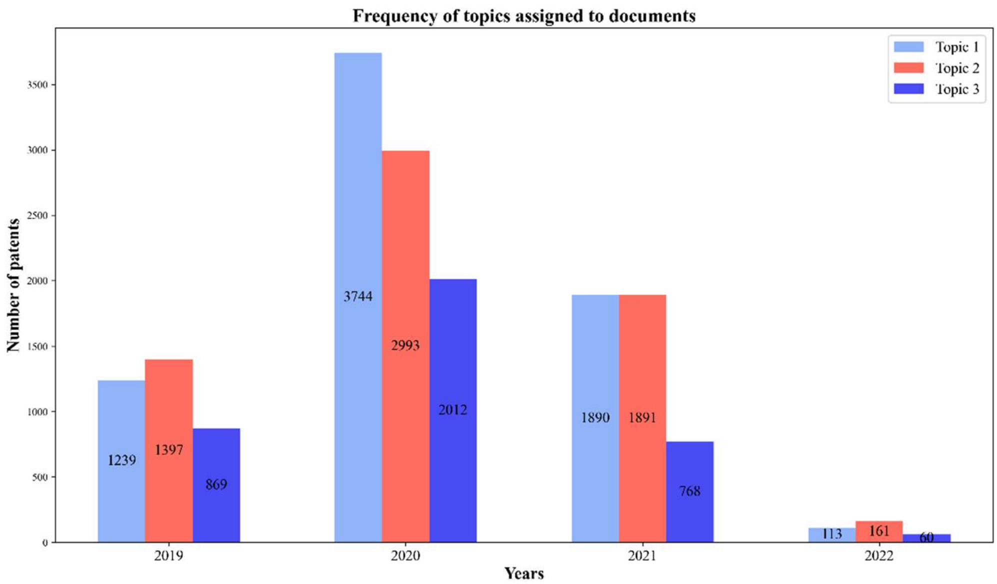

Figure 6. Frequency of the number of topics assigned to a document per year. Here, only the dominant topic per document is counted, i.e., if a topic consists of topic 1 and topic 2, the document is assigned solely to the topic with the highest frequency. The largest topic was topic 1, followed by topic 2 and topic 3. Most patents were filed in 2020.

Table 3. Word classification by topic

	Topic 1	Topic 2	Topic 3
1	datum	portion	vehicle
2	content	surface	control
3	base	include	sensor
4	video	body	unit
5	information	end	signal
6	include	member	configure
7	image	second	include
8	generate	comprise	power
9	receive	have	light
10	determine	layer	drive
11	network	support	wireless
12	provide	material	controller
13	associate	composition	base
14	model	form	apparatus
15	plurality	assembly	second
16	set	provide	module
17	identify	structure	communication
18	display	extend	determine
19	event	upper	battery
20	medium	position	mode
21	item	panel	motor
22	game	cover	receive
23	application	couple	electric
24	object	connect	operation
25	server	low	detect
26	select	outer	information
27	service	plate	state
28	device	mount	sound
29	request	frame	transmission
30	time	attach	speed

The thirty most relevant terms per topic. This is based on the raw LDAvis results ($\lambda = 0.6$), which can be observed in Appendix-B. An λ of 0.6 ensured the highest interpretability, based on Sievert & Shirley (2014).

Table 4. Topic identification by tokens

Topic	Description
Topic 1	Video analysis
Topic 2	Sports equipment
Topic 3	Sensor technology

Topics were identified by looking at the terms in table 3 and the associated patents to a topic (Appendix-D).

Assessing technical assets

Data for assessing the core competencies and technical assets of Demcon were collected from Demcons' website (https://demcon.com/), newspaper articles, and internal reports. Core competencies and technical assets were determined from the competence chart. The competence chart was filled in by assessing the found data. This was done by looking at the occurrence of keywords mentioned in the data. An initial competence chart was created for the discussions with the business developer and the communication specialist. Based on these discussions, the model was refined and is shown in figure 7. The model was only refined when both the business developer and the communication specialist agreed that an asset needed to be added or removed. It was found that Demcons' core competence is the conceptualization and development of (complex) mechatronic systems. Demcons' technical assets are described in its understanding of general phenomena (artificial intelligence, data science, mechatronics, mechanics, medical device certification, optics, production technology, robotics, software, sensor technology, system engineering, thermodynamics, and visual effects) and arise from eight different domains (chemical engineering, computer science, electrical engineering, industrial design, materials science, mathematics, mechanical engineering, and physics).

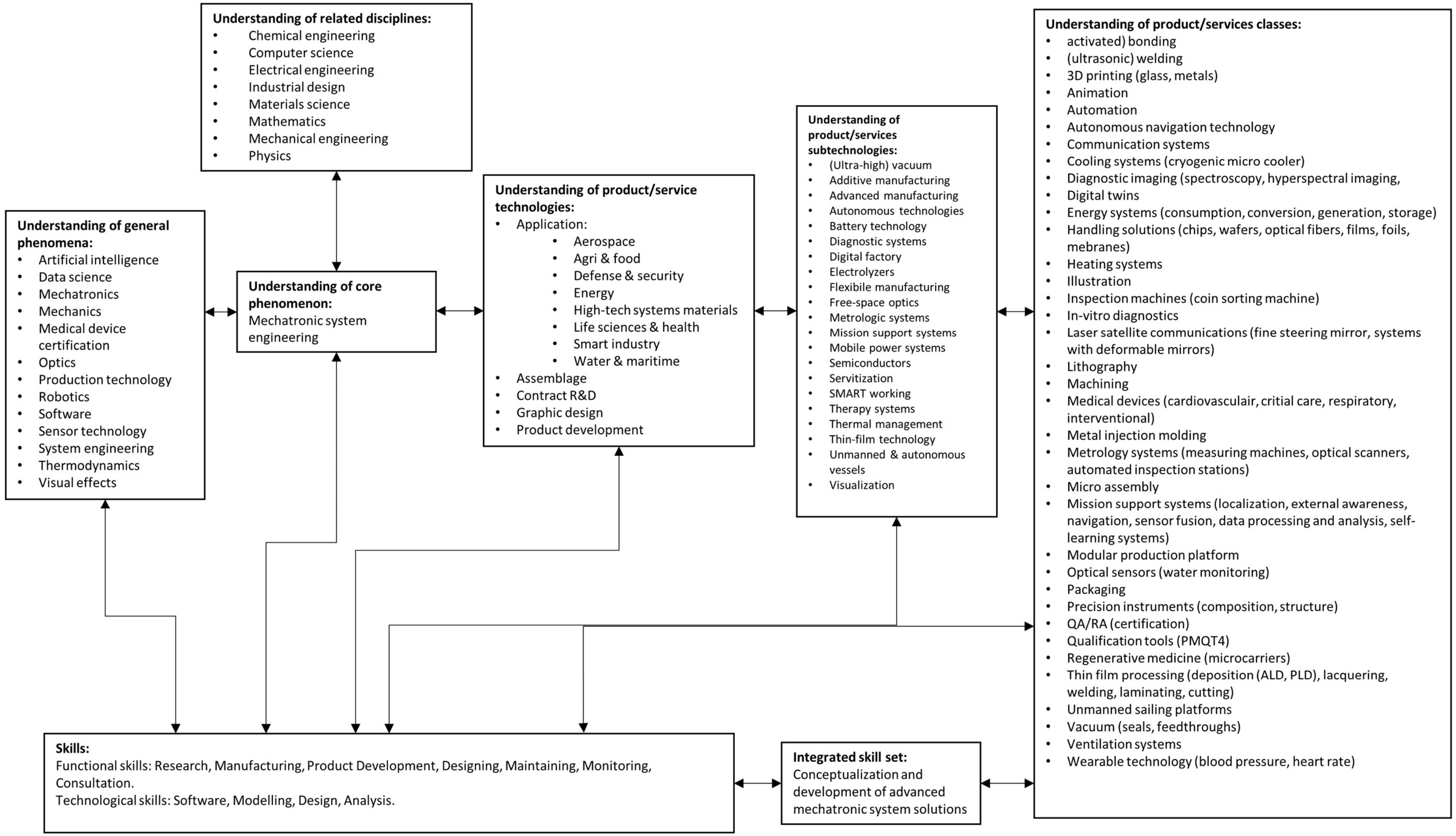

Figure 7. Core competence chart of Demcon. The chart was constructed based on the core competence chart by Edgar & Lockwood (2012) and used to identify the core competencies and technical assets of Demcon. Demcons' core competence was found to be the conceptualization and development of advanced mechatronic system solutions. Technical assets of Demcon were found in artificial intelligence, data science, mechatronics, mechanics, medical device certification, optics, production technology, robotics, software, sensor technology, system engineering, thermodynamics, and visual effects.

Discussion

This paper proposed a framework for identifying opportunities in the technological landscape of a new market for firms, based on the autoclassification of patent analysis with text mining. This paper was applied to the sports technology innovation market for Demcon. The research questions were (*RQ1*) What are current technological development trends in sports technology? (*RQ2*) What are the core competencies and technical assets of Demcon? and (*RQ3*) How do Demcons' assets match with the found sports technology development trends? To answer these questions, LDA modeling and the core competence chart by Edgar & Lockwood (2012) were used. Three sports technology development trends were found: video analysis, sports equipment, and sensor technology (see table 4). Furthermore, Demcons' core competence was found to be the conceptualization and development of mechatronic systems, associated with a wide range of technological assets in artificial intelligence, data science, mechanics, software, sensor technology, and system engineering. The found sports technology developments all match with the assets of Demcon. Opportunities for Demcon in sports technology are, therefore, suggested to lie in providing contract R&D to sports organizations with the suggested technology trends or in the acquisition of smaller companies that are active in video analysis, sports equipment, or sensor technology.

Development trends in sports technology

From the automated patent analysis, three trends in sports technology were found by the LDA model. These were video analysis, sports equipment, and sensor technology. All these categories were mentioned, to some extent, in the earlier literature review on sports technology. The overarching theme in the categories is the importance of data in sports.

The largest trend found in the patent database was video analysis. Video analysis refers to the use of images for analysis and is increasingly becoming important for sports (Shanbhag, 2021). This can be applied for different purposes. Athletes and coaches can use videos to review performances, either individual performances or in a team setting. By doing this, weaknesses can be identified and decisions made during the game can be assessed afterward (Tahan, Rady, Sleiman, Ghantous, & Merhi, 2018). Referees can use video analysis for objective decision-making in high stake situations (Michalopoulos, Raptis, & Katsini, 2019). Current challenges and developments in video analysis are focused on the enhancement of it with computer vision and artificial intelligence (Depoian, Jaques, Xie, Bailey, & Guturu, 2020; Stein et al., 2018; R. Zhang, 2022). Practical examples include robust object detection, player labeling, and trajectory visualization (Barris & Button, 2008; Mooney et al., 2015).

Sports equipment refers to all equipment used in sports. It can help athletes in improving their performance. Running shoes are an example of sports equipment and have increasingly been patented since 2017 when Nike introduced the Vaporfly with its innovative carbon-fiber plate in the midsole (Arderiu & de Fondeville, 2021; Bermon, 2021; Bermon, Garrandes, Szabo, Berkovics, & Adami, 2021). Nike was also assigned to 119 patents in this study, which places it in place nine for most frequent assignees in this study (see figure 4). Sports equipment can thus help athletes to enhance their performance directly (Muniz-Pardos et al., 2021). Sports equipment can also be used by athletes during training. Smart sports equipment, such as smart skis, golf clubs, or tennis rackets, uses sensors to gather data about the athlete and gives feedback to the athlete about their technique (Barbosa, 2018; Brown, 2020; Kos & Umek, 2018; Umek et al., 2017). Current challenges and developments in sports equipment are focused on improving the design of existing equipment to improve performance, e.g., running shoes, or on the development of smart equipment to gather data about performance.

Sensor technology refers to the manufacturing and application of sensors to conduct information about parameters. Examples of sensors include inertial sensors, optical sensors, and angle and displacement sensors. The main goal is to develop sensors that can be used by the athletes to improve efficiency and enhance performance, while the athlete is hindered as least as possible by the sensor (Menolotto, Komaris, Tedesco, O'flynn, & Walsh, 2020; Pueo & Jimenez-Olmedo, 2017). Current challenges and developments in sensor technology are focused on improving the quality of the signal, improving signal transmission, improving sensor design and manufacturing sensors to measure new parameters (James & Petrone, 2016; Kos, Milutinović, & Umek, 2019; Rein & Memmert, 2016; Salpavaara, Verho, Lekkala, & Halttunen, 2009).

The found trends in sports technology are in line with the results from the literature research. Video analysis, sports equipment, and sensor technology can be related to the found trends in wearable technology, virtual environments, and technology for refereeing. These trends also reflect and confirm the increasing importance of data in sports (Rajšp & Fister, 2020). Especially for professional athletes, data has become essential for success, such as winning the Tour de France or winning a gold medal at the world speed skating championships (Aarts, 2022; van Weel, 2022). In contrast to the literature review, eSports was not found to be a technology trend, despite it being named the biggest trend in sport in the twenty-first century (Hamari & Sjöblom, 2017). Patents in the eSports industry are mainly filed by game developers, who are behind the major driving forces of the eSports industry (Wong & Meng-Lewis, 2022). These patents are not directly related to sport, but more to the design of hardware and software used in video games. As a result, a search query focused on sports technology could thus possibly not have captured games used for eSports. It is worth mentioning here that the term 'game' was classified under topic 1 (see table 3). However, based on the other words found for topic 1, it is more likely that topic 1 represents video analysis than eSports.

Key competencies and assets of Demcon

The core competence chart by Edgar & Lockwood (2012) was used to determine the key competencies and technical assets of Demcon and can be viewed in figure 7. Competencies refer to the integration of multiple technologies to create synergy (Timokhina, 2021). For Demcon, this is mechatronic system engineering, which captures the widespread creation and diffusion of knowledge in Demcon.

Technical assets were found to arise from fields of artificial intelligence, data science, mechatronics, mechanics, medical device certification, optics, production technology, robotics, software, sensor technology, system engineering, thermodynamics, and visual effects. Technical assets held by Demcon are not isolated by project boundaries. For instance, Demcon has experience with conceptualizing and developing pressure sensors to measure blood pressure. This uses assets from medical device certification and sensor technology. Assets in sensor technology are also used for the development of autonomous systems. This free flow of technical assets fosters knowledge creation and diffusion, which is critical for gaining a competitive advantage (Garina, Garin, Kuznetsov, Romanovskaya, & Andryashina, 2020; Roth, 2003; Un & Cuervo-Cazurra, 2004). Moreover, the assets of Demcon tended to be dynamic. This is characteristic of firms in technological areas, which are subject to rapidly changing environments (Ambrosini & Bowman, 2009; Teece, Pisano, & Shuen, 2009).

Matching assets and development trends

The matching of assets with the development trends consisted of looking at what assets are needed for the development trends and looking at whether these assets were already held by Demcon. Sport technology trends for which Demcon already has assets that can be used are the most promising. This would enable Demcon an efficient use of its assets for successful business diversification.

Video analysis uses assets from different domains. These include image processing, data science, computer vision, and artificial intelligence. Demcon already has assets from artificial intelligence and data science. However, given that Demcon is a mechatronic system engineering firm, it mostly has experience in the implementation of these assets in hardware and less in algorithm development. Challenges such as robust object detection and player labeling are mostly focused on data algorithms and less on the hardware used in video analysis. A match with video analysis is therefore limited. Demcon could focus on the acquisition of companies focused on video algorithm development to gain the assets needed, but this is not really in line with its core competence. Alternatively, Demcon could focus on the development of hardware for video analysis, such as cameras or other sensors.

Sports equipment uses assets from (bio)mechanics, materials engineering, and design. Demcon has experience with all of these assets. The design of sports equipment, particularly used during races or games, seems a less promising option for Demcon. This is a highly specialized field and also has to deal with strict design rules and regulations (Brown, 2020; Grimshaw, 2019). Demcon could, however, focus on the manufacturing of equipment used during training, such as smart sports equipment, which is less restricted by rules and regulations. This could consist of developing smart sports equipment which can be used by athletes to track and improve their training.

Sensor technology refers to the development and manufacturing of sensors to gather data. It uses assets from electrical engineering, wearable technology, physics, and computer science. Demcon has experience in developing sensors, as it has not only developed sensors for other practices, such as smart industry but also has Johan Sports, which develops sensors for teams. Demcon could further advance in the field of sensor technology, in which the development of sensors seems to have the highest match for Demcon, such as done with Johan Sports. However, the improvement of signal quality and data transmission could also be interesting for Demcon. Sensor technology seems to be a promising option for Demcon in sports technology.

Technological roadmap for sports technology

Three sports technology trends were extracted during the patent analysis. These trends reflect the current technological developments in the sports technology field, which provides insight into the building blocks of this field (S. Lee, 2013; Paap, 2020; Song et al., 2016). As such, the found trends can be used to construct a business plan for Demcon in sports technology, which is for Demcon an unknown market (Arasti & Bagheri Moghaddam, 2010; H. Zhang et al., 2021). The found technology trends reflect what technologies are currently the most used in sports and which thus hold the most potential to be profitable. Moreover, Demcon operates in a business-to-business context and is employed by companies for contract R&D. When entering a new market, knowing which technology developments are aligned with its assets could guide Demcon in searching more targeted for potential clients. Alternatively, it could help Demcon in identifying those firms which could be interesting for acquisition as a means for external innovation, as was done with Johan Sports (Brueller & Capron, 2021). It is easier to search specifically for firms in a certain field than searching for sports technology firms in general, which is quite a broad field.

Based on the matches between the assets and the development trends, video analysis, sports equipment, and sensor technology could all be considered promising opportunities for Demcon. Demcons' core competence indicates that the development of data algorithms seems to be less of a match, but the development of hardware could be an interesting opportunity. Given that Demcon has acquired Johan Sports, Demcon already has experience with the sensor technology market. Further practical applications for Demcon in sports technology are provided under implications.

Assessment of framework

This study constructed a framework for identifying opportunities in the technological landscape of a new market for firms. This was done by using autoclassification of patent analysis with text mining (see also figure 2). Automated patent analysis can be used to create technological roadmaps, but research so far had not tested its applicability for strategic decision-making in a business-to-business context. To identify business opportunities from a technology roadmap, the core competencies and technical assets of a business need to be taken into account. This way, assets can be used more efficiently, which enhances successful business diversification. Therefore, this framework adopted an approach in which both technology development trends from a patent analysis and technical assets were used to analyze how successful business diversification could be realized. This framework was then applied in the sports technology market for Demcon.

The generalizability of the framework seems quite high. In most markets, specifically technological markets, patents are increasingly being used to ensure the rights of an invention, regardless of whether this is in a business-to-business or a business-to-consumer context (Ménière & Kendrick, 2021). For businesses that operate mainly in a service-oriented industry, the use of patents to identify trends seems less usable, as services are generally less patented (Dotzel & Shankar, 2019). However, for most other businesses it seems plausible that this framework could be used for business diversification. Besides that, unlike other sources of web scraping, ethical and legal issues are less of a concern when patents are being used, as patents are publicly available data (Krotov & Silva, 2018; Luscombe, Dick, & Walby, 2022). The framework can, thus, be used by various companies in various markets, without the risk of ethical or legal issues.

The quality of the extracted technology development trends and the technical assets of the companies is quite sensitive to subjectivity. First, the technology development trends are identified by interpreting associated words from the topic modeling. This interpretation is quite sensitive to human interpretability, as two users could extract different trends from the same set of keywords. Second, the construction of the core competence chart is also influenced by subjectivity. Therefore, it is recommended that for successful use of the framework for business diversification, multiple people are involved, both in the interpretation of the topic words and the construction of the core competence chart. Furthermore, it is also recommended, especially in the interpretation of the technology development trends, to first gain some background knowledge of common technologies used in the market.

The used framework captures the technologies which are currently being patented. As such, the framework shows the current push of technology in a market (Toro-Jarrín et al., 2016). However, this does not indicate whether these technologies are actually needed by end-users, i.e. the pull of the market. This questions the quality of generating business opportunities based on the push of technology in the market, as pushed technologies do not necessarily become profitable. The use of this framework, however, is recommended to be at the beginning stage of market research. It, therefore, rather serves to gain a general understanding of what a market looks like, which directions could provide potential business opportunities and what the next step could be, than to generate a very detailed business plan.

Overall, the framework in this study can be a valuable tool to assess new market opportunities with patents. It gives an indication of current technology development trends in a market and which of these trends have assets that are already held by the company. This can then be used for enhancing business diversification. Some issues exist with regard to subjectivity, but a potential solution for this would be to involve multiple people in the application of the framework.

Limitations

Several limitations exist in the current study. First, this study employed an LDA model to extract topics in sports technology from patent abstracts and titles. This is a widespread probabilistic method to extract topics from text (Jelodar et al., 2019; Rieger, Jentsch, & Rahnenführer, 2020). While patent analysis in itself is regarded as a reliable method, the reliability of employing LDA is limited, as the extracted topics are sensitive to human interpretation (Doogan & Buntine, 2021; Lau, Newman, & Baldwin, 2014). Objective measures to asses topics from LDA models exist and are based on coherence measured and automated labeling (Bhatia, Lau, & Baldwin, 2017). However, automated evaluations do not match human evaluations closely, which questions the validity of automated evaluations (Hoyle et al., 2021). Therefore, human evaluations are still considered superior to current automated analyses in terms of validity.

Second, research on the use of patents as a source of innovation is conflicted. On one hand, it is argued that when an invention is granted a patent, it does not necessarily mean that an invention will lead to innovation and technological progress (Boldrin & Levine, 2013; O'toole, 2021). Additionally, it is argued that only a small part of all inventions are being patented and often without the involvement of the actual inventors (Basberg, 1987; Gittelman, 2008). On the other hand, it is argued that patents foster innovation by enabling companies to utilize external knowledge (Comai, 2020). Patents are considered essential for the exchange of knowledge in an industry (Orsenigo & Sterzi, 2010). Besides that, even if not all inventions are being patented, it is thought that the majority of all inventions are (Mansfield, 1986). As a result, patents are considered the basis for measuring technology developments (Boone, Lokshin, Guenter, & Belderbos, 2019; Dobrzanski, Bobowski, Chrysostome, Velinov, & Strouhal, 2021; Rubilar-Torrealba, Chahuán-Jiménez, & de la Fuente-Mella, 2022; Taques, López, Basso, & Areal, 2021). This study adopts the view that patents could be a valuable tool for measuring technology developments, i.e., the push of technology in a market. The overall aim of this study was to use patent analysis to better understand a new market. It was thus less focused on directly measuring innovation and creating a detailed business plan, but more focused on building an understanding of what a market looks like.

Third, the use of web scraping and artificial intelligence could raise issues regarding the validity and ethics of the results. First, the use of web scraping often results in big datasets for analysis, i.e., big data, which could have negatively impacted the validity of the results. The large size and volume of big data have as a risk that the quality of the used data is lower, i.e., great quantity does not necessarily mean great quality (Boegershausen et al., 2022; Cai & Zhu, 2015; Fan, Han, & Liu, 2014; García-Gil, Luengo, García, & Herrera, 2019; J. Liu et al., 2016; Lovelace, Birkin, Cross, & Clarke, 2016). As a result, the included patents might not have all been related to sports technology, which would mean that the found topics might give a good indication of trends in this particular dataset, but not of technology development trends in sport in general. It can, for instance, be observed in table 6 (Appendix-D) that some patents are included of which it is not directly clear how they are related to sports technology. Second, the use of artificial intelligence might result in biased or unexplainable results. This knowledge transfer of the model to the user is a common problem in machine learning and is enforced by the black box in which the model operates (De Bruyn, Viswanathan, Beh, Brock, & von Wangenheim, 2020). It is often unclear how a model comes to a classification, which makes it harder to interpret the results. Third, the extraction of data from websites poses legal and ethical risks, specifically if the data may be considered private (Boegershausen et al., 2022). However, this is less of a concern for this study, as patent data is publicly available and can freely be accessed.

Future research

Further research on using patent analysis for the identification of business opportunities could focus on various directions. The framework could be applied in other markets to verify its generalizability. Besides that, the framework could be extended by also including other sources, instead of just patents. Patents only capture current trends in technology developments (i.e., the push of technology) but do not contain information about the needs of end-users (i.e., the pull of the market). This was considered sufficient for the first exploration of a new market but could be an interesting topic for future research. For example, for the sports technology market, it could have been interesting to include the needs of athletes, coaches, and data analysts. This could be done in a qualitative way, by conducting interviews, or by building further on the field of big data by web scraping reviews, news articles, and reports from the internet. Additionally, future research could focus on measures for assessing the quality of a patent dataset. The current study only analyzed the cooperative patent classifications (CPCs) retrospectively, but it could be interesting to see whether CPCs could be used as a sort of filter to improve the quality of the included patents. This filter would need to find a compromise between the inclusion of irrelevant patents and the exclusion of relevant patents, which could be a challenging task. This could potentially also address the aforementioned validity concerns on the quality of big data.

Implications

Video analysis, sports equipment, and sensor technology all have the potential to be promising market fields for Demcon. The found sports technology trends give insight into the current landscape for sports technology. Therefore, pursuing business in the found trends holds the potential to create a profitable business for Demcon. Because sports technology is a new market for Demcon, knowing what the current landscape looks like can give Demcon guidance on what is currently profitable in sports technology and can help Demcon in searching for clients for contract R&D or companies for acquisition by narrowing the market. It is generally easier to search for specific companies than to search for sports companies in general. However, it is worth emphasizing here that the found technology trends for Demcon are not the only options for Demcon in sports technology. Rather, it serves as a guide for Demcon in what is currently the most being used in sports technology and serves thus as an inspiration for what could be profitable endeavors for Demcon in sports technology.

The next step for Demcon would be to further research video analysis, sports equipment, and sensor technology. This could consist of not only getting a deeper understanding of the found trends but also on getting information which players exist in each of these fields. This could consist of interviewing experts of the technologies in the Netherlands. To contact these experts, it could be helpful to look at existing clients, in addition to desk research. It could also be interesting to interview potential end-users, such as athletes, coaches, and data scientists. This could give a better insight into the needs of end-users. The interviews, in addition to desk research, could then be used to create a list of potential leads to contact in video analysis, sports equipment, and sensor technology for projects. Based on earlier clients of Demcon, it would be preferred to focus on smaller companies. R&D is often not available in smaller companies, which is thus where opportunities could lie for Demcon. Sports organizations could also be interesting potential clients for Demcon, such as football clubs or cycling teams. For these organizations, sports technology holds the potential to enhance their performance, which Demcon could provide with custom-made solutions.

It is worth highlighting here that Demcon already has experience in the sensor technology market, given the acquisition of Johan Sports. Specifically for sensor technology, Demcon could thus focus on the growth of Johan Sports, for instance by product diversification, instead of searching more generally for business opportunities in sensor technology.

This study further contributes to the application of automated classification of patents for mapping technology. It provides a framework that can generally be applied for identifying opportunities in technological markets and aligning these with a firm's assets for business opportunities to construct a technological roadmap. It is the first research to test the applicability of automated patent analysis for strategic decision-making for business diversification by taking core competencies and technical assets into account. This study builds on the use of artificial intelligence in market research, specifically on the application of text mining for the identification of market trends (Kim & Woo, 2022; Lau et al., 2012; Park et al., 2018; Pek & Lim, 2019; Ploessl et al., 2021; Sung & Yeo, 2019; Xu et al., 2020; H. Zhang et al., 2021). This study could serve as inspiration for the application of unsupervised machine learning in patent analysis and could serve as a guide for marketers for the integration of automated patent analysis in the exploration of new markets. The framework in this study could also be used by other organizations to identify new markets.

Conclusion

This research aimed to identify opportunities for Demcon in the sports technology market. Based on an automated patent classification with topic modeling, three trends could be extracted from sports technology patents: video analysis, sports equipment, and sensor technology. Based on conceptual analysis, Demcons' core competence was found to be mechatronic system engineering, which is associated with a wide range of technical assets. As a result, future opportunities for Demcon in sports technology are recommended to focus on knowledge expansion and lead generation in video analysis, sports equipment, and sensor technology. This research has shown that automated patent classification could be a valid tool for capturing technology development trends, but that concerns could arise about the validity of results generated from big data and artificial intelligence. Further research is needed to address validity concerns of big data and integration of the needs of end-users in patent analysis. This research contributes to the expansion of knowledge on sports technology and the application of machine learning methods in market trend identification.

References

Aarts, D. (2022). Wielrennen als een prof van Jumbo-Visma met Leids datamodel. Retrieved August 4, 2022, from https://www.universiteitleiden.nl/nieuws/2022/07/wielrennen-als-een-prof-van-jumbo-visma-met-leids-datamodel

Adesida, Y., Papi, E., & Mcgregor, A. H. (2019). Exploring the Role of Wearable Technology in Sport Kinematics and Kinetics: A Systematic Review. *Sensors*. https://doi.org/10.3390/s19071597

Aharonson, B. S., & Schilling, M. A. (2016). Mapping the technological landscape: Measuring technology distance, technological footprints, and technology evolution. *Research Policy*, *45*(1), 81–96. https://doi.org/10.1016/J.RESPOL.2015.08.001

Ahn, S., Couture, S. V., Cuzzocrea, A., Dam, K., Grasso, G. M., Leung, C. K., … Wodi, B. H. (2019). A Fuzzy Logic Based Machine Learning Tool for Supporting Big Data Business Analytics in Complex Artificial Intelligence Environments. *IEEE International Conference on Fuzzy Systems*, *2019-June*. https://doi.org/10.1109/FUZZ-IEEE.2019.8858791

Alstott, J., Triulzi, G., Yan, B., & Luo, J. (2017). Mapping technology space by normalizing patent networks. *Scientometrics*, *110*(1), 443–479. https://doi.org/10.1007/s11192-016-2107-y

Ambrosini, V., & Bowman, C. (2009). What are dynamic capabilities and are they a useful construct in strategic management? *International Journal of Management Reviews*, *11*(1), 29–49. https://doi.org/10.1111/J.1468-2370.2008.00251.X

Ami, J., Ubaid, A. M., Abuseem, A. J., Alzaabi, Y. J., Alhammadi, A. H., & Dweiri, F. T. (2020). Technology Mapping: Definitions, Types, and Applications. In *Proceedings of the International Conference on Industrial Engineering and Operations Management* (pp. 377–386). Retrieved from www.sharjah.ac.ae

Ampornphan, P., & Tongngam, S. (2017). Patent knowledge discovery using data analytics. In *ACM International Conference Proceeding Series* (pp. 42–46). Association for Computing Machinery. https://doi.org/10.1145/3176653.3176721

Antons, D., Grünwald, E., Cichy, P., & Salge, T. O. (2020). The application of text mining methods in innovation research: current state, evolution patterns, and development priorities. In *R and D Management* (Vol. 50, pp. 329–351). John Wiley & Sons, Ltd. https://doi.org/10.1111/radm.12408

Arasti, M. R., & Bagheri Moghaddam, N. (2010). Use of technology mapping in identification of fuel cell sub-technologies. *International Journal of Hydrogen Energy*, *35*(17), 9516–9525. https://doi.org/10.1016/J.IJHYDENE.2010.05.071

Arderiu, A., & de Fondeville, R. (2021). Influence of Vaporfly shoe on sub-2 hour marathon and other top running performances. Retrieved from http://arxiv.org/abs/2104.08509

Aroganam, G., Manivannan, N., & Harrison, D. (2019). Review on Wearable Technology Sensors Used in Consumer Sport Applications. *Sensors 2019, Vol. 19, Page 1983*, *19*(9), 1983. https://doi.org/10.3390/S19091983

Bach, M. P., Krstič, Ž., Seljan, S., & Turulja, L. (2019). Text Mining for Big Data Analysis in Financial Sector: A Literature Review. *Sustainability 2019, Vol. 11, Page 1277*, *11*(5), 1277. https://doi.org/10.3390/SU11051277

Bal, B., & Dureja, G. (2012). Hawk Eye: A Logical Innovative Technology Use in Sports for Effective Decision Making. *Sport Science Review*, *21*(1–2), 107–119. https://doi.org/10.2478/v10237-012-0006-6

Barbosa, T. M. (2018). Smart sport equipment: reshaping the sports landscape. *Motricidade*, *14*(2–3), 1–2. https://doi.org/10.6063/motricidade.15232

Barde, B. V., & Bainwad, A. M. (2017). An overview of topic modeling methods and tools. In *Proceedings of the 2017 International Conference on Intelligent Computing and Control Systems, ICICCS 2017* (Vol. 2018-Janua, pp. 745–750). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICCONS.2017.8250563

Barris, S., & Button, C. (2008). A review of vision-based motion analysis in sport. *Sports Medicine*. https://doi.org/10.2165/00007256-200838120-00006

Basberg, B. L. (1987). Patents and the measurement of technological change: A survey of the literature. *Research Policy*, *16*(2–4), 131–141. https://doi.org/10.1016/0048-7333(87)90027-8

Bermon, S. (2021, July 13). Evolution of distance running shoes: Performance, injuries, and rules. *Journal of Sports Medicine and Physical Fitness*. Edizioni Minerva Medica. https://doi.org/10.23736/S0022-4707.21.12728-8

Bermon, S., Garrandes, F., Szabo, A., Berkovics, I., & Adami, P. E. (2021). Effect of Advanced Shoe Technology on the Evolution of Road Race Times in Male and Female Elite Runners. *Frontiers in Sports and Active Living*, *3*, 46. https://doi.org/10.3389/FSPOR.2021.653173/BIBTEX

Bhatia, S., Lau, J. H., & Baldwin, T. (2017). An automatic approach for document-level topic model evaluation. In *CoNLL 2017 - 21st Conference on Computational Natural Language Learning, Proceedings* (pp. 206–215). https://doi.org/10.18653/v1/k17-1022

Blei, D. M., Ng, A. Y., & Edu, J. B. (2003). Latent Dirichlet Allocation Michael I. Jordan. *Journal of Machine Learning Research*, *3*, 993–1022.

Boegershausen, J., Datta, H., Borah, A., & Stephen, A. T. (2022). Fields of Gold: Scraping Web Data for Marketing Insights. *Journal of Marketing*, 002224292211007. https://doi.org/10.1177/00222429221100750

Boldrin, M., & Levine, D. K. (2013). The case against patents. In *Journal of Economic Perspectives* (Vol. 27, pp. 3–22). https://doi.org/10.1257/jep.27.1.3

Bonaccorso, G. (2017). *Machine learning algorithms*. Packt Publishing Ltd.

Boone, C., Lokshin, B., Guenter, H., & Belderbos, R. (2019). Top management team nationality diversity, corporate entrepreneurship, and innovation in multinational firms. *Strategic Management Journal*, *40*(2), 277–302. https://doi.org/10.1002/SMJ.2976

Bosch, V. (2016). Big Data in Market Research: Why More Data Does Not Automatically Mean Better Information. *GfK Marketing Intelligence Review*, *8*(2), 56–63. https://doi.org/10.1515/gfkmir-2016-0017

Boyd-Graber, J., Mimno, D., & Newman, D. (2014). Care and feeding of topic models: Problems, diagnostics, and improvements. In *Handbook of Mixed Membership Models and Their Applications* (Vol. 39, pp. 225–254). https://doi.org/10.1201/b17520

Bregman, D. J. J. (2019). *Triple helix samenwerking in de sport*.

Brink, M. (2001). *Development of a method to forecast future systems in the forest engineering value chain*. Retrieved from http://scholar.sun.ac.za/handle/10019.1/52188

Brown, J. (2020). The Creation and Regulation of Sports Equipment: Implications for the Future. *ESLJ*, *18*, 1.

Brueller, N. N., & Capron, L. (2021). Acquisitions of Startups by Incumbents: The 3 Cs of Co-speCializaTion from sTarTup inCepTion To posT-merger inTegraTion. *California Management Review*, *63*(3), 70–93. https://doi.org/10.1177/0008125621996502

Buenano-Fernandez, D., Gonzalez, M., Gil, D., & Lujan-Mora, S. (2020). Text Mining of Open-Ended Questions in Self-Assessment of University Teachers: An LDA Topic Modeling Approach. *IEEE Access*, *8*, 35318–35330. https://doi.org/10.1109/ACCESS.2020.2974983

Cai, L., & Zhu, Y. (2015). The challenges of data quality and data quality assessment in the big data era. *Data Science Journal*, *14*(0). https://doi.org/10.5334/DSJ-2015-002/METRICS/

Cascini, G., Fantechi, A., & Spinicci, E. (2004). Natural language processing of patents and technical documentation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *3163*, 508–520. https://doi.org/10.1007/978-3-540-28640-0_48

Chan, C. K., Ng, K. W., Ang, M. C., Ng, C. Y., & Kor, A. L. (2021). Sustainable Product Innovation Using Patent Mining and TRIZ. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 13051 LNCS, pp. 287–298). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-030-90235-3_25

Chhaya, K., Khanzode, A., & Sarode, R. D. (2020). Advantages and Disadvantages of Artificial Intelligence and Machine Learning: A Literate Review. *International Journal of Library & Information Science (IJLIS)*, *9*, 3. Retrieved from http://www.iaeme.com/IJLIS/index.asp30http://www.iaeme.com/IJLIS/issues.asp?JType=IJLIS&VType=9&IType=1JournalImpactFactor

Choi, S., Yoon, J., Kim, K., Lee, J. Y., & Kim, C. H. (2011). SAO network analysis of patents for technology trends identification: A case study of polymer electrolyte membrane technology in proton exchange membrane fuel cells. *Scientometrics*, *88*(3), 863–883. https://doi.org/10.1007/S11192-011-0420-Z

Comai, A. (2020). A new approach for detecting open innovation in patents: the designation of inventor. *Journal of Technology Transfer*, *45*(6), 1797–1822. https://doi.org/10.1007/s10961-019-09763-8

Cruz, L. B., & Boehe, D. M. (2008). CSR in the global marketplace: Towards sustainable global value chains. *Management Decision*, *46*(8), 1187–1209. https://doi.org/10.1108/00251740810901381

De Bruyn, A., Viswanathan, V., Beh, Y. S., Brock, J. K. U., & von Wangenheim, F. (2020). Artificial Intelligence and Marketing: Pitfalls and Opportunities. *Journal of Interactive Marketing*, *51*, 91–105. https://doi.org/10.1016/j.intmar.2020.04.007

Depoian, A. C., Jaques, L. E., Xie, D., Bailey, C. P., & Guturu, P. (2020). Computer vision learning techniques for sports video analytics: removing overlays, 22. https://doi.org/10.1117/12.2560888

DiFrancisco-Donoghue, J., Werner, W. G., Douris, P. C., & Zwibel, H. (2020). Esports players, got muscle? Competitive video game players' physical activity, body fat, bone mineral content, and muscle mass in comparison to matched controls. *Journal of Sport and Health Science*. https://doi.org/10.1016/j.jshs.2020.07.006

Dobrzanski, P., Bobowski, S., Chrysostome, E., Velinov, E., & Strouhal, J. (2021). Toward innovation-driven competitiveness across african countries: An analysis of efficiency of R&D expenditures. *Journal of Competitiveness*. https://doi.org/10.7441/joc.2021.01.01

Doogan, C., & Buntine, W. (2021). Topic Model or Topic Twaddle? Re-evaluating Semantic Interpretability Measures (pp. 3824–3848). https://doi.org/10.18653/v1/2021.naacl-main.300

Dotzel, T., & Shankar, V. (2019). The Relative Effects of Business-to-Business (vs. Business-to-Consumer) Service Innovations on Firm Value and Firm Risk: An Empirical Analysis. *Journal of Marketing*, *83*(5), 133–152. https://doi.org/10.1177/0022242919847221

Doz, Y. (2011, June 15). Qualitative research for international business. *Journal of International Business Studies*. Palgrave. https://doi.org/10.1057/jibs.2011.18

Ducange, P., Pecori, R., & Mezzina, P. (2017). A glimpse on big data analytics in the framework of marketing strategies. *Soft Computing 2017 22:1*, *22*(1), 325–342. https://doi.org/10.1007/S00500-017-2536-4

Düking, P., Stammel, C., Sperlich, B., Sutehall, S., Muniz-Pardos, B., Lima, G., … Pitsiladis, Y. P. (2018). Necessary steps to accelerate the integration of wearable sensors into recreation and competitive sports. *Current Sports Medicine Reports*, *17*(6), 178–182. https://doi.org/10.1249/JSR.0000000000000495

Edgar, W. B., & Lockwood, C. A. (2012). Understanding, finding, and conceptualizing core competence depth: A framework, guide, and generalization for corporate managers and research professionals. *Academy of Strategic Management Journal*, *11*(2), 63–92.

Erevelles, S., Fukawa, N., & Swayne, L. (2015). Big Data Consumer Analytics and the Transformation of Marketing. *Article in Journal of Business Research*. https://doi.org/10.1016/j.jbusres.2015.07.001

Evans, M. S. (2014). A Computational Approach to Qualitative Analysis in Large Textual Datasets. *PLOS ONE*, *9*(2), e87908. https://doi.org/10.1371/JOURNAL.PONE.0087908

Fan, J., Han, F., & Liu, H. (2014). Challenges of Big Data analysis. *National Science Review*, *1*(2), 293–314. https://doi.org/10.1093/NSR/NWT032

Fattori, M., Pedrazzi, G., & Turra, R. (2003). Text mining applied to patent mapping: A practical business case. *World Patent Information*, *25*(4), 335–342. https://doi.org/10.1016/S0172-2190(03)00113-3

Feng, J., Liu, Z., & Feng, L. (2021). Identifying opportunities for sustainable business models in manufacturing: Application of patent analysis and generative topographic mapping. *Sustainable Production and Consumption*, *27*, 509–522. https://doi.org/10.1016/J.SPC.2021.01.021

Forestal, R. L., Liu, S. H., Pi, S. M., & Li, E. Y. (2021). Text mining-based patent analysis of blockchain technology applications. In *Proceedings of the International Conference on Electronic Business (ICEB)* (Vol. 21, pp. 476–483). Retrieved from https://www.researchgate.net/publication/358262475

Fortes, L. S., Almeida, S. S., Praça, G. M., Nascimento-Júnior, J. R. A., Lima-Junior, D., Barbosa, B. T., & Ferreira, M. E. C. (2021). Virtual reality promotes greater improvements than video-stimulation screen on perceptual-cognitive skills in young soccer athletes. *Human Movement Science*, *79*, 102856. https://doi.org/10.1016/J.HUMOV.2021.102856

Fossey, E., Harvey, C., McDermott, F., & Davidson, L. (2002). Understanding and evaluating qualitative research. *Australian & New Zealand Journal of Psychiatry*, *36*(6), 717–732. https://doi.org/10.1046/j.1440-1614.2002.01100.x

Gabbett, T. J., Nassis, G. P., Oetter, E., Pretorius, J., Johnston, N., Medina, D., … Ryan, A. (2017). The athlete monitoring cycle: A practical guide to interpreting and applying training monitoring data. *British Journal of Sports Medicine*. https://doi.org/10.1136/bjsports-2016-097298

García-Gil, D., Luengo, J., García, S., & Herrera, F. (2019). Enabling Smart Data: Noise filtering in Big Data classification. *Information Sciences*, *479*, 135–152. https://doi.org/10.1016/j.ins.2018.12.002

Garina, E. P., Garin, A. P., Kuznetsov, V. P., Romanovskaya, E. V., & Andryashina, N. S. (2020). Definition of Key Competences of Companies in the “Product-Production” System. In *Lecture Notes in Networks and Systems* (Vol. 73, pp. 737–746). Springer. https://doi.org/10.1007/978-3-030-15160-7_74

Gerken, J. M., & Moehrle, M. G. (2012, June). A new instrument for technology monitoring: Novelty in patents measured by semantic patent analysis. *Scientometrics*. Springer. https://doi.org/10.1007/s11192-012-0635-7

Gerken, J., Moehrle, M. G., & Walter, L. (2010). Semantische Patentlandkarten zur Analyse technologischen Wandels: Eine Längsschnittstudie aus der Allradtechnik. In *6. Symposium für Vorausschau und Technologieplanung*. Paderborn: Heinz Nixdorf Institut.

Gittelman, M. (2008, August 1). A note on the value of patents as indicators of innovation: Implications for management research. *Academy of Management Perspectives*. Academy of Management Briarcliff Manor, NY. https://doi.org/10.5465/AMP.2008.34587992

Gottschalk, C., Tewes, S., & Niestroj, B. (2020). The Innovation of Refereeing in Football through AI. *International Journal of Innovation and Economics Development*, *6*(2), 35–54. https://doi.org/10.18775/ijied.1849-7551-7020.2015.62.2004

Grimshaw, P. (2019). Sports equipment. In *Instant Notes in Sport and Exercise Biomechanics: Second Edition* (pp. 364–374). Taylor and Francis. https://doi.org/10.4324/9781315636764-39

Hague, P., Harrision, M., Cupman, J., & Truman, O. (2016). *Market research in practice: An introduction to gaining greater market insight*. (J. Volich, Ed.) (3rd ed.). London: Kogan Page Limited.

Hallmann, K., & Giel, T. (2018). eSports – Competitive sports or recreational activity? *Sport Management Review*. https://doi.org/10.1016/j.smr.2017.07.011

Hamari, J., & Sjöblom, M. (2017). What is eSports and why do people watch it? *Internet Research*, *27*(2), 211–232. https://doi.org/10.1108/INTR-04-2016-0085/FULL/XML

Hasan, M., Rahman, A., Karim, M. R., Khan, M. S. I., & Islam, M. J. (2021). Normalized approach to find optimal number of topics in latent dirichlet allocation (lda). *Advances in Intelligent Systems and Computing*, *1309*, 341–354. https://doi.org/10.1007/978-981-33-4673-4_27/FIGURES/7

Hennink, M. M., Hutter, I., & Bailey, A. (2020). *Qualitative research methods* (2. edition). London: SAGE Publications Ltd.

Hoyle, A., Goel, P., Peskov, D., Hian-Cheong, A., Boyd-Graber, J., & Resnik, P. (2021). Is Automated Topic Model Evaluation Broken?: The Incoherence of Coherence. In *Advances in Neural Information Processing Systems* (Vol. 3, pp. 2018–2033).

Huang, C. Y., Kuo, Y. T., Wang, L. C., Hsiao, Y. H., & Yang, C. L. (2021). Technology Roadmapping for the GAA-FETs Based on the Methods of Patent Mining, Fuzzy Knapsack Problem, and Fuzzy Competence Set Expansion. In *2021 International Conference on Fuzzy Theory and Its Applications, iFUZZY 2021*. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/iFUZZY53132.2021.9605088

Hung, J. L., He, W., & Shen, J. (2020). Big data analytics for supply chain relationship in banking. *Industrial Marketing Management*, *86*, 144–153. https://doi.org/10.1016/J.INDMARMAN.2019.11.001

Iwayama, M., & Tokunaga, T. (1994). A probabilistic model for text categorization: Based on a single random variable with multiple values. In *4th Conference on Applied Natural Language Processing, ANLP 1994 - Proceedings* (pp. 162–167).

James, D. A., & Petrone, N. (2016). Sensors and wearable technologies in sport: Technologies, trends and approaches for implementation. In *SpringerBriefs in Applied Sciences and Technology* (pp. 1–49). Retrieved from http://www.springer.com/series/8884

Janssen, M., Walravens, R., Thibaut, E., Scheerder, J., Brombacher, A., & Vos, S. (2020). Understanding different types of recreational runners and how they use running-related technology. *International Journal of Environmental Research and Public Health*, *17*(7). https://doi.org/10.3390/ijerph17072276

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, *78*(11), 15169–15211. https://doi.org/10.1007/s11042-018-6894-4

Jenny, S. E., Manning, R. D., Keiper, M. C., & Olrich, T. W. (2017). Virtual(ly) Athletes: Where eSports Fit Within the Definition of "Sport." *Quest*, *69*(1), 1–18. https://doi.org/10.1080/00336297.2016.1144517

Jha, A., Dave, M., & Madan, S. (2016). A review on the study and analysis of big data using data mining techniques. *International Journal of Latest Trends in Engineering and Technology (IJLTET)*. Retrieved from https://www.academia.edu/download/45797782/4.pdf

Kane, D. (2017). Recognizing ESports as a Sport. *Sport Journal*, *20*, 1–9. Retrieved from https://www.researchgate.net/publication/317929457

Kang, Y., Cai, Z., Tan, C. W., Huang, Q., & Liu, H. (2020). Natural language processing (NLP) in management research: A literature review. *Journal of Management Analytics*, *7*(2), 139–172. https://doi.org/10.1080/23270012.2020.1756939

Kasemsap, K. (2016). Text mining: Current trends and applications. In *Web Data Mining and the Development of Knowledge-Based Decision Support Systems* (pp. 338–358). IGI Global. https://doi.org/10.4018/978-1-5225-1877-8.ch017

Khamzina, M., Parab, K. V., An, R., Bullard, T., & Grigsby-Toussaint, D. S. (2020). Impact of Pokémon Go on Physical Activity: A Systematic Review and Meta-Analysis. *American Journal of Preventive Medicine*, *58*(2), 270–282. https://doi.org/10.1016/J.AMEPRE.2019.09.005

Kherwa, P., & Bansal, P. (2020). Topic Modeling: A Comprehensive Review. *EAI Endorsed Transactions on Scalable Information Systems*, *7*(24), 1–16. https://doi.org/10.4108/eai.13-7-2018.159623

Kim, S., & Woo, H. (2022). Big data analysis of trends in the second-hand apparel market: a comparison between 2014 and 2019. *Research Journal of Textile and Apparel*, *26*(2), 138–155. https://doi.org/10.1108/RJTA-12-2020-0139

Kong, D., Zhou, Y., Liu, Y., & Xue, L. (2017). Using the data mining method to assess the innovation gap: A case of industrial robotics in a catching-up country. *Technological Forecasting and Social Change*, *119*, 80–97. https://doi.org/10.1016/J.TECHFORE.2017.02.035

Kos, A., Milutinović, V., & Umek, A. (2019). Challenges in wireless communication for connected sensors and wearable devices used in sport biofeedback applications. *Future Generation Computer Systems*, *92*, 582–592. https://doi.org/10.1016/j.future.2018.03.032

Kos, A., & Umek, A. (2018). Smart sport equipment: SmartSki prototype for biofeedback applications in skiing. *Personal and Ubiquitous Computing*, *22*(3), 535–544. https://doi.org/10.1007/S00779-018-1146-1/FIGURES/12

Krotov, V., & Silva, L. (2018). Legality and ethics of web scraping. In *Americas Conference on Information Systems 2018: Digital Disruption, AMCIS 2018*. Retrieved from https://www.researchgate.net/publication/324907302

Kuitert, G. (2016). De Ondernemer | "Uitvindersfabriek" Demcon in Enschede wordt zelf…. *De Ondernemer*. Retrieved from https://www.deondernemer.nl/actueel/uitvindersfabriek-demcon-in-enschede-wordt-zelf-merk-in-medische-robots~490566?referrer=https%3A%2F%2Fwww.google.com%2F

Lau, J. H., Collier, N., & Baldwin, T. (2012). On-line trend analysis with topic models: Twitter trends detection topic model online. In *24th International Conference on Computational Linguistics - Proceedings of COLING 2012: Technical Papers* (pp. 1519–1534). Retrieved from http://twendr.com/

Lau, J. H., Newman, D., & Baldwin, T. (2014). Machine reading tea leaves: Automatically evaluating topic coherence and topic model quality. In *14th Conference of the European Chapter of the Association for Computational Linguistics 2014, EACL 2014* (pp. 530–539). https://doi.org/10.3115/v1/e14-1056

Le, D.-N., Le, C. Van, Tromp, J. G., & Nguyen, G. N. (2018). *Emerging technologies for health and medicine virtual reality, augmented reality, artificial intelligence, internet of things, robotics, industry 4.0 (PDFDrive)*. Wiley-Scrivener.

Le Noury, P., Polman, R., Maloney, M., & Gorman, A. (2022). A Narrative Review of the Current State of Extended Reality Technology and How it can be Utilised in Sport. *Sports Medicine*. https://doi.org/10.1007/S40279-022-01669-0

Lee, C., & Lee, G. (2019). Technology opportunity analysis based on recombinant search: patent landscape analysis for idea generation. *Scientometrics*, *121*(2), 603–632. https://doi.org/10.1007/s11192-019-03224-7

Lee, S. (2013). Linking technology roadmapping to patent analysis. In *Technology Roadmapping for Strategy and Innovation: Charting the Route to Success* (pp. 267–284). https://doi.org/10.1007/978-3-642-33923-3_17

Lee, S., Yoon, B., Lee, C., & Park, J. (2009). Business planning based on technological capabilities: Patent analysis for technology-driven roadmapping. *Technological Forecasting and Social Change*, *76*(6), 769–786. https://doi.org/10.1016/J.TECHFORE.2009.01.003

Lee, Y., Kim, S. Y., Song, I., Park, Y., & Shin, J. (2014). Technology opportunity identification customized to the technological capability of SMEs through two-stage patent analysis. *Scientometrics*, *100*(1), 227–244. https://doi.org/10.1007/S11192-013-1216-0

Leveaux, R. (2010). Facilitating Referee's Decision Making in Sport via the Application of Technology. *Communications of the IBIMA*, *2010*, 1–8. https://doi.org/10.5171/2010.545333

Li, R. (2017). Good luck have fun: The rise of eSports. Retrieved from https://books.google.com/books?hl=nl&lr=&id=gF2CDwAAQBAJ&oi=fnd&pg=PT8&dq=+Li,+R.+(2016).+Good+luck+have+fun:+The+rise+of+eSports.+New+York,+NY:+Skyhorse+Publish&ots=pu_qnI3x4r&sig=T5QPHMGa8ypYb4SfyzrJq-Wp5o0

Li, X., & Lei, L. (2021). A bibliometric analysis of topic modelling studies (2000–2017). *Journal of Information Science*, *47*(2), 161–175. https://doi.org/10.1177/0165551519877049

Liao, L., Le, F., & Zhu, Y. (2017). The application of LDA model in patent text classification. *Journal of Modern Information*, *37*(3), 35–39.

Liu, C., Feng, Y., Lin, D., Wu, L., & Guo, M. (2020). Iot based laundry services: an application of big data analytics, intelligent logistics management, and machine learning techniques. *International Journal of Production Research*, *58*(17), 5113–5131. https://doi.org/10.1080/00207543.2019.1677961

Liu, J., Li, J., Li, W., & Wu, J. (2016, May 1). Rethinking big data: A review on the data quality and usage issues. *ISPRS Journal of Photogrammetry and Remote Sensing*. Elsevier. https://doi.org/10.1016/j.isprsjprs.2015.11.006

Lovelace, R., Birkin, M., Cross, P., & Clarke, M. (2016). From Big Noise to Big Data: Toward the Verification of Large Data sets for Understanding Regional Retail Flows. *Geographical Analysis*, *48*(1), 59–81. https://doi.org/10.1111/GEAN.12081

Lu, Y., Mei, Q., & Zhai, C. X. (2011). Investigating task performance of probabilistic topic models: An empirical study of PLSA and LDA. *Information Retrieval*, *14*(2), 178–203. https://doi.org/10.1007/s10791-010-9141-9

Luscombe, A., Dick, K., & Walby, K. (2022). Algorithmic thinking in the public interest: navigating technical, legal, and ethical hurdles to web scraping in the social sciences. *Quality and Quantity*, *56*(3), 1023–1044. https://doi.org/10.1007/s11135-021-01164-0

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., … Adam, S. (2018). Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology. *Communication Methods and Measures*, *12*(3), 93–118. https://doi.org/10.1080/19312458.2018.1430754

Mann, D. (2004). Hands - On Systematic Innovation Hands - On Systematic Innovation, 1–52. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.514.3435

Mansfield, E. (1986). Patents and Innovation: An Empirical Study. *Management Science*, *32*(2), 173–181. https://doi.org/10.1287/mnsc.32.2.173

Mayne, R. S., Bleakley, C. M., & Matthews, M. (2021). Use of monitoring technology and injury incidence among recreational runners: a cross-sectional study. *BMC Sports Science, Medicine and Rehabilitation*, *13*(1), 1–7. https://doi.org/10.1186/s13102-021-00347-4

Meire, M., Ballings, M., & Van den Poel, D. (2017). The added value of social media data in B2B customer acquisition systems: A real-life experiment. *Decision Support Systems*, *104*, 26–37. https://doi.org/10.1016/j.dss.2017.09.010

Mendes, J. J. A., Vieira, M. E. M., Pires, M. B., & Stevan, S. L. (2016). Sensor Fusion and Smart Sensor in Sports and Biomedical Applications. *Sensors 2016, Vol. 16, Page 1569*, *16*(10), 1569. https://doi.org/10.3390/S16101569

Ménière, Y., & Kendrick, A. (2021). EPO - Expert analysis: patenting trends in 2020. Retrieved August 7, 2022, from https://www.epo.org/about-us/annual-reports-statistics/statistics/2020/patenting-trends.html

Menolotto, M., Komaris, D. S., Tedesco, S., O'flynn, B., & Walsh, M. (2020). Motion capture technology in industrial applications: A systematic review. *Sensors (Switzerland)*, *20*(19), 1–25. https://doi.org/10.3390/S20195687

Michalopoulos, N., Raptis, G. E., & Katsini, C. (2019). Using Video-based Technology in Powerlifting Sport to Support Referees' Decision Making. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3290607.3312991

Michalski, S. C., Szpak, A., & Loetscher, T. (2019). Using virtual environments to improve real-world motor skills in sports: A systematic review. *Frontiers in Psychology*, *10*(SEP), 2159. https://doi.org/10.3389/FPSYG.2019.02159/BIBTEX

Milshina, Y., & Vishnevskiy, K. (2019). Roadmapping in fast changing environments – the case of the Russian media industry. *Journal of Engineering and Technology Management*, *52*, 32–47. https://doi.org/10.1016/J.JENGTECMAN.2017.10.004

Moehrle, M. G., & Geritz, A. (2007). Developing acquisition strategies based on patent maps. *Management of Technology : New Directions in Technology Management : Selected Papers from the Thirteenth International Conference on Management of Technology.*

Mooney, R., Corley, G., Godfrey, A., Osborough, C., Quinlan, L. R., & ÓLaighin, G. (2015). Application of Video-Based Methods for Competitive Swimming Analysis: A Systematic Review. *Sports and Exercise Medicine - Open Journal*, *1*(5), 133–150. https://doi.org/10.17140/semoj-1-121

Morkhat, P. M. (2018). Artificial intelligence in sports: Potentialities and implementation ways and tools. *Teoriya i Praktika Fizicheskoy Kultury*, (10), 95–97.

Moro, S., Cortez, P., & Rita, P. (2015). Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation. *Expert Systems with Applications*, *42*(3), 1314–1324. https://doi.org/10.1016/J.ESWA.2014.09.024

Moro, S., Pires, G., Rita, P., & Cortez, P. (2019). A text mining and topic modelling perspective of ethnic marketing research. *Journal of*

Business Research, *103*, 275–285. https://doi.org/10.1016/j.jbusres.2019.01.053

Muniz-Pardos, B., Sutehall, S., Angeloudis, K., Guppy, F. M., Bosch, A., & Pitsiladis, Y. (2021). Recent Improvements in Marathon Run Times Are Likely Technological, Not Physiological. *Sports Medicine*, *51*(3), 371–378. https://doi.org/10.1007/S40279-020-01420-7

Neumann, D. L., Moffitt, R. L., Thomas, P. R., Loveday, K., Watling, D. P., Lombard, C. L., … Tremeer, M. A. (2018). A systematic review of the application of interactive virtual reality to sport. *Virtual Reality*, *22*(3), 183–198. https://doi.org/10.1007/s10055-017-0320-5

Newbert, S., Walsh, S., Kirchhoff, B., & Chavez, V. (2006). Technology-driven entrepreneurship: Muddling through and succeeding with the second product. *Entrepreneurship: The Engine of Growth*, *3*, 291–312.

Nlandu, T. (2012). The fallacies of the assumptions behind the arguments for goal-line technology in soccer. *Sport, Ethics and Philosophy*, *6*(4), 451–466. https://doi.org/10.1080/17511321.2012.745896

Noh, H., Jo, Y., & Lee, S. (2015). Keyword selection and processing strategy for applying text mining to patent analysis. *Expert Systems with Applications*, *42*(9), 4348–4360. https://doi.org/10.1016/J.ESWA.2015.01.050

Noruzi, A., & Abdekhoda, M. (2014). Google patents: The global patent search engine. *Webology*, *11*(1). Retrieved from http://www.webology.org/2014/v11n1/a122.pdf

Nti, I. K., Quarcoo, J. A., Aning, J., & Fosu, G. K. (2022, June 1). A mini-review of machine learning in big data analytics: Applications, challenges, and prospects. *Big Data Mining and Analytics*. Tsinghua University Press. https://doi.org/10.26599/BDMA.2021.9020028

O'toole, J. (2021). Questioning Authority: Patents and Source Evaluation in an Era of Misinformation. *Journal of the Patent and Trademark Resource Center*, *31*, 3.

Ofoghi, B., Zeleznikow, J., MacMahon, C., & Raab, M. (2013, July 1). Data mining in elite sports: A review and a framework. *Measurement in Physical Education and Exercise Science*. Routledge . https://doi.org/10.1080/1091367X.2013.805137

Orsenigo, L., & Sterzi, V. (2010). Comparative Study of the Use of Patents in Different Industries. *KITeS*, *33*. Retrieved from www.kites.unibocconi.it

Ozcan, S., Homayounfard, A., Simms, C., & Wasim, J. (2022). Technology Roadmapping Using Text Mining: A Foresight Study for the Retail Industry. *IEEE Transactions on Engineering Management*, *69*(1), 228–244. https://doi.org/10.1109/TEM.2021.3068310

Paap, J. (2020). Mapping the technological landscape to accelerate innovation. *Foresight and STI Governance*, *14*(3), 41–54. https://doi.org/10.17323/2500-2597.2020.3.41.54

Pagé, C., Bernier, P. M., & Trempe, M. (2019). Using video simulations and virtual reality to improve decision-making skills in basketball. *Https://Doi.Org/10.1080/02640414.2019.1638193*, *37*(21), 2403–2410. https://doi.org/10.1080/02640414.2019.1638193

Park, J. S., Kim, N. R., Choi, H. R., & Han, E. (2018). A new forecasting system using the latent dirichlet allocation (LDA) topic modeling technique. *WSEAS Transactions on Environment and Development*, *14*, 363–373.

Passarelli, M., Cariola, A., & Vecellio, P. (2018). Beyond multidirectional technology transfer: The case of a 'proof-of-concept network.' *Industry and Higher Education*, *32*(5), 312–325. https://doi.org/10.1177/0950422218790538

Passos, J., Lopes, S. I., Clemente, F. M., Moreira, P. M., Rico-González, M., Bezerra, P., & Rodrigues, L. P. (2021). Wearables and Internet of Things (IoT) Technologies for Fitness Assessment: A Systematic Review. *Sensors 2021*, *21*(16), 5418. https://doi.org/10.3390/S21165418

Pek, Y. N., & Lim, K. H. (2019). Identifying and Understanding Business Trends using Topic Models with Word Embedding. In *Proceedings - 2019 IEEE International Conference on Big Data, Big Data 2019* (pp. 6177–6179). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/BigData47090.2019.9005497

Ploessl, F., Just, T., & Wehrheim, L. (2021). Cyclicity of real estate-related trends: topic modelling and sentiment analysis on German real estate news. *Journal of European Real Estate Research*, *14*(3), 381–400. https://doi.org/10.1108/JERER-12-2020-0059/FULL/XML

Pueo, B., & Jimenez-Olmedo, J. M. (2017). Application of motion capture technology for sport performance analysis. *Retos*, (32), 241–247. Retrieved from https://dialnet.unirioja.es/servlet/articulo?codigo=6352311

Purwar, S., & Daim, T. U. (2021). Technology Roadmap: Smartwatches (pp. 209–223). Springer, Cham. https://doi.org/10.1007/978-3-030-50502-8_5

Qi, Y., Zhu, N., Zhai, Y., & Ding, Y. (2018). The mutually beneficial relationship of patents and scientific literature: topic evolution in nanoscience. *Scientometrics*, *115*(2), 893–911. https://doi.org/10.1007/S11192-018-2693-Y

Raj, R., & Mishra, A. (2020). Machine Learning for Big Data Analytics, Interactive and Reinforcement. In *Artificial Intelligence Trends for Data Analytics Using Machine Learning and Deep Learning Approaches* (pp. 215–232). CRC Press. https://doi.org/10.1201/9780367854737-13

Rajšp, A., & Fister, I. (2020). A systematic literature review of intelligent data analysis methods for smart sport training. *Applied Sciences (Switzerland)*. https://doi.org/10.3390/app10093013

Rao, P., Seshadri, D. R., & Hsu, J. J. (2021). Current and Potential Applications of Wearables in Sports Cardiology. *Current Treatment Options in Cardiovascular Medicine*, *23*(10), 1–15. https://doi.org/10.1007/S11936-021-00942-1/FIGURES/3

Ratten, V. (2017). Sports innovation management. *Sports Innovation Management*, 1–127. https://doi.org/10.4324/9781315177977/SPORTS-INNOVATION-MANAGEMENT-VANESSA-RATTEN

Rauter, G., Sigrist, R., Koch, C., Crivelli, F., Van Raai, M., Riener, R., & Wolf, P. (2013). Transfer of complex skill learning from virtual to real rowing. *PLoS ONE*, *8*(12). https://doi.org/10.1371/JOURNAL.PONE.0082145

Rein, R., & Memmert, D. (2016, December 1). Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science. *SpringerPlus*. SpringerOpen. https://doi.org/10.1186/s40064-016-3108-2

Rejeb, A., Rejeb, K., & Keogh, J. G. (2020). Potential of Big Data for Marketing: A Literature Review. *Management Research and Practice*, *12*(3), 60–73. Retrieved from https://search.proquest.com/scholarly-journals/potential-big-data-marketing-literature-review/docview/2447004745/se-2?accountid=38809

Repanovici, R. M. (2018). Multi-criteria analysis of wearable activity-fitness trackers and 3D concept models of smart jewellery. In *MATEC Web of Conferences* (Vol. 178). https://doi.org/10.1051/matecconf/201817805017

Rieger, J., Jentsch, C., & Rahnenführer, J. (2020). Assessing the Uncertainty of the Text Generating Process Using Topic Models. *Communications in Computer and Information Science*, *1323*, 385–396. https://doi.org/10.1007/978-3-030-65965-3_26/FIGURES/6

Ringuet-Riot, C. J., Hahn, A., & James, D. A. (2013). A structured approach for technology innovation in sport. *Sports Technology*, *6*(3), 137–149. https://doi.org/10.1080/19346182.2013.868468

Robertshaw, H. (2021). Zwift: Everything you need to know about the online training and racing platform. Retrieved March 18, 2022, from https://www.cyclingweekly.com/news/product-news/zwift-turbo-trainer-game-171798

Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. In *WSDM 2015 - Proceedings of the 8th ACM International Conference on Web Search and Data Mining* (pp. 399–408). https://doi.org/10.1145/2684822.2685324

Rodriguez, H., Haag, M., Abner, S., Johnson, W., Glassel, A., Musselman, R., & Wyatt, R. (2016). *OpTic gaming: The making of eSports champions*. HarperCollins.

Roth, J. (2003). Enabling knowledge creation: Learning from an R&D organization. *Journal of Knowledge Management*, *7*(1), 32–48. https://doi.org/10.1108/13673270310463608

Rubilar-Torrealba, R., Chahuán-Jiménez, K., & de la Fuente-Mella, H. (2022). Analysis of the Growth in the Number of Patents Granted and Its Effect over the Level of Growth of the Countries: An Econometric Estimation of the Mixed Model Approach. *Sustainability (Switzerland)*, *14*(4), 2384. https://doi.org/10.3390/su14042384

Saidani, S., Haddad, R., Mezghani, N., & Bouallegue, R. (2018). A survey on smart shoe insole systems. In *2018 International Conference on Smart Applications, Communications and Networking, SmartNets 2018*. https://doi.org/10.1109/SMARTNETS.2018.8707391

Salpavaara, T., Verho, J., Lekkala, J., & Halttunen, J. (2009). Wireless insole sensor system for plantar force measurements during sport events. In *19th IMEKO World Congress 2009* (Vol. 2, pp. 835–840).

Saura, J. R., Ribeiro-Soriano, D., & Palacios-Marqués, D. (2021). Setting B2B digital marketing in artificial intelligence-based CRMs: A review and directions for future research. *Industrial Marketing Management*, *98*, 161–178. https://doi.org/10.1016/J.INDMARMAN.2021.08.006

Scataglini, S., Moorhead, A. P., & Feletti, F. (2020). A systematic review of smart clothing in sports: Possible applications to extreme sports. *Muscles, Ligaments and Tendons Journal*, *10*(2), 333–342. https://doi.org/10.32098/mltj.02.2020.19

Schumaker, R. P., & Chen, H. (2006). Textual analysis of stock market prediction using financial news articles. In *Association for Information Systems - 12th Americas Conference On Information Systems, AMCIS 2006* (Vol. 3, pp. 1422–1430). AMCIS.

Seshadri, D. R., Thom, M. L., Harlow, E. R., Gabbett, T. J., Geletka, B. J., Hsu, J. J., … Voos, J. E. (2021). Wearable Technology and Analytics as a Complementary Toolkit to Optimize Workload and to Reduce Injury Burden. *Frontiers in Sports and Active Living*, *2*, 228. https://doi.org/10.3389/FSPOR.2020.630576/BIBTEX

Shanbhag, A. (2021). Trend forecast for 2022: How will Video Analytics technology change the industry? Retrieved July 19, 2022, from https://www.cioandleader.com/article/2021/11/24/trend-forecast-2022-how-will-video-analytics-technology-change-industry#1

Shibata, N., Kajikawa, Y., Takeda, Y., & Matsushima, K. (2008). Detecting emerging research fronts based on topological measures in citation networks of scientific publications. *Technovation*, *28*(11), 758–775. https://doi.org/10.1016/j.technovation.2008.03.009

Siepmann, C., & Kowalczuk, P. (2021). Understanding continued smartwatch usage: the role of emotional as well as health and fitness factors. *Electronic Markets*, *31*(4), 795–809. https://doi.org/10.1007/S12525-021-00458-3/FIGURES/2

Sievert, C., & Shirley, K. E. (2014). LDAvis: A method for visualizing and interpreting topics, 63–70.

Song, C., Yoon, J., Ko, N., & Han, J.-W. (2016). Mapping the Technological Knowledge Landscape: The Case of Epigenetics. *Recent Patents on Anti-Cancer Drug Discovery*, *11*(4), 424–433. https://doi.org/10.2174/1574892811666160825143006

Spitz, J., Wagemans, J., Memmert, D., Mark Williams, A., & Helsen, W. F. (2020). Video assistant referees (VAR): The impact of technology on decision making in association football referees. https://doi.org/10.1080/02640414.2020.1809163

Steiber, A., & Alänge, S. (2020). Corporate-startup Co-creation for Increased Innovation and Societal Change. *Triple Helix*, *7*(2–3), 227–249. https://doi.org/10.1163/21971927-BJA10004

Stein, M., Janetzko, H., Lamprecht, A., Breitkreutz, T., Zimmermann, P., Goldlücke, B., … Keim, D. A. (2018). Bring It to the Pitch: Combining Video and Movement Data to Enhance Team Sport Analysis. *IEEE Transactions on Visualization and Computer Graphics*, *24*(1), 13–22. https://doi.org/10.1109/TVCG.2017.2745181

Sun, Y., Deng, H., & Han, J. (2012). Probabilistic models for text mining. In *Mining Text Data* (Vol. 9781461432, pp. 259–295). https://doi.org/10.1007/978-1-4614-3223-4_8

Sung, W., & Yeo, H. (2019). Smart city research trend analysis: Text mining, topic analysis, network analysis. *Journal of Advanced Research in Dynamical and Control Systems*, *11*(7 Special Issue), 1302–1309.

Tahan, O., Rady, M., Sleiman, N., Ghantous, M., & Merhi, Z. (2018). A computer vision driven squash players tracking system. In *19th IEEE Mediterranean Eletrotechnical Conference, MELECON 2018 - Proceedings* (pp. 155–159). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/MELCON.2018.8379085

Taques, F. H., López, M. G., Basso, L. F., & Areal, N. (2021). Indicators used to measure service innovation and manufacturing innovation. *Journal of Innovation & Knowledge*, *6*(1), 11–26. https://doi.org/10.1016/J.JIK.2019.12.001

Teece, D. J., Pisano, G., & Shuen, A. (2009). Dynamic capabilities and strategic management. In *Knowledge and Strategy* (Vol. 18, pp. 77–116). https://doi.org/10.1093/0199248540.003.0013

Thiel, A., & John, J. M. (2018, October 2). Is eSport a ‘real’ sport? Reflections on the spread of virtual competitions. *European Journal for Sport and Society*. Routledge. https://doi.org/10.1080/16138171.2018.1559019

Thompson, W. R. (2022). Worldwide Survey of Fitness Trends for 2022. *ACSM’S Health & Fitness Journal*, *26*(1), 11–20. https://doi.org/10.1249/fit.0000000000000732

Timokhina, G. (2021). Improving the Company Competitiveness through the Development of Key Competencies: Marketing - Management Technologies and Practices. *Revista Gestão Inovação e Tecnologias*, *11*(4), 917–934. https://doi.org/10.47059/revistageintec.v11i4.2157

Tirp, J., Steingröver, C., Wattie, N., Baker, J., & Schorer, J. (2015). Virtual realities as optimal learning environments in sport-A transfer study of virtual and real dart throwing. *Psychologie-Aktuell.Com*, *57*(1), 57–69. Retrieved from https://www.psychologie-aktuell.com/fileadmin/download/ptam/1-2015_20150327/04_Tirp.pdf

Toro-Jarrín, M. A., Ponce-Jaramillo, I. E., & Güemes-Castorena, D. (2016). Methodology for the of building process integration of Business Model Canvas and Technological Roadmap. *Technological Forecasting and Social Change*, *110*, 213–225. https://doi.org/10.1016/j.techfore.2016.01.009

Toth, A. J., Conroy, E., & Campbell, M. J. (2021). Beyond action video games: Differences in gameplay and ability preferences among gaming genres. *Entertainment Computing*, *38*, 100408. https://doi.org/10.1016/J.ENTCOM.2021.100408

Umek, A., Zhang, Y., Tomažič, S., Kos, A., Sun, Y., Cai, Z., & Jara, A. (2017). Suitability of Strain Gage Sensors for Integration into Smart Sport Equipment: A Golf Club Example. *Sensors 2017, Vol. 17, Page 916*, *17*(4), 916. https://doi.org/10.3390/S17040916

Un, C. A., & Cuervo-Cazurra, A. (2004). Strategies for Knowledge Creation in Firms. *British Journal of Management*, *15*(S1), 27–41. https://doi.org/10.1111/j.1467-8551.2004.00398.x

Van Hilvoorde, I., & Pot, N. (2016). Sport, Ethics and Philosophy Embodiment and fundamental motor skills in eSports Ivo van Hilvoorde & Niek Pot Embodiment and fundamental motor skills in eSports. *EthicS and PhiloSophy*, *10*(1), 14–27. https://doi.org/10.1080/17511321.2016.1159246

van Weel, K. (2022). Meetschaats geeft inzicht in schaatstechniek. Retrieved August 4, 2022, from https://www.universiteitleiden.nl/nieuws/2022/01/meetschaats-geeft-inzicht-in-schaatstechniek

Vijayarani, S., Ilamathi, M. ., & Nithya, M. (2015). Preprocessing Techniques for Text Mining-An Overview Privacy Preserving Data Mining View project. *International Journal of Computer Science & Communication Networks*, *5*(1), 7–16. Retrieved from https://www.researchgate.net/publication/339529230

Walkowiak, T., Datko, S., & Maciejewski, H. (2019). Bag-of-words, bag-of-topics and word-to-vec based subject classification of text documents in Polish - A comparative study. In *Advances in Intelligent Systems and Computing* (Vol. 761, pp. 526–535). Springer Verlag. https://doi.org/10.1007/978-3-319-91446-6_49

Wang, K. J., & Lestari, Y. D. (2013). Firm competencies on market entry success: Evidence from a high-tech industry in an emerging market. *Journal of Business Research*, *66*(12), 2444–2450. https://doi.org/10.1016/J.JBUSRES.2013.05.033

Weber, K. M., Harper, J. C., Könnölä, T., & Barceló, V. C. (2012). Coping with a fast-changing world: Towards new systems of future-oriented technology analysis. *Science and Public Policy*, *39*(2), 153–165. https://doi.org/10.1093/SCIPOL/SCS012

Westmattelmann, Daniel Grotenhermen, J.-G., Stoffers, B., & Schewe, G. (2021). Exploring the adoption of mixed-reality sport platforms: A qualitative study on ZWIFT. In *Exploring the adoption of mixed-reality sport platforms: A qualitative study on ZWIFT* (pp. 1–18). Retrieved from https://aisel.aisnet.org/ecis2021_rp/48

Winand, M., & Fergusson, C. (2016). More decision-aid technology in sport? An analysis of football supporters’ perceptions on goal-line technology. *Https://Doi-Org.Ezproxy2.Utwente.Nl/10.1080/14660970.2016.1267629*, *19*(7), 966–985. https://doi.org/10.1080/14660970.2016.1267629

Wong, D., & Meng-Lewis, Y. (2022). Esports: an exploration of the advancing esports landscape, actors and interorganisational relationships. *Https://Doi-Org.Ezproxy2.Utwente.Nl/10.1080/17430437.2022.2086458*, 1–27. https://doi.org/10.1080/17430437.2022.2086458

Xu, G., Dong, F., & Feng, J. (2020). Mapping the Technological Landscape of Emerging Industry Value Chain Through a Patent Lens: An Integrated Framework With Deep Learning. *IEEE Transactions on Engineering Management*. https://doi.org/10.1109/TEM.2020.3041933

Yafooz, W. M. S., Bakar, Z. B. A., Fahad, S. K. A., & Mithun, A. M. (2020). Business Intelligence Through Big Data Analytics, Data Mining and Machine Learning. *Advances in Intelligent Systems and Computing*, *1016*, 217–230. https://doi.org/10.1007/978-981-13-9364-8_36

Yeap, T., Loo, G. H., & Pang, S. (2003). Computational patent mapping: Intelligent agents for nanotechnology. In *Proceedings - International Conference on MEMS, NANO and Smart Systems, ICMENS 2003* (pp. 274–278). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICMENS.2003.1222008

Yoon, J., & Kim, K. (2011). Identifying rapidly evolving technological trends for R&D planning using SAO-based semantic patent networks. *Scientometrics*, *88*(1), 213–228. https://doi.org/10.1007/s11192-011-0383-0

Zglinski, J. (2022). Rules, Standards, and the Video Assistant Referee in Football. *Sport, Ethics and Philosophy*, *16*(1), 3–19. https://doi.org/10.1080/17511321.2020.1857823

Zhang, H., Daim, T., & Zhang, Y. (Peggy). (2021). Integrating patent analysis into technology roadmapping: A latent dirichlet allocation based technology assessment and roadmapping in the field of Blockchain. *Technological Forecasting and Social Change*, *167*, 120729. https://doi.org/10.1016/j.techfore.2021.120729

Zhang, R. (2022). Research on intelligent analysis technology of football video based on computer vision, 84. https://doi.org/10.1117/12.2636761

Zimmer, R. T., Haupt, S., Heidenreich, H., & Schmidt, W. F. J. (2022). Acute Effects of Esports on the Cardiovascular System and Energy Expenditure in Amateur Esports Players. *Frontiers in Sports and Active Living*, *4*. https://doi.org/10.3389/fspor.2022.824006

Appendix

A: Scripts used for extracting the data

main.py

```
import sys
sys.path.append("../")
import pandas as pd
from patentClass import *
from wordcloud import WordCloud
import gensim
from nltk import WordNetLemmatizer
import spacy
import gensim.corpora as corpora
from pprint import pprint

if __name__ == '__main__':

    data = pd.read_csv(r'C:\Users\Path\To\CSVFile.csv',sep=',',header=1)
    # Add dataframe with abstracts and titles
    data['Text'] = getAbstracts(data)
    # Save to csv file for faster loading later on
    data.to_csv('SportTechUSEU3y.csv')
    #   Remove files with no text
    data = data[data.Text != 'no abstract found']
    data.reset_index(drop=True,inplace=True)
    #   Clean data
    data['Processed_Text'] = data['Text'].apply(lambda x: cleanText(x,stem='None'))
    print(data['Processed_Text'][2])

    data_words = data['Processed_Text'].values.tolist()
    # higher threshold fewer phrases.
    bigram = gensim.models.Phrases(data_words, min_count=5, threshold=100)
    trigram = gensim.models.Phrases(bigram[data_words], threshold=100)


    # Remove Stop Words
    data_words_nostops = remove_stopwords(data_words)

    # Form Bigrams
    data_words_bigrams = make_bigrams(data_words_nostops,bigram=bigram,trigram=trigram)

    # Initialize spacy 'en' model, keeping only tagger component (for efficiency)
    # Do lemmatization keeping only noun, adj, vb, adv
    nlp = spacy.load("en_core_web_sm", disable=['parser', 'ner'])
    data_lemmatized  =  lemmatization(data_words_bigrams,  allowed_postags=['NOUN',  'ADJ',
'VERB', 'ADV'])
    # Create WordCloud to see how we're doing
    abstracts = data_lemmatized
    allText = "".join(str(abstracts))
    allText = allText.replace('\'',' ')
    wordcloud = WordCloud(background_color="white").generate(allText)
    wordcloud.to_file("firstCloud.png")

    # Create inputs for LDA model (dictionary and corpus)
    id2word = corpora.Dictionary(data_lemmatized)
    texts = data_lemmatized
    corpus = [id2word.doc2bow(text) for text in texts]# View
    print(corpus[:1])
    # Build LDA model
    lda_model = gensim.models.LdaMulticore(corpus=corpus,
                                       id2word=id2word,
                                       num_topics=5,
                                       random_state=100,
```

```
                                   chunksize=100,
                                   passes=10,
                                   per_word_topics=True)

    # Print the Keyword in the topics
    pprint(lda_model.print_topics())
    doc_lda = lda_model[corpus]

    import numpy as np
    # Calculate best number of topics from coherence score
    best_num = float('NaN')
    best_score = 0
    from gensim.models.coherencemodel import CoherenceModel
      # compute the coherence scores for each number of topics
    for i in range(2,15):

        # create lda model with i topics
        lda = gensim.models.LdaMulticore(corpus=corpus,
                                    id2word=id2word,
                                    num_topics=i,
                                    random_state=100,
                                    chunksize=100,
                                    passes=10,
                                    per_word_topics=True)

        # obtain the coherence score
        coherence_model       =       CoherenceModel(model=lda,       texts=data_lemmatized,
dictionary=id2word, coherence='c_v')
        coherence_score = np.round(coherence_model.get_coherence(),2)
        if coherence_score > best_score:
            best_num = i
            best_score = coherence_score

    print(f'The coherence score is highest ({best_score}) with {best_num} topics.')

    # Visualize results from LDA with LDAvis
    import pyLDAvis.gensim_models
    import pyLDAvis# Visualize the topics
    # pyLDAvis.enable_notebook()
    LDAvis_prepared = pyLDAvis.gensim_models.prepare(lda_model, corpus, id2word)
    pyLDAvis.save_html(LDAvis_prepared,'ldaTest.htm')

    # Assign every patent to topic
    # column names
    topicnames = ["Topic" + str(i) for i in range(lda_model.num_topics)]
    # index names
    docnames = ["Doc" + str(i) for i in range(len(data))]

    # Make the pandas dataframe
    #   df_document_topic   =   pd.DataFrame(np.round(lda_output,   2),   columns=topicnames,
index=docnames)

    def format_topics_sent(ldamodel, corpus, texts):
        sent_topics_df = pd.DataFrame()
        for i, row in enumerate(ldamodel[corpus]):
            row = sorted(row[0], key=lambda x: x[1], reverse=True)

            for j, (topic_num, prop_topic) in enumerate(row):
                if j == 0:
                    wp = ldamodel.show_topic(topic_num)
                    topic_keywords = ", ".join([word for word, prop in wp])
                    sent_topics_df     =     sent_topics_df.append(pd.Series([int(topic_num),
round(prop_topic,4), topic_keywords]), ignore_index=True)
                else:
                    break
        sent_topics_df.columns = ['Dominant_topic', 'Perc_Contrib', 'Topic_Keywords']
        contents = pd.Series(texts)
```

```
        sent_topics_df = pd.concat([sent_topics_df, contents], axis=1)
        sent_topics_df.rename(columns={0: "Text"}, inplace=True)
        return sent_topics_df


    sent_topics_df = format_topics_sent(lda_model, corpus, data_lemmatized)
```

patentClass.py

```
# Script containing functions to process patent data
from bs4 import BeautifulSoup
import requests as rq
import sys
import pandas as pd
import nltk
import re
from gensim.parsing.preprocessing import STOPWORDS
import string
from nltk.stem import PorterStemmer
from nltk.stem.wordnet import WordNetLemmatizer
import spacy
import gensim
from gensim.utils import simple_preprocess


def getAbstracts(data):
    """ Function to extract abstracts from Google Patents """
    text = []
    for index,row in data.iterrows():
        try:
            #   Get html data from links
            page = rq.get((row['result link'])).text
            soup = BeautifulSoup(page,features = 'html.parser')
            #   Update text to include title in abstract
            title = row['title']
            abstract = soup.select_one('[itemprop="abstract"]').text
            abstract = abstract.replace('Abstract', '')
            abstract = abstract.strip()
            alltext = [title,abstract]
            alltext = " ".join(alltext)
            if 'Translated from' in abstract:
                alltext = 'no abstract found'
            text.append(alltext)
            print(index)
        except:
            text.append('no abstract found')
    return(text)


def cleanText(text,stem='None'):
    """ Function to clean text:
        1. Remove non-English characters;
        2. Lower characters
        3. Segment sentences into words
        4. Remove stop words
        5. Remove morphological affixes from words;
        Function takes string object as an input """

    try:
        #   Remove non-English characters and numbers
        text1 = text.encode('ascii', 'ignore').decode('ascii')
        text1 = ''.join([i for i in text1 if not i.isdigit()])
        #   Lower characters
```

```
        text2 = text1.lower()
        #   Segment sentences into words by removing line breaks and punctuation
        text3 = re.sub(r'\n', '', text2)
        translator = str.maketrans('', '', string.punctuation)
        text3 = text3.translate(translator)
        text4 = text3.split()
        # print(text3)
        return(text4)
    except:
        raise Exception("Text needs to be a string")

def remove_stopwords(texts):
    """ Function to remove stopwords from text """
#   Remove stop words
    all_stopwords_gensim  =  STOPWORDS.union(set(['method',  'subject',  'system','player',
'use','sport','training','equipment','technologies','technology','invention','methods','syst
ems','players','user','users','sports','device']))
    return [[word for word in simple_preprocess(str(doc)) if word not in all_stopwords_gensim]
for doc in texts]

def make_bigrams(texts,bigram,trigram):
    """ Function to create bigrams """
    bigram_mod = gensim.models.phrases.Phraser(bigram)
    trigram_mod = gensim.models.phrases.Phraser(trigram)
    return [bigram_mod[doc] for doc in texts]

def lemmatization(texts, allowed_postags=['NOUN', 'ADJ', 'VERB', 'ADV']):
    """ Function to lemmatize text """
    # Initialize spacy 'en' model, keeping only tagger component (for efficiency)
    nlp = spacy.load("en_core_web_sm", disable=['parser', 'ner'])
    """https://spacy.io/api/annotation"""
    texts_out = []
    for sent in texts:
        doc = nlp(" ".join(sent))
        texts_out.append([token.lemma_ for token in doc if token.pos_ in allowed_postags])
    return texts_out
```

B: LDAvis results for every topic

Results from the LDA model were visualized with the LDAvis package by Sievert & Shirley (2014). This package visualizes every topic as a circle with a number, see figures 8-10 below. The number in the circle indicates the serial number of the topic and topics are arranged to the amount of associated text data. It consists of an intertopic distance map, in which each topic is plotted as a circle on a two-dimensional plane. This shows the prevalence of each topic (denoted by the size of the circle, a larger circle indicates a higher prevalence) and how the topics are related to each other. It also consists of a horizontal bar chart, in which the individual terms are shown for each topic, which can be used to derive the meaning of each topic. The bar chart consists of blue bars, which represent the frequency of the term in all documents, and a red bar, which represents the estimated frequency of the term within a topic. The topics are ranked by the relevance metric, λ, which ranges from 0 to 1. A smaller λ means that more topic-specific words will be at the top of the bar chart. Here, $\lambda = 0.6$ was used, which was found by Sievert & Shirley (2014) to give the best interpretable results. The bar charts on the right in each figure were used to construct table 3.

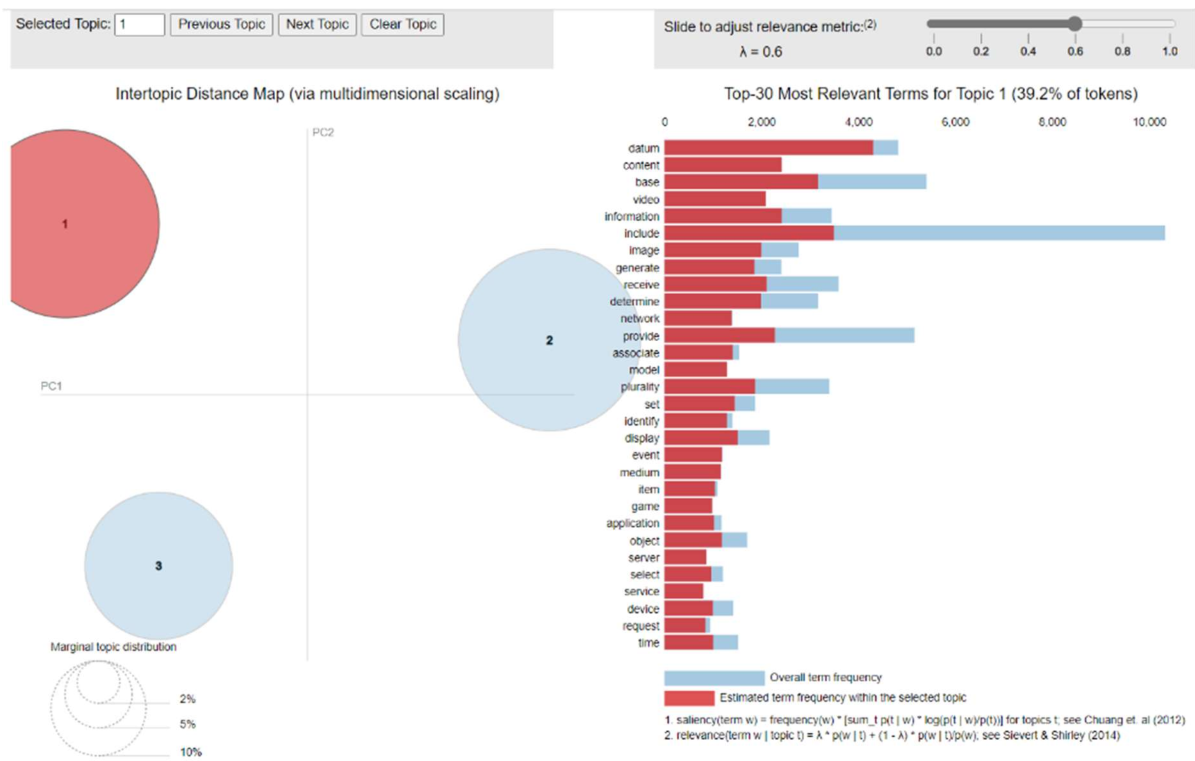

Figure 8. LDAvis results for topic 1. Topic 1 has the highest distribution. The terms ‘content’, ‘video’, ‘information’, ‘image’, ‘receive’, ‘determine’, ‘network’, ‘identify’, ‘display’, ‘medium’, ‘application’, ‘service’, and ‘device’ were used to identify topic 1, video analysis.

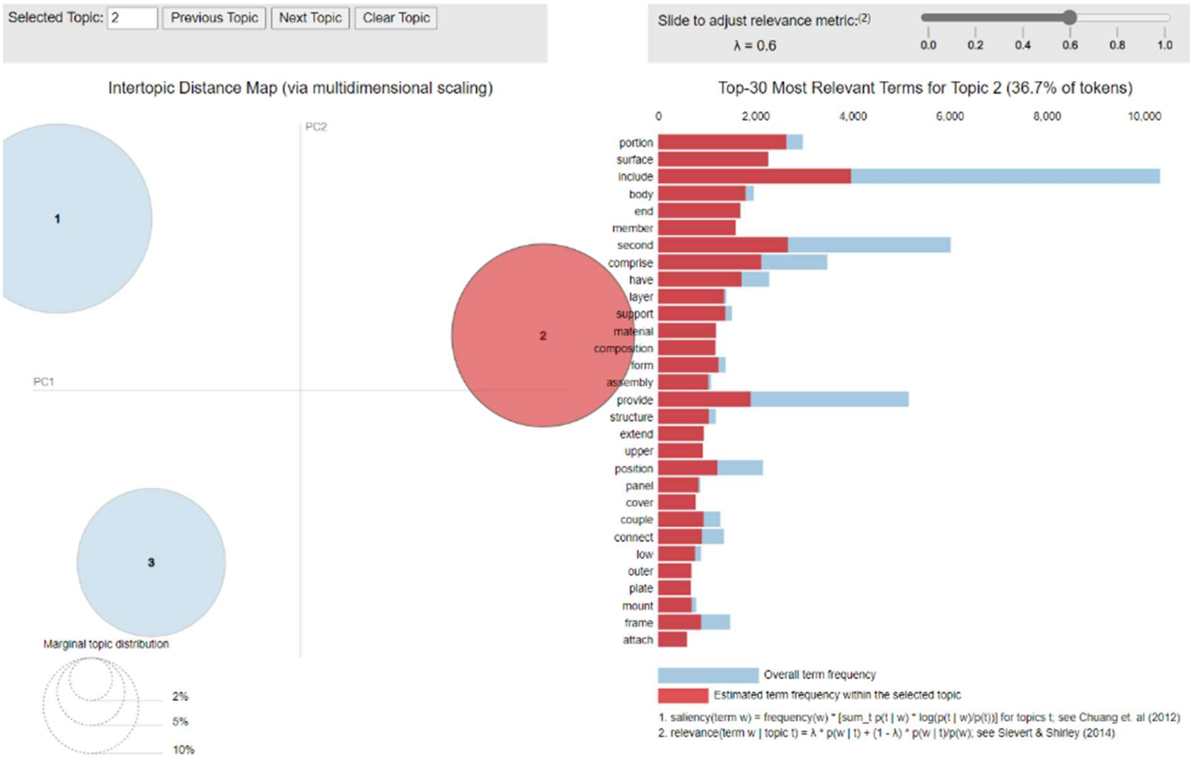

Figure 9. LDAvis results for topic 2. The terms ‘surface’, ‘body’, ‘layer’, ‘support’, ‘material’, ‘composition’, ‘form’, ‘structure’, ‘position’, ‘panel’, ‘cover’, ‘plate’, and ‘frame’ were used to identify topic 2, sports equipment.

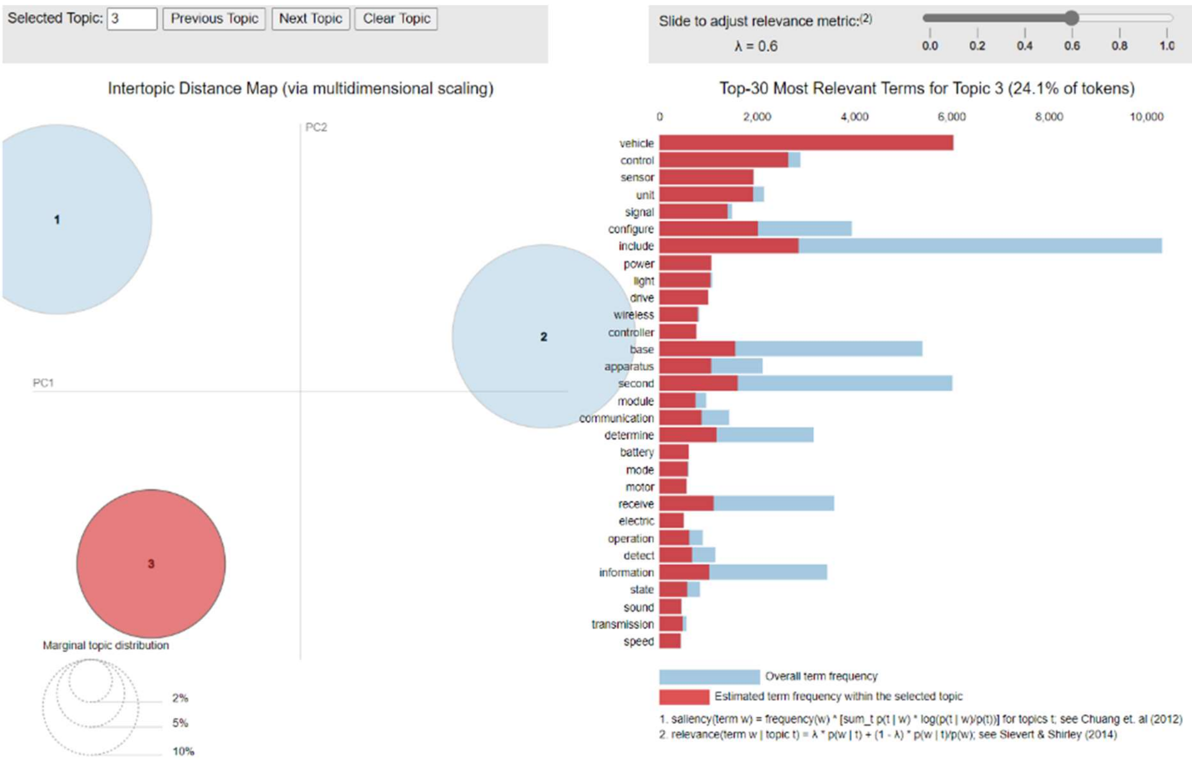

Figure 10. LDAvis results for topic 3. The terms ‘control’, ‘sensor’, ‘signal’, ‘configure’, ‘power’, ‘light’, ‘wireless’, ‘controller’, ‘apparatus’, ‘module’, ‘communication’, ‘battery’, ‘electric’, and ‘transmission’ were used to identify topic 3, sensor technology.

C: Cooperative Patent Classification (CPC)

The cooperative patent classification system is an extension of the international patent classification system, managed by the European Patent Office and the United States Patent and Trademark Office. There are 9 sections, A-H, Y, which can be further divided.

Table 5. Overview of CPC sections

Section	Meaning
A	Human necessities
B	Performing arts; transporting
C	Chemistry; metallurgy
D	Textiles; paper
E	Fixed constructions
F	Mechanical engineering; lighting; heating; weapons; blasting engines or pumps
G	Physics
H	Electricity
Y	General tagging of new technological developments

A complete overview of all CPC sections can be accessed via https://www.cooperativepatentclassification.org/cpcSchemeAndDefinitions/table

D: Patents for every topic

Table 6. Overview of patents for every topic

Topic	ID	Patent	CPC
1	US-2021264436-A1	Systems, methods, and apparatuses for implementing a consumer data aggregation platform for seamless product recall and consumer alert management	G06Q, G06F, H04L, H04W
1	US-2021326010-A1	Methods, systems, and media for navigating user interfaces	G06F
1	US-2020334711-A1	Online e-commerce and networking system with an instant payment and settlement digital currency application for realizing internet of values	G06Q
1	EP-4002147-A1	Digital content data generation systems and methods	G06F, G06N
1	US-11113744-B2	Personalized item recommendations through large-scale deep-embedding architecture with real-time inferencing	G06Q, G06F
2	US-11147903-B2	Wear-resistant joint arthroplasty implant devices	A61L, A61F
2	EP-3917349-A1	Liner for a ski boot and tongue having improved ventilation and pressure distribution on a foot	A43B
2	US-2021127857-A1	Memorabilia ball simulacrum stand and system	A47F, G09F
2	EP-3913158-A1	A load-bearing frame for a single-nave hall	E04B, E04C
2	EP-3978087-A1	An ice skate with exchangeable blade	A63C
3	EP-3749060-A1	Self-configuring lighting control	H05B
3	US-2022159163-A1	Image pickup device, solid-state image pickup element, camera module, drive control unit, and image pickup method	H05N, G02B, G03B, H04N
3	US-2022089061-A1	Controller, vehicle, and method	B60L, B60H, B60W, Y02T
3	EP-3893085-A1	Power supply device, power supply system, power supply control method, and carrier means	G06F, Y02D
3	US-2019351895-A1	INTEGRATED PROPULSION & STEERING For Battery Electric Vehicles (BEV), Hybrid Electric Vehicles (HEV), Fuel Cell Electric Vehicles (FCEV), AV (Autonomous Vehicles); Electric Trucks, Buses, and Semi-Trailers	B62D, B60W, B60K, B60L, B60Q, B60Y, F16D, G05D, Y02T