Winter 12-3-2021

# Text Mining-Based Patent Analysis of Blockchain Technology Applications

Roberto Louis Forestal

Su-Houn Liu

Shih-Ming Pi

Eldon Y. Li

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, Volume 21* (pp. 476-483). ICEB'21, Nanjing, China, December 3-7, 2021.



# Text Mining-Based Patent Analysis of Blockchain Technology Applications

Roberto Louis Forestal [1,*]
Su-Houn Liu [2]
Shih-Ming Pi [3]
Eldon Y. Li [4]

*Corresponding author
[1] Doctoral student, College of Business, Chung Yuan Christian University, Zhongli, Taoyuan, Taiwan, boolfrivie@gmail.com
[2] Professor, Department of Information Management, Chung Yuan Christian University, Zhongli, Taoyuan, Taiwan, vandy1020@gmail.com
[3] Professor, College of Business, Chung Yuan Christian University, Zhongli, Taoyuan, Taiwan, happypi@gmail.com
[4] Professor, Department of Information Management, National Chung Cheng University, Chiayi, Taiwan, miseli@ccu.edu.tw

## ABSTRACT

This study focuses on the emergence of blockchain-related technologies regarding patenting activity. Blockchain technology has gained the attention of the general public. It has intensified in recent years, making it a fascinating subject of study for a patent analysis to scrutinize the evolution of this technology. However, research using the patent landscape to study the evolution of blockchain technologies is scarce. This article follows a unique methodology and comprehensive search strategy based on patent mapping and text mining to identify and categorize Blockchain patent documents extracted from the United States Patent and Trademark Office and the World Intellectual Property Organization databases. This methodology and dataset can be used for patent landscaping exercises or bibliometric analysis.

*Keywords*: Blockchain technology, patent landscape, text mining, topic identification.

## INTRODUCTION

The idea of Bitcoin and Blockchain dates back to 2008 when a person (or group) with the pseudonym Satoshi Nakamoto proposed bitcoin as the first cryptocurrency to solve the double-spending problem without the need of a trusted authority or central server (Kakarlapudi & Mahmoud, 2021; Paulavičius, Grigaitis, Igumenov, & Filatovas, 2019). Today, hundreds of digital currencies have been created using a secure blockchain-based technique. These innovations make it possible to manage a distributed database and maintain records for digital transactions of all types, as blockchain-based transactions create quick and inexpensive public records. But, the innovative importance of blockchain and the fact that it is an open-source technology has strongly encouraged the development and use of this technology in many applications that go beyond cryptocurrencies (Abou Jaoude & Saade, 2019; Xu, Chen, & Kou, 2019). Given this enormous potential, many companies using Blockchain technologies have opted to protect their inventions by patenting their innovations. So, the last few years have seen a rush to patent Blockchain technology. This makes this hot topic a fascinating field of study for a patent analysis to know the trends and developments in a given technology (Clarke, Jürgens, & Herrero-Solana, 2020; Evangelista *et al.*, 2020).

Patents are formal agreements between all relevant parties involved in the technological innovation process. Those parties include the inventor, the government body, the assignee, and any entity designated by the government regarding information about technological innovations and their rights of use. Typically, organizations rely on intellectual property rights through patents to develop other advanced technologies (Chen, James Lin, Chang, & Liu, 2009; Kim & Bae, 2017) or to determine an efficient technology management strategy and identify technological trends in a specific field of technology (Choi & Song, 2018; Lee, Yoon, Lee, & Park, 2009; Li, Xie, Jiang, Zhou, & Huang, 2019). Hence, patent landscaping may be considered an ideal tool to identify the evolution of emerging technologies such as blockchain. This tool uses a detailed search strategy, including keywords and a patent filtering strategy. While noting the paucity of researches using patent landscaping in studying Blockchain technologies, Clarke *et al.* (2020) present a comprehensive search methodology and strategy for identifying Blockchain patent documents using a combination of keywords and specific patent classification. This query was built in cooperation with the European Patent Office (EPO). They showed that although in 2014, more than 70% of patents were related to cryptocurrencies, this percentage fell to 40% in 2017. It concludes that blockchain is slowly becoming a General-Purpose Technology. Unlike the previous work, which focuses on the EPO database, our study employs an automated patent landscape analysis coupled with text mining technology to scrutinize data coming from the United States Patent and Trademark Office (USPTO) and the World Intellectual Property Organization (WIPO) databases. In a case study, we applied the search strategy to analyze worldwide blockchain patenting from 2007 till 2020.

## MATERIALS AND METHODS

### Patent Landscape

Patent landscape analysis is a methodology using multiple sets of indicators to identify the evolution of a specific technological field (Bubela *et al.*, 2013; Carbonell, Gök, Shapira, & Faulon, 2016; Dhankhar, Gulati, Kumar, & Kapoor, 2018; Stoffels, Klauck, Hamadi, Glorius, & Leker, 2020). Although the notion of "patent landscape" may vary in scope and scale, it is

increasingly used to map trends in science and technology. In addition, text mining modeling, an artificial intelligence technology using natural language processing, can be applied in researching and clustering important technology-related thematic (Hassani, Beneki, Unger, Mazinani, & Yeganegi, 2020; Kumar, Kar, & Ilavarasan, 2021). Therefore, automated patent landscaping analysis and text mining modeling appear to be the appropriate methodology to study the evolution of blockchain technology. We employ ClusterMap, a text mining technology provided by IPTech, to exhibit the layout of technical topics.

**Search Parameters**

The starting point of the analysis is the consideration of the query list proposed by Clarke *et al.* (2020). The search of patent documents was conducted in InnoVue platform (https://www.innovue.ltd/) via its comprehensive search tool, IPTech, that gives access to patent applications filed with national and regional patent offices around the world such as the United States Patent and Trademark Office (USPTO) and World Intellectual Property Organization (WIPO). All searches were done on September 22, 2021. The search strategy used in this work is presented in Table 1.

Table 1: Search query.

| Candidate | Search string |
|---|---|
| Blockchain patents | all-US;WO-TA:(Blockchain* "Block-Chain" Bitcoin* "Bit-coin" Blocksign Codius "Colored-Coin" "Colored-Coin" Cryptocurrenc* "Crypto-Currenc" "Distributed Ledger" Dogecoin "Doge-Coin" Ethereum Factom Litecoin "Lite-Coin" "Pay-To-Script-Hash" P2SH "Proof-Of-Stake" Sidechain* "Smart-Contract" Smartcontract* Zerocash Zcash Chaincod* Counterparty XCP Digitalcurrenc* "Digital-Currenc" "Hash-Tree" Hashtree "Meta-Coin" Metacoin* "Name-Coin" Namecoin* "Proof-of-Work" "Hash-Cash" Hashcash Rootstock RSK Ripple Stellar Symbiont "Type-Coin" Typecoin* Zerocoin "Zero-Coin" Zeroknowledge "Zero-Knowledge") |

US: United States Patent and Trademark Office; WO: World Intellectual Property Organization; TA: Title and abstracts

**Data Analysis**

From the search query, 12514 patent documents whose publication date starts from January 11, 2007, to December 31, 2020, were retrieved using manual screening and duplicate removal. Documents are then organized by type of ownership and content. Therefore, the analysis is limited to the document content and the period specified in the research. Fig 1 displays the accumulated number of patents applications published between 2007 and 2020.

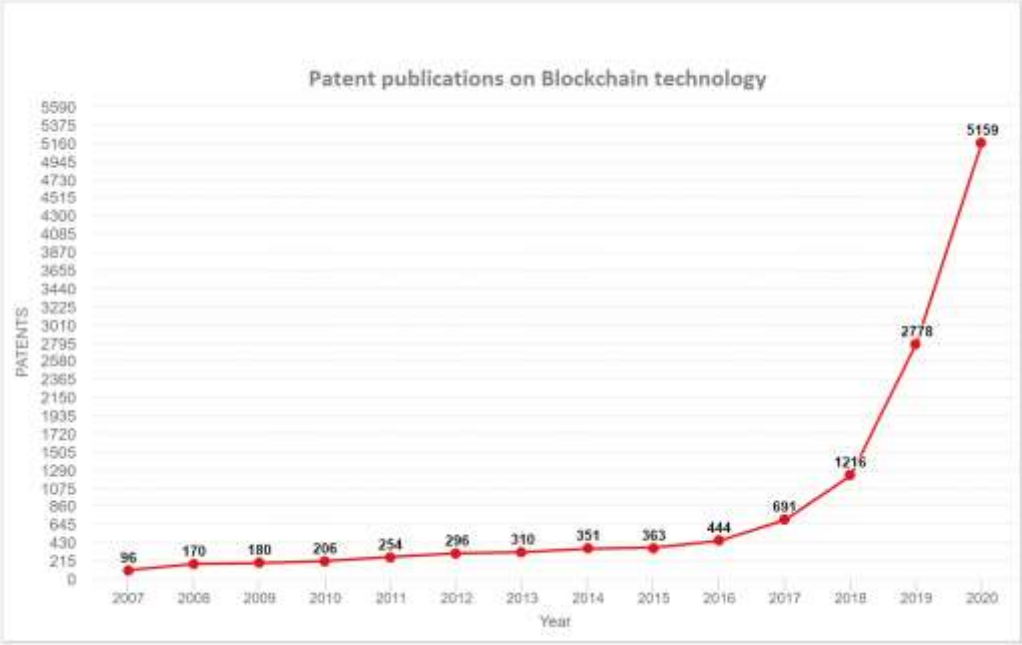


Figure 1: The trend in patent publications for blockchain technology applications.

## RESULTS

**Topic Identification**

We extract potential topic identification from the title, abstract, claims, and descriptions of all the gathered patents (Fig 2). The top 10 possible English topics are Blockchain network (2283 patents), Smart Contract (1252 patents), Computing Device (919 patents), Storage medium (483 patents), Digital Asset (281 patents), Hash tree (170 patents), Financial transaction (157

patents), Electronic document (67 patents), Machine learning (65 patents) and Virtual currency (65 patents). Blockchain networks refer to patents describing a process related to different blockchain networks. For instance, the patent application WO2020020674 introduces a communication interface to a public network and a controller to manage a transaction, while US10764259 details transaction processing for a consortium blockchain network. Smart contracts are simply programs stored on a blockchain that execute when predetermined conditions are met. An electronic document is exchanged between business partners in an electronic format. For example, the patent with the public number WO2020001103 describes a blockchain-based electronic signature method and apparatus. An electronic device receives a target transaction initiated by a subscribed user. Machine learning refers to patent documents that describe intelligent provisioning systems of a distributed ledger technology network that can be configured to receive provisioning requests from authorized users on a plurality of predefined parameters. For instance, US20200174463 presents a predictive system that involves a facility for analyzing surveillance data via a collection network. The subsequent classification includes a public key and central counterparty, as shown in Figure 3. Overall, this topic identification could be a starting point for further interpretation and in-depth analysis.

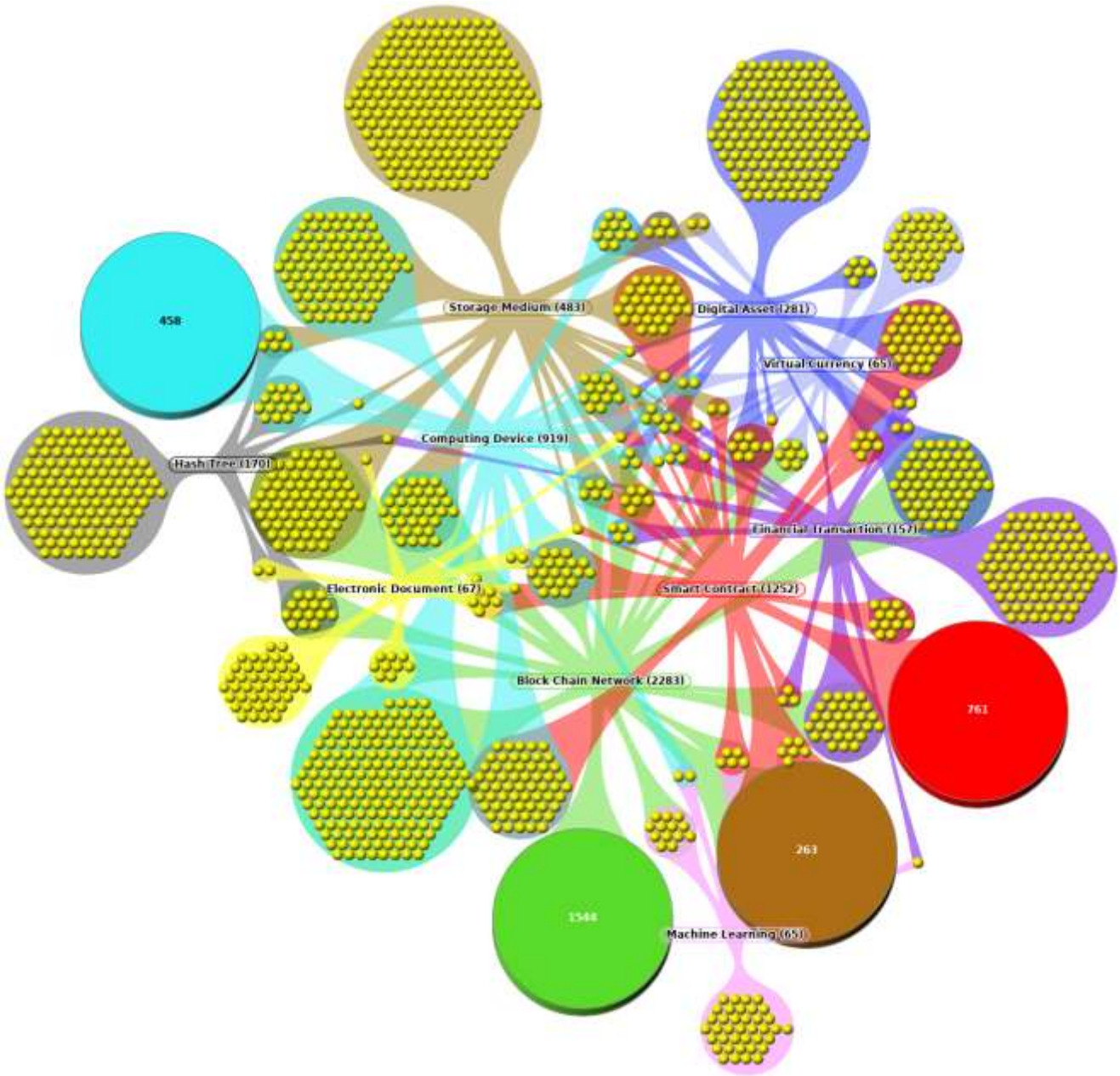


Figure 2: A network analysis of the top 10 English topics.

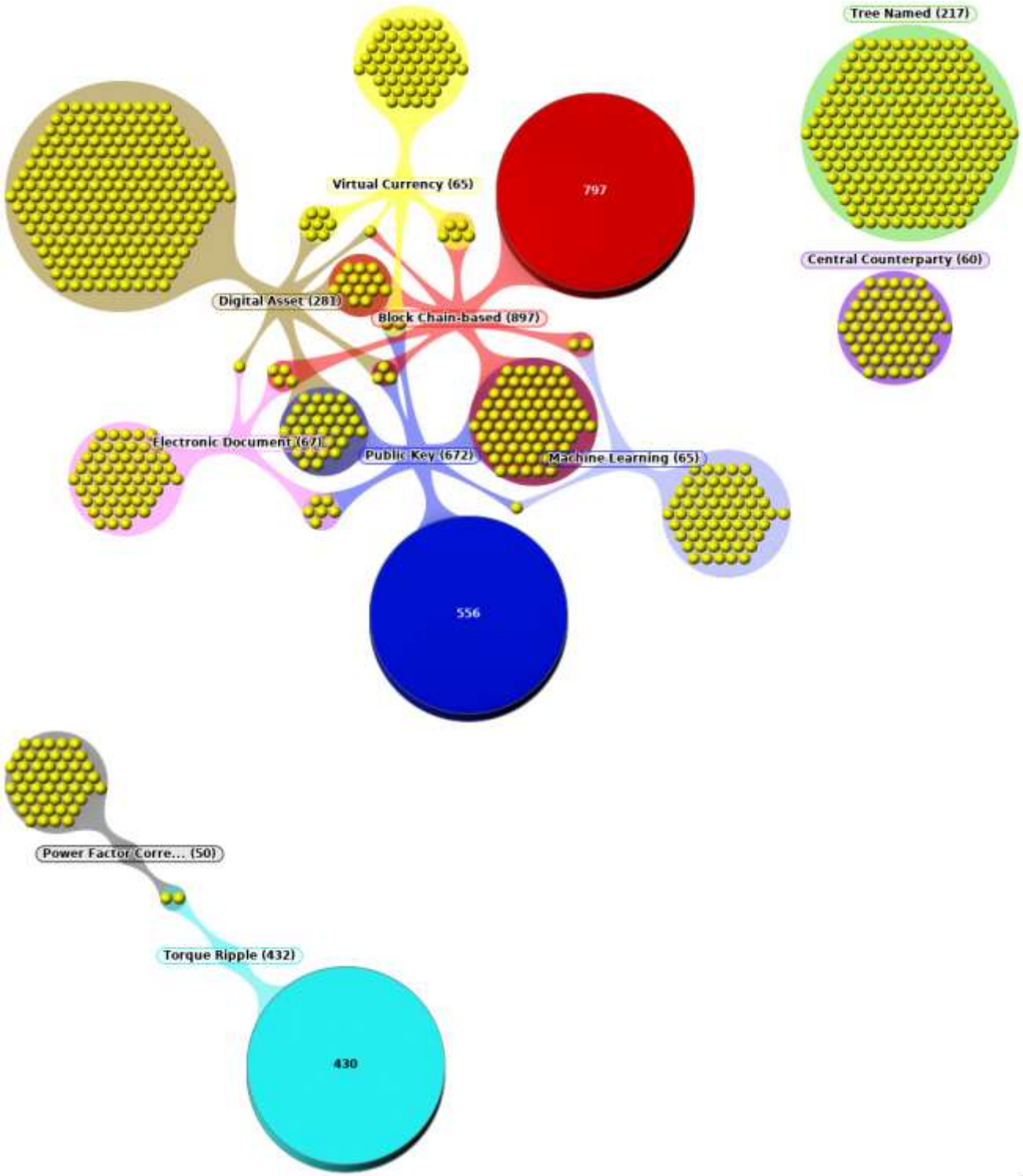


Figure 3: ClusterMap of subsequent topics.

**Classifications**

Based on data mining results and International Patent Classification (IPC) level 5 analysis, the patent documents could be organized into four groups: type of networks, IPC, crypto-assets, and industrial applications, as shown in Fig. 3. Blockchain networks involve public, private, permissioned, and consortium blockchains, while IPC includes computing devices and digital information. Crypto-assets refer to cryptocurrency and utility tokens, whereas industrial applications involve information technology (IT) services, banking, financial services, Healthcare, Media and entertainment, and Supply chain management.

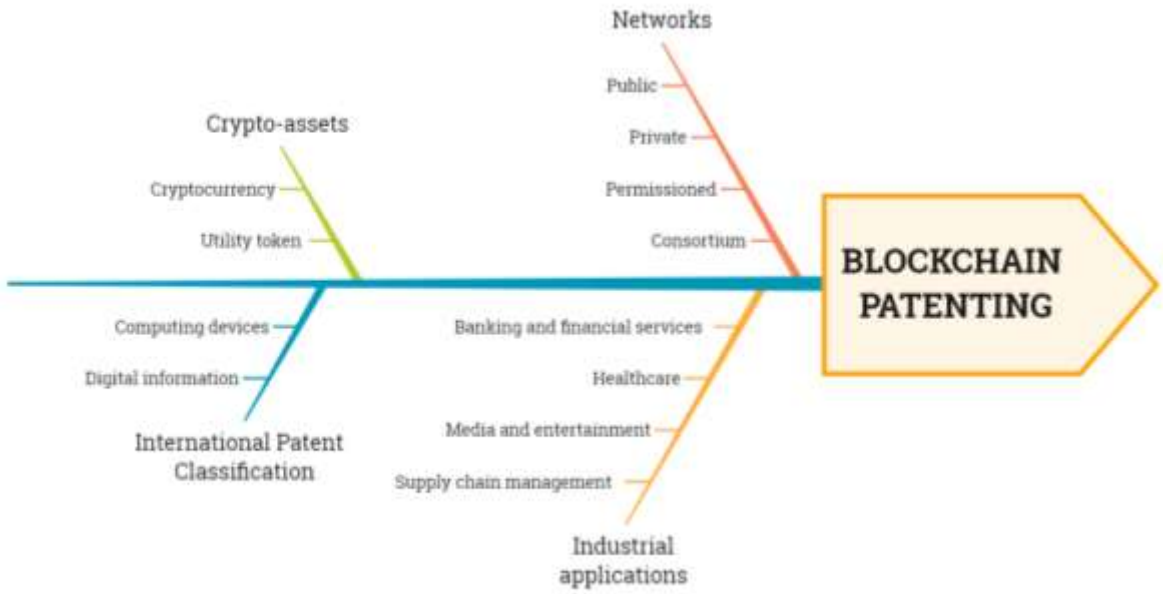


Figure 4: Blockchain patent classification.

***Blockchain networks***

Patents related to blockchain networks include public, private, permissioned, and consortium blockchains (Table 3). Public blockchain networks are blockchain networks where anyone can join as there are no real restrictions on participation. It was found that 158 patents are related to public blockchain networks. Those technologies are mostly related to systems and apparatus for blockchain-based consensus process to enhance security. For example, the invention with the publication number WO2020098837 discloses methods, systems, and apparatus for processing blockchain-based guarantee information.

In contrast, private blockchains are not open for the public people to join in. A single organization or institution controls them. Like public blockchain, patents related to private networks mostly cover security enhancement and encryption algorithms. For instance, WO2020098836 describes a method to verify whether one or more zero-knowledge proofs are correct after receiving a ciphertext.

Permissioned blockchains differ from public and private blockchains. They maintain a layer of access control to allow specific actions only by sure identifiable participants. Patents related to those technologies mostly describe techniques to record, manage, and transfer ownership rights. For instance, the patent identified as US20200074470 details a process to transfer an asset from an asset provider to an asset requester while determining whether the request is valid or not. A consortium blockchain is a platform where multiple organizations control and manage the network. Patents related to consortium blockchain consortium include methods and apparatus for controlling authorization to access user data. It should be noted that some patent documents such as US10880077, WO2020098840, and US10872170 belong to more than on blockchain networks.

Table 2: Distribution of patents based on the type of blockchain networks.

| Network type | No. of patents | Publication No.[a] |
|---|---|---|
| Public blockchain networks | 158 | US10880077; WO2020098840; WO2020256831; US10872170; WO2020098837; US20200394085; WO2020098836; WO2020098838; WO2020098834; US10855449 |
| Private blockchain networks | 147 | US10880077; WO2020098840; US10872170; WO2020098837; WO2020098836; WO2020098838; WO2020098834; US20200382372; US20200382509; US10855447 |
| Permissioned blockchain networks | 105 | US1087852; US20200389301; US20200380825; US20200380826; US10848622; US20200364735; US10833845; US10834095; US20200344073; WO2020216536 |
| Consortium blockchains | 272 | US10880077; WO2020098839; WO2020098840; WO2020256831; US10872170; WO2020098837; US20200394085; US10867299; US10868673; WO2020098836 |

[a] Our table includes the ~~10~~ ten more recent applications based on publication date.

***International Patent Classification (IPC)***

IPC includes digital information and computing devices, as presented in Table 4. The digital information subgroup can be further divided into two components: arrangements for secret or secure communication (H04L 9/32) and arrangements for communication processing (H04L 29/06). Patents related to those technologies mostly describe systems and methods for cryptographic zero-knowledge proof and Merkle-tree-based Proof-of-Work aimed to enhance information security and accessibility. For instance, US10848312 describes systems and methods relating to a zero-knowledge architecture between multiple systems. At the same time, US10873461 discloses systems and methods for zero-knowledge multiparty secure sharing of voiceprints.

Computing devices include two components which are payment architectures and protocols (G06Q 20/38) and Finance and exchange category (G06Q 40/04). Patents related to payment architectures and protocols mostly detail methods and devices for publishing smart contracts and checking their accuracy. For instance, US20190318349 details a method for controlling a smart contract in a network. This method involves accepting the details of the smart contract and evaluating the fulfillment of the conditions. On the other hand, finance, insurance, and detailed exchange information about systems and methods for providing financial data to financial instruments in a distributed ledger technology. For example, US20190164223 provides a system for tracking information related to a financial asset via a distributed ledger.

Table 3: Analysis of blockchain patents using IPC level 5.

| Classification | | IPC code | No. of patents |
|---|---|---|---|
| Digital information | Arrangements for secret or secure communication | H04L 9/32 | 797 |
| | Arrangements for communication processing | H04L 29/06 | 783 |
| Computing device | Payment architectures and protocols | G06Q 20/38 | 568 |
| | Finance; insurance, and exchange | G06Q 40/04 | 297 |

***Crypto-assets***

This category includes cryptocurrency and utility tokens, as presented in Table 4. It was found that 1711 patents with the term cryptocurrency are mentioned in their abstract, title, claim, or description. Those technologies are mostly related to systems and methods for using a shared platform to transmit cryptocurrency. For example, the invention with the publication number WO2020091103 relates to applying a blockchain-based cryptocurrency transaction rule and a terminal device for executing it. The method includes receiving a transmission request for a cryptocurrency transaction.

Results show that 460 patent documents are directly related to a utility token. This blockchain-based asset can help finance future projects. For instance, the patent identified as US20200328891 describes using intelligent, blockchain-compatible asset tokens for non-fungible assets. The intelligent asset tokens can be customized within a three-layer hierarchical identifier representing a hierarchy of asset data.

Table 4: Distribution of patents based on the type of crypto-asset categories.

| Crypto-asset category | No. of patents | Publication No.[a] |
|---|---|---|
| Cryptocurrency | 1711 | US20200410480; WO2020258125; WO2020260864; WO2020258126; WO2020263212; US20200402061; US20200402091; US20200394619; US20200394620; US20200395761 |
| Utility token | 460 | WO2020255372; WO2020252036; US10848449; WO2020212452; US20200328891; WO2020199135; US20200311698; US20200302429; WO2020176228; US10762506 |

[a] Our table includes the ~~10~~ ten more recent applications based on publication date.

***Industrial applications***

Applications include IT services, banking, financial services, healthcare, and supply chain management, as shown in Table 5. Our search strategy reports that IT companies are the leading patent holders of blockchain technology. For instance, the international business machines (IBM) corporation reportedly filled 504 patents during the period starting from July 2006 to April 2020. Those technologies are mostly related to smart contracts and methods for a model-driven extraction of event data that represents an event on a blockchain network. Another IBM invention refers to a blockchain Uniform Resource Indicator (URI) that has access to a blockchain distributed among a plurality of peer blockchain nodes. Moreover, Mitsubishi electric returns 96 blockchain-related innovations. For example, the patent with the application number JP2020016034 details an entry/exit management system that eliminates the possibility of unauthorized access. This invention enables the sharing of a plurality of nodes connected by a P2P network. The authentication conditions are recorded in a block of the blockchain by a management node supervising all the nodes.

Additional results point out that MasterCard, Bank of America, and Visa are among financial companies that massively invest in the field of blockchain technology. Those patents are mostly related to methods and systems for trustworthiness using digital certificates. For instance, the patent application identified as US16947545 provides a way to generate digital certificates for anonymous users in blockchain transactions. This technique involves storing blockchain and transaction values. The central node determines a confidence level based on the data included in each transaction value. Similarly, the application US2020033077 describes a method of transmitting verifiable information regarding the provenance of a cryptographically accurate product while maintaining complete product and participant anonymity on a blockchain. This method includes receiving a product identifier, then generating a digital token by applying a hash algorithm and a digital signature using a private key of a cryptographic key pair.

We also found that companies such as Alibaba Group and Walmart hold inventions related to blockchain technology. Those patents disclose methods, devices, and apparatuses for executing cross-chain anonymous contracts. For example, the invention (US16938650) relates to a method and apparatus for identifying the authenticity of evidence from two parties based on a blockchain ledger. For a specific case event that the accused claims to be accurate, the accused's indictment evidence supporting the case event is identified through the target procedure. In addition, evidence of the respondent's response to falsify the event of the case is identified through the target procedure. At the same time, the patent identified by the application number US2019045627 discloses the system and method for product recall using blockchain. The sales transaction is recorded in a blockchain ledger, and the sales transaction includes customer identification information. A recall report is received regarding the product. The blockchain ledger is accessible, and customers who purchased the recalled product are identified using customer credentials. An action to be taken is determined. The action uses the contact details of customers who purchased the recalled product.

Table 5: Distribution of patents based on the type of crypto-asset categories.

| Industry/Sector | Assignee/Company | Patents retrieved |
|---|---|---|
| Electronics and IT services | IBM | 504 |
| | Mitsubishi Electric | 96 |
| | Microsoft | 94 |
| | Siemens AG | 93 |
| | Intel | 70 |
| Banking and financial services | MasterCard | 133 |
| | Bank of America | 117 |
| | Capital One Services | 73 |
| | Visa | 63 |
| Healthcare | Ping An Technology | 72 |
| Supply chain management | Alibaba Group | 450 |
| | Walmart | 81 |

## DISCUSSIONS AND CONCLUSIONS

The technologies relating to the blockchain are in a dynamic of solid growth and innovation. Some authors nickname it “General Purpose Technology” (GPT), that is, a key technology for the evolution of humanity. Such an essential technological field has a significant impact on patent applications. The patent landscape is an ideal tool to identify the evolution of emerging technologies such as blockchain. It uses a detailed search strategy, including keywords and patent filtering. However, research using the patent landscape to study the evolution of blockchain technologies is scarce. This article proposes a methodology based on the patent landscape and text mining to identify blockchain-related patents.

Based on data mining and international patent classification, the patent documents have been organized into four groups: network types, IPC, crypto assets, and industrial applications. Blockchain networks involve public, private, permissioned, and consortium blockchains, while IPC includes computing devices and digital information. Crypto-assets refer to the cryptocurrency and utility token, while industrial applications involve banking and financial services, healthcare, media and entertainment, and supply chain management.

## REFERENCES


Abou Jaoude, J., & Saade, R. G. (2019). Blockchain applications-usage in different domains. *IEEE Access,* 7, 45360-45381. https://doi.org/10.1109/ACCESS.2019.2902501

Bubela, T., Gold, E. R., Graff, G. D., Cahoy, D. R., Nicol, D., & Castle, D. (2013). Patent landscaping for life sciences innovation: Toward consistent and transparent practices. *Nature Biotechnology,* 31(3), 202-206. https://doi.org/10.1038/nbt.2521

Carbonell, P., Gök, A., Shapira, P., & Faulon, J. L. (2016). Mapping the patent landscape of synthetic biology for fine chemical production pathways. *Microbial Biotechnology,* 9(5), 687-695. https://doi.org/10.1111/1751-7915.12401

Chen, Y.S., Lin, M.J.J., Chang, C.H., & Liu, F.M. (2009). Technological innovations and industry clustering in the bicycle industry in Taiwan. *Technology in Society,* 31(3), 207-217. https://doi.org/10.1016/j.techsoc.2009.06.001

Choi, D., & Song, B. (2018). Exploring technological trends in logistics: Topic modeling-based patent analysis. *Sustainability,* 10(8), 2810. Retrieved from https://www.mdpi.com/2071-1050/10/8/2810.

Clarke, N. S., Jürgens, B., & Herrero-Solana, V. (2020). Blockchain patent landscaping: An expert based methodology and search query. *World Patent Information,* 61, 101964. https://doi.org/10.1016/j.wpi.2020.101964

Dhankhar, R., Gulati, P., Kumar, S., & Kapoor, R. K. (2018). Arginine-lowering enzymes against cancer: A technocommercial analysis through patent landscape. *Expert Opinion on Therapeutic Patents,* 28(8), 603-614. https://doi.org/10.1080/13543776.2018.1508452

Evangelista, A., Ardito, L., Boccaccio, A., Fiorentino, M., Petruzzelli, A. M., & Uva, A. E. (2020). Unveiling the technological trends of augmented reality: A patent analysis. *Computers in Industry,* 118,103221. https://doi.org/10.1016/j.compind.2020.103221

Hassani, H., Beneki, C., Unger, S., Mazinani, M. T., & Yeganegi, M. R. (2020). Text mining in big data analytics. *Big Data and Cognitive Computing,* 4(1), 1. https://doi.org/10.3390/bdcc4010001

Kakarlapudi, P. V., & Mahmoud, Q. H. (2021). A systematic review of blockchain for consent management. *Healthcare,* 9(2), 137. http://dx.doi.org/10.3390/healthcare9020137

Kim, G., & Bae, J. (2017). A novel approach to forecast promising technology through patent analysis. *Technological Forecasting and Social Change,* 117, 228-237. https://doi.org/10.1016/j.techfore.2016.11.023

Kumar, S., Kar, A. K., & Ilavarasan, P. V. (2021). Applications of text mining in services management: A systematic literature review. *International Journal of Information Management Data Insights,* 1(1), 100008. https://doi.org/10.1016/j.jjimei.2021.100008

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

Li, X., Xie, Q., Jiang, J., Zhou, Y., & Huang, L. (2019). Identifying and monitoring the development trends of emerging technologies using patent analysis and Twitter data mining: The case of perovskite solar cell technology. *Technological Forecasting and Social Change,* 146, 687-705. https://doi.org/10.1016/j.techfore.2018.06.004

Paulavičius, R., Grigaitis, S., Igumenov, A., & Filatovas, E. (2019). A decade of blockchain: Review of the current status, challenges, and future directions. *Informatica,* 30(4), 729-748. https://doi.org/10.15388/Informatica.2019.227

Stoffels, M. A., Klauck, F. J., Hamadi, T., Glorius, F., & Leker, J. (2020). Technology trends of catalysts in hydrogenation reactions: A patent landscape analysis. *Advanced Synthesis & Catalysis,* 362(6), 1258-1274. https://doi.org/10.1002/adsc.201901292

Xu, M., Chen, X., & Kou, G. (2019). A systematic review of blockchain. *Financial Innovation,* 5(1), 1-14. https://doi.org/10.1186/s40854-019-0147-z