

Article

A Survey on Blockchain-Based Search Engines

Esmaeel Rezaee^{1,*} , Ali Mohammad Saghiri²  and Agostino Forestiero^{3,*} ¹ Department of Computer Engineering, Faculty of Computer and Industrial Engineering, The Birjand University of Technology, Birjand 009856, Iran² Information and Communications Technology Research Department, Niroo Research Institute, Tehran 009821, Iran; amsaghiri@nri.ac.ir³ Institute for High Performance Computing and Networking (ICAR), National Research Council of Italy (CNR), 00185 Rende (CS), Italy

* Correspondence: rezaei@birjandut.ac.ir (E.R.); agostino.forestiero@icar.cnr.it (A.F.)

Abstract: With the increasing growth of different types of data, search engines have become an essential tool on the Internet. Every day, billions of queries are run through few search engines with several privacy violations and monopoly problems. The blockchain, as a trending technology applied in various fields, including banking, IoT, education, etc., can be a beneficial alternative. Blockchain-based search engines, unlike monopolistic ones, do not have centralized controls. With a blockchain-based search system, no company can lay claims to user's data or access search history and other related information. All these data will be encrypted and stored on a blockchain. Valuing users' searches and paying them in return is another advantage of a blockchain-based search engine. Additionally, in smart environments, as a trending research field, blockchain-based search engines can provide context-aware and privacy-preserved search results. According to our research, few efforts have been made to develop blockchain use, which include studies generally in the early stages and few white papers. To the best of our knowledge, no research article has been published in this regard thus far. In this paper, a survey on blockchain-based search engines is provided. Additionally, we state that the blockchain is an essential paradigm for the search ecosystem by describing the advantages.



Citation: Rezaee, E.; Saghiri, A.M.; Forestiero, A. A Survey on Blockchain-Based Search Engines. *Appl. Sci.* **2021**, *11*, 7063. <https://doi.org/10.3390/app11157063>

Academic Editor: Gianluca Lax

Received: 24 June 2021

Accepted: 28 July 2021

Published: 30 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: search engine; monopoly; decentralization; blockchain; security; privacy

1. Introduction

Every day, billions of queries are run through search systems that a few companies own. These companies now control almost all of the search market. Additionally, these few companies monopolize the information served up on the web, which means they control any discourse. Additionally, these companies gather user information during searches to improve their systems. They do not pay users and do not care for privacy concerns. Conventional search engines have several problems such as privacy violation, monopoly, a limited selection of search space, and limited object indexing.

The mentioned issues have led to alternatives such as peer-to-peer search engines. These search systems are developed on a peer-to-peer network that is distributed and, in this regard, have the infrastructure to provide privacy-preserving search engines. However, since peer-to-peer applications do not have the required mechanism to support all dimensions of privacy, they cannot solve privacy issues. Blockchain technology is another beneficial alternative that can provide a novel solution to handle privacy and monopoly issues by using cryptography and consensus mechanisms. Cryptography by encrypting users' data and creating a private key provides a privacy preserved solution to protect users' data. Additionally, the consensus mechanisms that use a distributed voting approach avoid monopoly. These mechanisms protect data from change by few nodes.

Blockchain-based search engines, unlike conventional, do not have centralized controls. Instead, they are distributed across multiple networks. With a blockchain-based

search system, no company can lay claims to users' data or access search history and other related information. All these data will be encrypted and stored on a blockchain. Therefore, instead of a monopolist company controlling data, they will be in total control of their owners. Blockchain-based search engines can further provide users with a private key to protect their data. Valuing users' searches and paying them in return for use is another advantage of a blockchain-based search system.

Today, the blockchain has grown to be a trending technology, finding applications in various fields, including banking [1], IoT [2], healthcare [3], media [4], education [5], etc. On the other hand, blockchain technology improves several aspects in various applications such as file sharing [6], video streaming [7], cloud [8], fog and edge computing [9], smart grid [10], message passing [11], and database [12]. Therefore, considering the potential of the blockchain, using it in developing search systems will result in a new generation of search engines with new capabilities which solve the problems of the conventional ones.

Conventional search engines and their monopolized market, user's concerns, and the popularity of blockchain to solve problems indicate the need to establish a blockchain-based search system. The monopolized market includes ranking and advertising, and the users' concerns, include privacy and security. In this regard, few efforts have been made to provide blockchain-based search engines, which are Nebulas [13], Bit Clave [14], Presearch [15], and DeSearch [16]. These systems are generally in early stages, and their structures have been reported only as a few white papers.

In this paper, first, the blockchain applications are reviewed to show the popularity of this technology and its potential to provide new solutions. Then, a survey on blockchain applications in developing search systems is provided. Describing the advantages of the blockchain-based search system, we elucidate that blockchain technology is an essential paradigm for search ecosystems. To the best of our knowledge, no research article has been published in this regard thus far.

To show the motivations, the innovations of this research are expressed. The main contributions of this paper include the following:

- The search engine paradigms are described, and the problems of centralized as a dominated paradigm are provided;
- It provides a review of blockchain applications to show the popularity of this technology;
- A survey on blockchain-based search engines is provided to show the validity of blockchain-based generation;
- The blockchain characteristics are described and the way that it can address the problems of centralized search engines.

This paper is structured as follows: Section 2 provides an overview of the wide range of blockchain applications. Section 3 describes search engines and their approaches. Section 4 focuses on efforts made to propose blockchain-based search systems. Section 5 describes the role of blockchain-based search engines in smart environments. Section 6 discusses integration using the blockchain to improve the search ecosystem. Section 7 concludes the paper and offers future research directions.

2. Blockchain and Its Applications

In this section, first, the architecture of the blockchain is described. Subsequently, an overview of the wide range of using blockchain as infrastructure, including file sharing, video streaming, cloud, fog and edge computing, messaging system, grid computing database, and web hosting, is provided.

2.1. Blockchain Architecture

Blockchain is a shared public ledger. The information of the actual transaction is recorded in this ledger. A blockchain consists of blocks, each of which contains a set of valid trades. Each block has a hash code of the previous block, and the link between the blocks is created through this hash; finally, the linked blocks form a chain. The risk

of centralized data retention is eliminated with blockchain [17]. The general blockchain architecture is presented in Figure 1 [18].

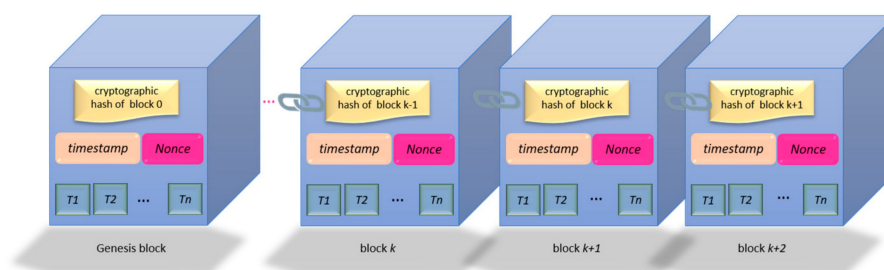


Figure 1. Blockchain architecture.

The blockchain is built on a peer-to-peer network. The blockchain architecture uses two critical characteristics of these networks: distributed and self-organized computation. Different peer-to-peer network technologies have recently been widely used to develop novel blockchain applications [19].

2.2. Blockchain Applications

Nowadays, blockchain has grown to be a trending and promising technology. Blockchain proponents argue that the technology has nearly limitless applications in various fields, including banking, electricity, IoT, health, media, voting, education, etc. [20]. On the other hand, blockchain technology is used as infrastructure to improve several aspects in various applications. The main reason for this broad application of blockchain as infrastructure is the cryptographic techniques and consensus mechanisms; the blockchain handles the users' transactions as a sequence of blocks in a ledger and protects them by encrypting [19]. In this section, an overview of the wide range of blockchain applications is provided as the research framework. Figure 2 depicts a taxonomy of the applications of the blockchain. In the following, the applications of blockchain technology are discussed.

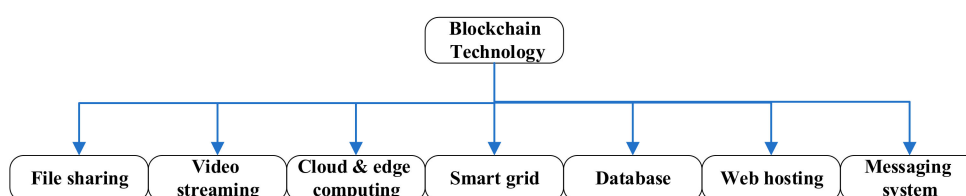


Figure 2. The applications of blockchain technology.

2.2.1. File Sharing

In an autonomous distributed ecosystem, such as peer-to-peer file-sharing networks, some selfish nodes can cause the system to collapse. In such an ecosystem, blockchain technology can meet the needs of multiple users [6]. In [21], a robust decentralized application platform for file sharing and data provenance on blockchain technology was introduced. The proposed architecture solves the integrity and ownership problems, which are the main issues in the standard file-sharing systems. In the proposed scheme, user registration and provenance are handled by a decentralized application (DApp) based on Ethereum. From the beginning to the most recent version, the Ethereum smart contract governs, monitors, and provides traceability and visibility into the shared content's history. Its data storage layer is IPFS, a distributed file system that avoids the drawbacks of centralized storage solutions. The proposed framework uses an inbuilt editor to access and change files only by the owner.

2.2.2. Video Streaming

The video streaming industry is being reshaped by blockchain technology. The blockchain changes the way people make, distribute, and consume videos. Some emerging video streaming platforms leverage blockchain technology to develop peer-to-peer content delivery architectures and cryptocurrency-based payment systems to remove centralized video streaming servers. The source video streams must be transcoded into multiple versions to satisfy heterogeneous user demands on these blockchain-based platforms, a computationally and energy-intensive process [22]. In [22], a new MEC-enabled transcoding platform for blockchain-based video streaming was presented, while the underlying blockchains are built with an adaptive block size scheme. In [23], a new algorithm for using blockchain in live stream video transfers was suggested. The new proposed algorithm creates a safe link that sends and receives data packets in any format (boxes or containers of information) in a secure and encrypted manner while keeping the sender and receiver anonymous.

2.2.3. Cloud, Fog, and Edge Computing

Cloud computing is a computing paradigm that also provides collaboration and facilitates ubiquitous access to computing according to a user's request. This technology offers new data processing and services in various industries while also lowering computing and storage costs for users and improving ease of use [24]. Combining blockchain technology with existing cloud services has great potential for improving functionality, efficiency, and security/privacy [8]. One of the most critical protection technologies for securing confidential data stored in the cloud by businesses and individuals is access control. Since the cloud uses a centralized access control system, personal data in the cloud are easily tampered with or leaked by hackers or cloud internal managers. In [24], AuthPrivacyChain, a blockchain-based access control system with privacy and security, to address this problem was presented. In [25], a new distributed virtual machine agent model was implemented in the cloud using mobile agent technology to solve data protection and trusted computing, which are significant challenges in current cloud computing applications. Multi-tenants will collaborate to ensure data trust verification using the virtual machine agent—the tasks of storing, tracking, and securing secure data. In [26], a new blockchain-based database was proposed to face data integrity threats.

Edge computing is a new distributed computing paradigm that vows to bring processing and data storage closer to where they are needed, saving network traffic and lowering request latency [27]. Decentralized administration and security are currently the main issues for this technology. The integration of blockchain and edge computing into a single system can provide safe network access, storage, and computation at the edges [9]. In [28], the blockchain was used as a framework for hierarchical and distributed control systems. Hyperledger Fabric was chosen as the blockchain solution in this study, and the interaction with the edge nodes was performed using micro-services architecture. Additionally, [29] provided a mobile edge computing-enabled blockchain system and a revolutionary notion of edge computing for mobile blockchain. IoT devices or mobile users can access and employ resources or computing services from an edge computing service provider in the proposed network to enable their blockchain applications.

2.2.4. Smart Grid

The advancement of network technology assists the growing trend of smart grid implementation, as the wired environment provides a variety of options for electrical data collection. Allowing intelligent controls in the smart grid and providing various data-sharing methods are necessary. However, while versatile communication solutions are offered, security and privacy issues emerge, such as energy depletion and infrastructure mapping attacks. By integrating blockchain and edge computing techniques, [30] proposed a permission-based blockchain edge model for intelligent grid networks (PBEM-SGN) to solve the two major issues in the smart grid—privacy rights and energy security. For

edge-computing-based intelligent grid networks, in [31], blockchain-based shared authentication and key agreement protocol were proposed. Without using other complex cryptographic primitives, the proposed protocol supports practical conditional anonymity and critical management.

2.2.5. Messaging System

Instant Messaging (IM) allows two or more people to communicate in real-time over the Internet. Most IMs nowadays take place on mobile apps such as WhatsApp, WeChat, Viber, and Facebook Messenger, which have a more extensive user base than social media sites such as Twitter and Facebook. Online shopping has become a part of our daily lives, especially for those who are the busiest. Online shopping, however, often results in transaction conflicts. The messaging between users and owners of online shops is unreliable and traceable since most IMs are centralized and message history is not kept in the middle. In [32], new methods for securing instant messaging by combining blockchain and machine learning algorithms were presented. A blockchain-based instant messaging scheme based on Chinese cryptographic bases was proposed in this study. In [33], an Ethereum-based privacy-preserving social media platform is presented. The proposed platform overcomes some issues in traditional social media, such as privacy and censorship.

2.2.6. Database

The distributed replicated database is one of the most popular Blockchain applications. In [34], BigchainDB as a scalable blockchain-based database was proposed. BigchainDB fills a gap in the decentralization environment by providing a scalable decentralized database. It boasts a throughput of one million writes per second, storage of petabytes of data, and latency of less than a second. BigchainDB begins with a distributed database (DB) and introduces blockchain features such as decentralized power, immutability, and the development and movement of digital assets through a series of actions. BigchainDB has many of the same features as modern distributed databases, such as linear scaling of throughput and storage. The number of nodes increases, along with a full-featured NoSQL query language, effective querying, and managing permissions. In [35], a new resilient distributed database system was submitted. Geo-Scale Byzantine Fault Tolerant Consensus Protocol (GeoBFT) is a new geo-scale Byzantine fault-tolerant consensus protocol proposed in this paper. By grouping replicas in local clusters in a topologically aware manner, implementing parallelization of consensus at the local level, and minimizing communication between clusters, GeoBFT is optimized for scalability. In [36], a new decentralized replicated relational database with blockchain properties was designed and implemented. The new database is termed a blockchain relational database. This new database is implemented on PostgreSQL. Initial research into the challenges and benefits of using blockchain as a database for Internet-of-Things applications was discussed [12]. This paper focuses on bitcoin's enabling technology, the Bitcoin Backbone Protocol (BBP), and proposes a clear method for ensuring that a BBP-based database meets both read–write and eventual compatibility requirements.

2.2.7. Web Hosting

Web applications, because of their capabilities such as infrastructure independence, are popular in many fields such as banking services, markets, social networks, web search, e-mail, and interactive information. Almost all web applications run based on client-server architecture. In this architecture, source codes and website data are stored in centralized systems, and users use browsers to send requests to these web servers. This architecture has several limitations, such as security, privacy, and access control. A decentralized solution for web hosting is a new technology process that provides an effective mechanism for storing and accessing websites [37]. Blockchain and encryption technology is a good solution for decentralized hosting, eliminating current centralized web hosting systems. In [38], based on blockchain technology, encryption, and the framework of the inter-planet

file system, a new protocol for web hosting was proposed that provides three critical features. The first feature ensures anonymous customer information. The second feature offers confidentiality and authentication to transfer source handles of websites between customers and service providers. Additionally, the latest is responsible for supplying websites from other nodes in the network.

Concluding the blockchain advantages, Table 1 depicts the capabilities of the blockchain technology considered in each application. As can be inferred from this table, security, privacy, access control, and anonymity are the most popular capabilities used by almost all applications. Data integrity trusted computing, visibility, scalability, immutability, and fault tolerance are the applications used for special applications. This review demonstrates the vast capabilities of blockchain technology that can be used in any field to solve similar problems. As mentioned in Section 1, privacy and monopoly are the most critical issues in conventional search engines that blockchain capabilities can solve.

Table 1. Applications and blockchain capabilities.

	File Sharing [6,21]	Video Streaming [22,23]	Cloud, Fog and Edge Computing [8,24–26]	Smart Grid [29,30]	Messaging System [31,32]	Database [33–35]	Web Hosting [36,37]
Security		✓	✓	✓	✓		✓
Privacy			✓	✓	✓		✓
Access control	✓		✓			✓	✓
Anonymity	✓	✓		✓			✓
Data	✓		✓				
Integrity	✓						
Visibility	✓						
Data protection			✓				
Safety		✓					
Trusted Computing			✓				
Scalability						✓	
Immutability						✓	
Fault tolerance						✓	

3. Search Engines and Their Approaches

Search engines are the most powerful tools for finding information on the Internet. Since traditional search systems were no longer sufficient and suitable due to the increase in web pages and the amount of data on the web, the semantic search paradigm was born. The point of semantic search is to use meaning to improve the user's search experience. There are four approaches to semantic search, including contextual analysis, reasoning, natural language understanding, and ontology-based knowledge representation [39]. According to another taxonomy, search engines classify as centralized (standard) and decentralized [40]. The dominant approach for conventional search is centralized. This approach has several limitations that are described in the following sections.

3.1. Centralized Search Engines and Their Problems

The primary approach for conventional search systems is centralized. The centralization in this context refers to the controlling methods. The centralized are controlled by a single group [18]. Large commercial search engines (such as Google, Yahoo, and Bing) use this paradigm. According to the architecture of search systems, almost all of them consist of three parts, including a web crawler for collecting information, a file system or

database to store information, and a searching subsystem to find the relevant document to a query [40,41].

The following is a brief description of two primary subsystems, including the crawler and the searcher:

- The crawler: This offline subsystem retrieves web pages and creates a text-based index. A web crawler is a type of software that, given one or more seed URLs, downloads the web pages associated with these URLs, extracts any hyperlinks, and then downloads the web pages recursively [42,43].
- The searcher: This subsystem, which is an online component, processes a stream of queries. The search algorithm ranks documents depending on various parameters, including the quantity of anchor text hits and the document's PageRank. Counts of hits of different types at various proximity levels are computed for each matching document. These counts are then routed through many lookup tables before being translated into a rank. The purpose of searching is to deliver high-quality results quickly.

From the point of view of architectural structure, these subsystems can be distributed. For example, as stated in [44], Google uses several distributed crawlers to handle hundreds of millions of web pages and save the retrieved data in multiple distributed servers. Although these companies may use distributed architecture, they have a centralized/monopolistic control that leads to many problems such as privacy violation and monopoly, as explained below.

- Privacy: user privacy violation is one of the most critical problems of popular search engines, with the typical track user searches and other online activities. They make user profiles by storing user search information and analyzing it to improve their [45–47], processes.
- Monopoly: Monopolization about search engines means that they manipulate the search results. In other words, monopoly means the deviations from some kind of “objective” or “unbiased” standard for delivering search results. The enormous market power of companies such as Google provides monopolization for these companies [48]. This monopoly results in information distortion (disinformation) [49], loss of social welfare, and, for sites not covered, the loss of economic value [50].

3.2. Decentralized Search Engines

A decentralized search engine does not have a single server. In this kind of search system, the tasks such as crawling, data mining, indexing, and query processing are spread among several peers in a decentralized way, with no single point of control, in contrast to conventional centralized. FAROO [51] and YaCy [52,53] are two examples of distributed search engines. Both are based on peer-to-peer (P2P) network concepts. In order to address the privacy problems that exist in centralized search systems, distributed search systems have been proposed. The privacy challenges include censorship, the exposure of personal queries to the search, the lack of transparency of proprietary search algorithms and the server, and any third parties with whom the search service operator can share data. There are three critical criteria for a successful search from the user's perspective: a high-quality answer, a quick response time, and a comprehensive web collection available in the index. Distributed search can satisfy the user expectations using the network topology, caching layers, high concurrency, and the local query [41]. The architecture of YaCy, a decentralized search engine, is presented in Figure 3.

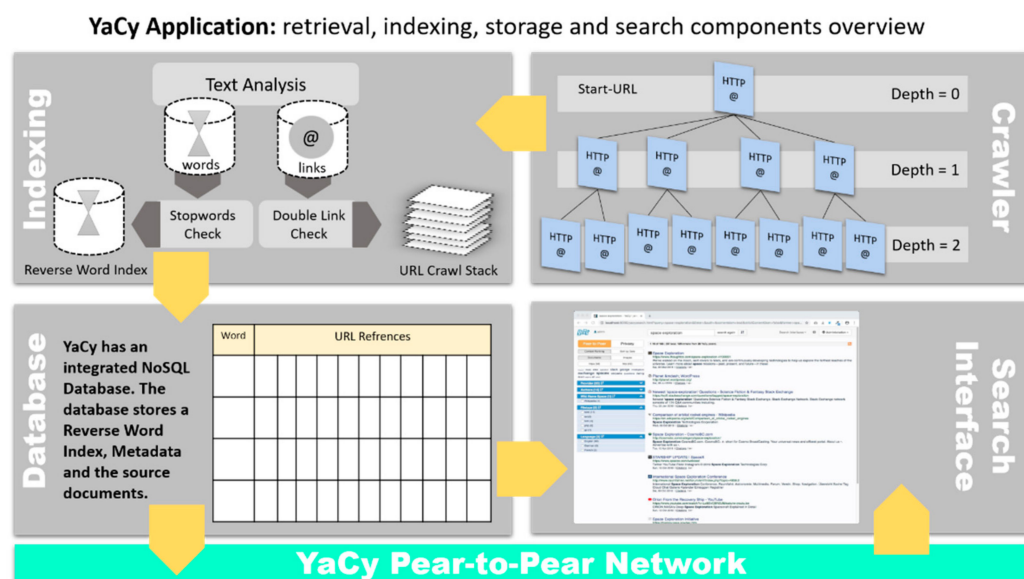


Figure 3. The YaCy architecture [52].

The YaCy framework is divided into five parts, namely, YaCy P2P network, crawler, indexing, database, and search interface. YaCy harvests web pages with a web crawler, indexes, parses, and saves the search index locally. When a peer joins a peer network, its local search index is combined with the network's mutual index. When a search is initiated, the local index and a global search index both contribute.

3.3. Blockchain-Based Search Engine to Address the Problems

Blockchain is a chain of blocks that store information securely in a decentralized and distributed network using cryptography techniques. Decentralization, immutability, transparency, and auditability are all features of blockchain that make it more secure and tamperproof. The following are four essential characteristics of blockchain, including decentralization, anonymity, persistency, and auditability.

- **Decentralization:** Each transaction in a traditional centralized transaction system must be validated by a central trusted agency. This centralization results in cost and performance bottleneck and hence monopolization in the delegated agency decisions. A transaction in the blockchain network can be completed between any two peers (P2P) without central agency authentication. In this manner, blockchain can reduce the trust concern and prevent monopoly in any dimension of the trusted agency.
- **Anonymity:** With a created address, each user can communicate with the blockchain network. Furthermore, a user could generate a large number of lessons to protect their identity. There is no longer a central authority in charge of users' personal information. This approach ensures that the transactions in the blockchain are kept private to some extent.
- **Persistency:** It is nearly impossible to tamper with the network because each transaction must be validated and recorded in blocks distributed throughout the network. Other nodes would also validate each broadcasted block, and transactions would be examined. As a result, any falsification would be easily detectable.
- **Auditability:** A digital distributed ledger records all transactions in a blockchain network and verifies them with a digital timestamp. As a result, any node in the network can be accessed to audit and trace records.

According to the characteristics of the blockchain, blockchain-based search engines, unlike their conventional counterparts, do not have centralized controls. Instead, they are distributed across multiple networks. By decentralization, the trust concerns are reduced, and centralized control's side effects, such as monopoly, are eliminated.

Additionally, considering anonymity, with a blockchain-based, no company can lay claims to users' data or violate user privacy. The persistency makes it impossible to tamper with the data because all data will be validated, encrypted, and stored on a network of blocks. Blockchain-based search ecosystems can further provide users with a private key to protect their data.

4. The Blockchain-Based Search Engines

This section focuses on efforts made to propose a blockchain-based search engine, including Nebulas, Bit Clave, Presearch, and DeSearch. All these engines are still in improvement phases.

4.1. Nebulas

Nebulas [13], is a blockchain framework committed to the vision “to allow everyone to benefit fairly from decentralized collaboration.” It also presented novel technologies such as Nebulas Rank (NR), Developer Incentive Protocol (DIP), Nebulas Force (NF), and Nebulas' unique approach to on-chain consensus entitled Proof of Devotion (PoD). The Nebulas platform will include a blockchain that will index and rank the platform's Dapps, smart contracts, and info, as well as external blockchains. The aim of Nebulas, the world's first global blockchain search engine, is to discover new blockchain utilities [13].

Nebulas Rank is a tool that lets users find and use valuable information in the ever-growing number of data on the blockchain. Nebulas Rank is developing a metric to rank blockchain applications based on a universal measure of worth. The open source core-ranking algorithm ranks addresses, smart contracts, decentralized applications (Dapps), and other blockchain entities based on liquidity, asset propagation, and interactivity between users by using on-chain transaction records as source data. Nebulas ranking attempts to establish a trustful, computable, and deterministic measurement approach. The novel ranking algorithm Nebulas Rank simplifies access to useful blockchain info. More and more outstanding applications will emerge on the Nebulas platform as the importance ranking system is implemented.

4.2. BitClave

BASE (BitClave [14], Active Search Ecosystem) is an open, public, decentralized platform that allows users to use and share their data selectively. This platform allows for direct customer-to-business communications without the use of intermediaries. End users, companies, and ecosystem members are the key participants in BASE. End users have power over their identity and who has access to their data in BASE. End users can save time and effort by receiving deals that are ideally suited to their needs, owing to BASE. The BASE will provide users with highly targeted promotions based on personal information they voluntarily share. Concentrating their promotional efforts on closely matched end users helps them achieve productivity in consumer acquisition and retention. By engaging in BASE, BASE partners will generate additional revenue streams. When a successful transaction between companies and consumers occurs due to services rendered, ecosystem members are compensated. In other words, the BASE is a decentralized marketplace [14].

BitClave's solution allows large and small businesses to collaborate in an ecosystem. Users and other data providers are rewarded on the value of their data, which are determined by conversions (not clicks), owing to the creative search technology, which allows sellers to pay directly for outcomes, maximizing marketing spend for performance. Consumers benefit from cash rewards and lower-cost commodities while also having control of their data.

For the first time, BitClave allows for the optimization of the seller–consumer value chain. Third parties are encouraged to provide data that maximize seller–consumer business efficiency, advertising, and deals, resulting in the lowest-cost outcome in the BitClave ecosystem. Analytics providers will participate in a diverse app ecosystem and be rewarded for their creativity, funded by improved overall business performance—all benefits

from removing the waste from today's massive advertisement network solutions. As technology advances, BitClave's open ecosystem ensures that competition in the advertising industry continues.

Advertising, a multibillion-dollar industry, is disrupted. Businesses today have to deal with a slew of intermediaries in the advertisement stack. The dominant advertisement firms (e.g., Google, Amazon, Facebook) charge exorbitant fees to reach customers. BitClave uses Blockchain technology to allow mutually distrusting parties to participate in mutually beneficial co-enterprise without a central authority. Decentralized search is the next step in the evolution of ads as a tool for connecting people and businesses. BitClave reimagines a more open and meaningful platform for people and companies to participate in value creation using these influential design trends. BitClave technology allows for scalable, safe, and anonymous search through large datasets and data privacy and a mechanism to transact valuable data with attestation. BitClave was established to reimagine business–customer relationships based on the confidence and transparency of smart contracts. The proposed solution can destabilize one of the world's biggest industries, the ad network, which is currently monopolized by multinational companies and dominated by middlemen.

4.3. Presearch

Presearch [15], builds a complete ecosystem to support the PRE token and provide the world with a decentralized search. The ecosystem includes research tokens, decentralized search, Presearch nodes, keyword staking, Presearch marketplace, Ethereum blockchain [15]. Presearch has developed a modern distributed architecture, which addresses several specific challenges that centralized search systems do not face, such as preventing cybercriminals from running nodes and either extracting user information or delivering dangerous or unwanted content, achieving quick response times across a massively distributed network of servers with drastic performance and reliability variability, enabling people to run nodes, and fairly balancing supply and demand for both nodes and searches within the network are all challenges.

Users lose ownership of information about themselves when using conventional search engines, but they also encourage others to profit from that information. Presearch alters the equation by allowing users to benefit from their actions—whether or not personal information is revealed, searches create tokens for the searching user. After all, even though no personal information is exchanged, search queries are helpful. Users can receive Presearch reward tokens as members of the Presearch group when searching, running a node, and referring others to enter Presearch. In other words, Presearch provides a world-class search experience that rewards user searches based on the actual value of advertising revenue generated by those searches.

Another feature of Presearch is transparent ads. The lack of clarity surrounding the cost per click is one of advertisers' significant concerns. Fraud is another primary concern for advertisers. The Presearch keyword marketing platform, similar to Google AdWords, is a tool for greater advertisement transparency. However, there are a few key differences due to the project's early stages and the unique ability to account for and pass value using blockchain tokens. Due to the high importance of search traffic, which is highly qualified based on the keyword entered, there is an enormous opportunity, similar to AdWords. Presearch can use its property to draw advertisers and kickstart the launch of a keyword staking ad network. Other apps, advertisers, and search engines would ultimately use the platform to extend keyword staking on their assets, resulting in more use cases for PRE and opportunities to generate value for the network. Staking your PRE tokens on a particular word or phrase is known as keyword staking. Your ad will appear when someone searches for that term on Presearch if you stake the most tokens on it. Although you retain control of your PRE while they are staked, your ad will no longer be visible if you want to bet them from a term or if anyone stakes more tokens than you. Since they are extremely attractive

for advertisers, keyword advertising on search systems is some of the most lucrative ads throughout the world.

4.4. DeSearch

DeSearch [16], is the first cryptocurrency-specific search engine in the world. The BitClave mission in motion is DeSearch. DeSearch's powerful search capability also searches the top blockchain pages, including Telegram networks, BT, company and influencer medium material, and Reddit, for the most up-to-date information on ICOs and all crypto. The features of DeSearch are prioritized, which emphasize accessibility, relevant information, and privacy, making it an excellent crypto resource. DeSearch is currently in alpha development [16].

Table 2 shows the state of research by reviewing the existing blockchain-based search engines from three primary criteria—system scale, searching content, and inclusion of cryptocurrency.

Table 2. Comparison of existing blockchain-based search engines.

	Nebulas	BitClave	Presearch	DeSearch
System Scale	Decentralized ecosystem	Decentralized marketplace	Decentralized ecosystem	Decentralized Crypto Search Engine
Searching Content	DApps, smart contracts, and info.	Customer, businesses, and ads.	Web information, ads.	Cryptocurrencies, Web pages, Telegram networks, BT, company, and influencer medium material.
Cryptocurrency	Not reported	Not reported	PRE	Not reported

As shown in Table 2 according to system scale, some proposed systems such as Nubulas and Presearch provide a decentralized ecosystem. In these systems, the decentralized search engine is a subsystem in contract with the ecosystem. However, some search engines such as Desearch provide only a decentralized search system with a special mission. Additionally, according to the searching content, almost all the proposed systems index and search more content types than the conventional. These systems index DApps, smart contracts, customers and businesses, cryptocurrencies, and the content of social networks such as Telegram and BitTorrent. Finally, according to cryptocurrency, Presearch reported the PRE, but others were not reported. The cryptocurrency is used to raise capital and offer tokens due to search and trade-in markets.

5. Blockchain-Based Search Engines in Smart Environments

A smart environment is a sophisticated setting with dynamic devices that create an “intelligent entity”. In these environments, because computational power is incorporated in the space, it is critical to explore its perspective and interactions with the surrounding world [54,55]. In this context, edge computing provides sound solutions to support analytic data applications, high-performance computing, and big data storage. Edge computing is a new distributed computing paradigm that promises to bring processing and data storage closer to where they're needed, saving network traffic and lowering request latency [56].

Search engines for smart environments are generally different from standards. Heterogeneity, interacting with other apps, a wide range of usage scenarios, a vast number of data exchange protocols, and data privacy are the critical challenges in smart settings [57]. Therefore, we need a distributed search system with unique capabilities to provide context-aware and personalized search results.

Blockchain-based search engines can provide an excellent solution to support personalized search results to the privacy principles. In edge computing, an emerging distributed computing paradigm in smart environments, considering the continuous increase in edge

capabilities such as processing power and storage space, the approach transfers the processing from the center of the cloud to the edges. More context-aware and personalized search results can be provided by processing the user's background data and the information about the user context. In context-based services, the disclosure of user's data and privacy violations are the fundamental problems. In the blockchain-based search engines, encrypting the data, storing it on edge, and processing locally, user privacy will be provided. Additionally, using the edge processing power and storage capacity, the overall search performance will be improved.

6. Discussion

Conventional search engines have several problems, such as privacy violations and monopoly. Considering the blockchain's popularity and capabilities, such as distribution and cryptography, this technology can solve these problems. The blockchain is used in cloud computing, smart grid, and messaging systems to solve privacy and security issues. Additionally, it is used in file sharing and grid computing to solve problems such as access control, data integrity, and visibility related to monopoly. Using the blockchain as the primary approach in developing search results in powerful search capabilities that solve the mentioned problems. In the following, for better discussion, some critical questions have been answered.

- Q1: What are the benefits of integrating blockchain and search engines technologies?
 - Give users control of their data and the ability to share it selectively to protect user privacy;
 - Disrupt the existing ad network, which is monopolized by large companies and dominated by middlemen, and create a competitive market that benefits all;
 - Enable scalable, secure, and anonymous search over massive datasets;
 - Allow everyone the opportunity of seeing the search results;
 - Enable users to profit from their actions—whether or not personal information is revealed, searches create tokens for the searching user;
 - Allow everyone to benefit significantly from decentralized collaboration;
 - Eliminate the need for intermediaries by allowing direct customer-to-business interactions;
 - Provide the context-aware and privacy-preserved search results in smart environments.

- Q2: What are the future dynamics of research in the blockchain-based search engines arena?

As a future dynamic of research in this area, we may refer to possible combinations among search engine technology with digital twins, digital money, and IoT. Merging digital money platforms such as those reported in the blockchain technology domain with search engines might be the first substantial evolution of search technologies in the era of digital transformation. For example, blockchain-based search systems might be customized for IoT to bring a powerful search considering IoT's security and privacy issues.

- Q3: Why do we need blockchain-based search engines?

It is worth noting that the computational power of computers of the users increases day by day, and therefore, we must design efficient distributed search systems considering this potential. All the distributed search engines, such as blockchain-based technology, are not widespread for the users. This is because some of them attempt to save a centralized control. In other words, search engines may manipulate the results according to some arguments. On the other hand, decentralized architecture will be eliminated by decentralized architectures such as blockchain because the users prefer to utilize those technologies that protect their privacy.

- Q4: What development hurdles can we face while shifting from centralized to distributed?

As the main development hurdles that we can face while shifting from existing search ecosystems to blockchain-based ecosystems, blockchain infrastructures are costly. Therefore, the development of search systems based on them may lead to complex tasks.

- Q5: Who can welcome such search engines, and why is there a need for blockchain-based search engines, to name a few?

Appearing in a blockchain-based search system can lead to an important phenomenon during digital transformation that may involve many new users and critical participants in search ecosystems. Presearch projects [15] prove the benefits of these systems for the users. Every user that has high activity in search space can obtain a valuable token. In other words, the economic cycle of the search system will not be restricted to a company or government.

7. Conclusions and Future Research Directions

Considering the importance of the blockchain-based search engines, in this section, conclusions are drawn, after which future research directions are provided.

7.1. Conclusions

Privacy and monopoly are the most critical problems of conventional search engines, which lead to several emerging alternatives such as p2p search systems. Considering the blockchain's characteristics and capabilities, this technology has the potential to solve these problems. With the blockchain-based, no company can lay claims to users' data or access their search history and other related information. All these data will be encrypted and stored on a blockchain. Blockchain-based can further provide users with a private key to protect their data entirely. In this paper, the blockchain applications and their characteristics are reviewed to show the blockchain potentials. Additionally, the blockchain-based search engines are studied through a comparative survey. Additionally, we elucidate that blockchain technology is an essential paradigm for the search ecosystem. To the best of our knowledge, no research article has been published in this regard thus far.

7.2. Future Research Directions

According to the future research directions, blockchain-based can provide context-aware and privacy-preserved search results in smart environments. As an emerging distributed computing paradigm in smart environments, edge computing attempts to bring computation and data storage to the edges to save network bandwidth and reduce the latency of requests. In this context, using the blockchain-based, encrypting the data, storing it on edge, and processing locally, user privacy will be provided. Additionally, using the edge processing power and storage capacity, the overall search performance will be improved.

In addition to the monopoly and privacy, other issues such as selective search space and limited object indexing are two important open issues of conventional. These search systems select a specific part of the web for indexing. Thus, many data with different origins in the deep web and dark web will be ignored. Additionally, they commonly index only the web pages and links and ignored other searchable contents such as businesses, customers, and contracts, which can be indexed to improve the search performance. Many open problems need to be reviewed and explored to develop more usable and efficient industrial applications. Alternate architectures, possible consensus algorithms, and integration with other systems are an example of many open issues. This survey is intended to be a helpful resource for learning more about various blockchain-based search engines.

Author Contributions: Conceptualization, E.R., A.M.S. and A.F.; methodology, E.R. and A.M.S.; writing, E.R., A.M.S. and A.F.; original draft, E.R. and A.M.S.; review and editing, E.R., A.M.S. and A.F.; Investigation E.R. and A.M.S.; paper writing management, E.R.; validation A.F.; visualization A.F.; supervision A.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Arjun, R.; Suprabha, K.R. Innovation and Challenges of Blockchain in Banking: A Scientometric View. *Int. J. Interact. Multimed. Artif. Intell.* **2020**, *6*, 7–14. [CrossRef]
2. Wu, J.; Dong, M.; Ota, K.; Li, J.; Yang, W. Application-aware consensus management for software-defined intelligent blockchain in IoT. *IEEE Netw.* **2020**, *34*, 69–75. [CrossRef]
3. Hasselgren, A.; Kralevska, K.; Gligoroski, D.; Pedersen, S.A.; Faxvaag, A. Blockchain in healthcare and health sciences—A scoping review. *Int. J. Med. Inf.* **2020**, *134*, 104040. [CrossRef]
4. Shrestha, B.; Halgamuge, M.N.; Treiblmaier, H. Using Blockchain for Online Multimedia Management: Characteristics of Existing Platforms. In *Blockchain and Distributed Ledger Technology Use Cases*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 289–303.
5. Aini, Q.; Lutfiani, N.; Santoso, N.P.L.; Sulistiawati, S.; Astriyani, E. Blockchain For Education Purpose: Essential Topology. *Aptisi Trans. Manag. ATM* **2021**, *5*, 112–120.
6. Vimal, S.; Srivatsa, S.K. A new cluster p2p file sharing system based on ipfs and blockchain technology. *J. Ambient Intell. Humaniz. Comput.* **2019**, 1–7. [CrossRef]
7. Barman, N.; Deepak, G.C.; Martini, M.G. Blockchain for Video Streaming: Opportunities, Challenges, and Open Issues. *Computer* **2020**, *53*, 45–56. [CrossRef]
8. Gai, K.; Guo, J.; Zhu, L.; Yu, S. Blockchain meets cloud computing: A survey. *IEEE Commun. Surv. Tutor.* **2020**, *22*, 2009–2030. [CrossRef]
9. Yang, R.; Yu, F.R.; Si, P.; Yang, Z.; Zhang, Y. Integrated blockchain and edge computing systems: A survey, some research issues and challenges. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 1508–1532. [CrossRef]
10. Mollah, M.B.; Zhao, J.; Niyato, D.; Lam, K.Y.; Zhang, X.; Ghias, A.M.; Koh, L.H.; Yang, L. Blockchain for future smart grid: A comprehensive survey. *IEEE Internet Things J.* **2020**, *8*, 18–43. [CrossRef]
11. Kavya, K.R.; Kavitha, M. Military Message Passing using Consortium Blockchain Technology. In Proceedings of the 2020 5th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 10–12 June 2020; pp. 1273–1278.
12. Tseng, L.; Yao, X.; Otoum, S.; Aloqaily, M.; Jararweh, Y. Blockchain-based database in an IoT environment: Challenges, opportunities, and analysis. *Clust. Comput.* **2020**, *23*, 2151–2165. [CrossRef]
13. Nebulas. Nebulas Technical Whitepaper. 2017. Available online: <https://nebulas.io/docs/NebulasTechnicalWhitepaper.pdf> (accessed on 29 July 2021).
14. Bitclave. Active Search Ecosystem. 2017. Available online: [https://cryptorating.eu/whitepapers/BitClave/BitClaveWhitePapernewdesign\(current\).pdf](https://cryptorating.eu/whitepapers/BitClave/BitClaveWhitePapernewdesign(current).pdf) (accessed on 29 July 2021).
15. Presearch. Presearch Whitepaper. 2017. Available online: <https://whitepaper.io/coin/presearch> (accessed on 29 July 2021).
16. DeSearch. DeSearch—Once Source for all Things Crypto. 2018. Available online: <https://medium.com/bitclave> (accessed on 29 July 2021).
17. Kaushik, A.; Choudhary, A.; Ektare, C.; Thomas, D.; Akram, S. Blockchain—literature survey. In Proceedings of the 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, 19–20 May 2017; pp. 2145–2148.
18. Zheng, Z.; Xie, S.; Dai, H.-N.; Chen, X.; Wang, H. Blockchain challenges and opportunities: A survey. *Int. J. Web Grid Serv.* **2018**, *14*, 352–375. [CrossRef]
19. Saghir, A.M. Blockchain Architecture. In *Advanced Applications of Blockchain Technology*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 161–176.
20. Berdik, D.; Otoum, S.; Schmidt, N.; Porter, D.; Jararweh, Y. A survey on blockchain for information systems management and security. *Inf. Process. Manag.* **2021**, *58*, 102397. [CrossRef]
21. Khatal, S.; Rane, J.; Patel, D.; Patel, P.; Busnel, Y. FileShare: A Blockchain and IPFS Framework for Secure File Sharing and Data Provenance. In *Advances in Machine Learning and Computational Intelligence*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 825–833.
22. Liu, M.; Teng, Y.; Yu, F.R.; Leung, V.C.; Song, M. A mobile edge computing (MEC)-enabled transcoding framework for blockchain-based video streaming. *IEEE Wirel. Commun.* **2020**, *27*, 81–87. [CrossRef]
23. Khalaf, O.I.; Abdulsahib, G.M.; Kasmaei, H.D.; Ogudo, K.A. A new algorithm on application of blockchain technology in live stream video transmissions and telecommunications. *Int. J. E-Collab. IJec* **2020**, *16*, 16–32. [CrossRef]
24. Yang, C.; Tan, L.; Shi, N.; Xu, B.; Cao, Y.; Yu, K. AuthPrivacyChain: A blockchain-based access control framework with privacy protection in cloud. *IEEE Access* **2020**, *8*, 70604–70615. [CrossRef]
25. Wei, P.; Wang, D.; Zhao, Y.; Tyagi, S.K.S.; Kumar, N. Blockchain data-based cloud data integrity protection mechanism. *Future Gener. Comput. Syst.* **2020**, *102*, 902–911. [CrossRef]
26. Ravishankar, B.; Kulkarni, P.; Vishnudas, M.V. Blockchain-based Database to Ensure Data Integrity in Cloud Computing Environments. In Proceedings of the 2020 International Conference on Mainstreaming Block Chain Implementation (ICOMBI), Bengaluru, India, 21–22 February 2020; pp. 1–4.
27. Shi, W.; Cao, J.; Zhang, Q.; Li, Y.; Xu, L. Edge computing: Vision and challenges. *IEEE Internet Things J.* **2016**, *3*, 637–646. [CrossRef]

28. Stanciu, A. Blockchain based distributed control system for edge computing. In Proceedings of the 2017 21st International Conference on Control. Systems and Computer Science (CSCS), Bucharest, Romania, 7 July 2017; pp. 667–671.
29. Xiong, Z.; Zhang, Y.; Niyato, D.; Wang, P.; Han, Z. When mobile blockchain meets edge computing. *IEEE Commun. J.* **2018**, *56*, 33–39. [\[CrossRef\]](#)
30. Gai, K.; Wu, Y.; Zhu, L.; Xu, L.; Zhang, Y. Permissioned blockchain and edge computing empowered privacy-preserving smart grid networks. *IEEE Internet Things J.* **2019**, *6*, 7992–8004. [\[CrossRef\]](#)
31. Wang, J.; Wu, L.; Choo, K.-K.R.; He, D. Blockchain-based anonymous authentication with key management for smart grid edge computing infrastructure. *IEEE Trans. Ind. Inform.* **2019**, *16*, 1984–1992. [\[CrossRef\]](#)
32. Yi, H. Securing instant messaging based on blockchain with machine learning. *Saf. Sci.* **2019**, *120*, 6–13. [\[CrossRef\]](#)
33. De Castro Oyama, P.I.; Ueyama, J.; Matias, P. EtherYou: Ethereum-based privacy-preserving social media platform. In Proceedings of the 3rd Blockchain Workshop: Theory, Technology, and Applications, Rio de Janeiro, Brazil, 10 May 2020; pp. 114–119.
34. McConaghy, T.; Marques, R.; Müller, A.; De Jonghe, D.; McConaghy, T.; McMullen, G.; Henderson, R.; Bellemare, S.; Granzotto, A. *Bigchaindb: A Scalable Blockchain Database*; White Pap. BigChainDB. Available online: <https://www.bigchaindb.com/whitepaper/> (accessed on 29 July 2021).
35. Gupta, S.; Rahnama, S.; Hellings, J.; Sadoghi, M. Resilientdb: Global scale resilient blockchain fabric. *arXiv* **2020**, arXiv:2002.00160. [\[CrossRef\]](#)
36. Nathan, S.; Govindarajan, C.; Saraf, A.; Sethi, M.; Jayachandran, P. Blockchain meets database: Design and implementation of a blockchain relational database. *arXiv* **2019**, arXiv:1903.01919. [\[CrossRef\]](#)
37. Huynh, T.T.; Nguyen, T.D.; Tan, H. A decentralized solution for web hosting. In Proceedings of the 2019 6th NAFOSTED Conference on Information and Computer Science (NICS), Hanoi, Vietnam, 12–13 December 2019; pp. 82–87.
38. Huynh, T.T.; Nguyen, T.D.; Nguyen, N.T.; Tan, H. Privacy-Preserving for Web Hosting. In Proceedings of the International Conference on Industrial Networks and Intelligent Systems, Hanoi, Vietnam, 27–28 August 2020; pp. 314–323.
39. Sudeepthi, G.; Anuradha, G.; Babu, M.S.P. A survey on semantic web search engine. *Int. J. Comput. Sci. Issues IJCSI* **2012**, *9*, 241.
40. Trotman, A.; Zhang, J. Future web growth and its consequences for web search architectures. *arXiv* **2013**, arXiv:1307.1179.
41. Baeza-Yates, R. Towards a distributed search engine. In Proceedings of the International Conference on Algorithms and Complexity, Rome, Italy, 26–28 May 2010; pp. 1–5.
42. Najork, M. Web Crawler Architecture. Available online: <http://research.microsoft.com/pubs/102936/EDSWebCrawlerArchitecture.pdf> (accessed on 29 July 2021).
43. Peshave, M.; Dezhgosha, K. How Search Engines Work: And a Web Crawler Application. Ph.D. Thesis, University of Illinois Springfield, Springfield, IL, USA, 2005.
44. Brin, S.; Page, L. Reprint of: The anatomy of a large-scale hypertextual web search engine. *Comput. Netw.* **2012**, *56*, 3825–3833. [\[CrossRef\]](#)
45. Shou, L.; Bai, H.; Chen, K.; Chen, G. Supporting privacy protection in personalized web search. *IEEE Trans. Knowl. Data Eng.* **2012**, *26*, 453–467. [\[CrossRef\]](#)
46. Castellà-Roca, J.; Viejo, A.; Herrera-Joancomartí, J. Preserving user’s privacy in web search engines. *Comput. Commun.* **2009**, *32*, 1541–1551. [\[CrossRef\]](#)
47. Amudha, S.; Phil, M. Web crawler for mining web data. *Int. Res. J. Eng. Technol.* **2017**, *3*, 128–136.
48. Patterson, M.R. Google and search engine market power. *Harv. J. Law Technol. Occas. Paper Ser.* **2013**. Available online: https://ir.lawnet.fordham.edu/faculty_scholarship/515/ (accessed on 29 July 2021).
49. Bradshaw, S. Disinformation optimised: Gaming search engine algorithms to amplify junk news. *Internet Policy Rev.* **2019**, *8*, 1–24. [\[CrossRef\]](#)
50. Sheu, T.-R.; Carley, K. Monopoly power on the web—A preliminary investigation of search engines. *arXiv* **2001**, arXiv:Cs0109054.
51. Search API | Search as a Service | SeekStorm.1. Available online: <https://seekstorm.com/> (accessed on 2 June 2021).
52. YaCy Project. 2017. Available online: <https://yacy.net/> (accessed on 29 July 2021).
53. Herrmann, M.; Ning, K.-C.; Diaz, C.; Preneel, B. Description of the YaCy distributed Web search engine. *Tech. Rep. KU Leuven ESATCOSIC IMinds*. **2014**.
54. Forestiero, A.; Papuzzo, G.; de Simone, R.; Forestiero, F.; Giordano, F. Multiagent approach for resource management in Smart Environments. In Proceedings of the 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 1–3 July 2020; pp. 1–5.
55. Diane, C.; Sajal, D. *Smart Environments: Technology, Protocols and Applications*; Wiley-Interscience: New York, NY, USA, 2004; pp. 153–174.
56. Preuveneers, D.; Tsingenopoulos, I.; Joosen, W. Resource usage and performance trade-offs for machine learning models in smart environments. *Sensors* **2020**, *20*, 1176. [\[CrossRef\]](#) [\[PubMed\]](#)
57. Matz, A.; Götz, C. Designing Human-Centered Interactions for Smart Environments Based on Heterogeneous, Interrelated Systems: A User Research Method for the ‘Age of Services’ (URSERVe). In Proceedings of the International Conference on Human-Computer Interaction, Copenhagen, Denmark, 19–24 July 2020; pp. 99–116.