Journal of Digital Science

ISSN 2686-8296

Volume 4 Issue 2 **December 2022**

CONTENTS

Cyber-Security Attacks, Prevention and Malware Detection Application 3
Darius Moldovan, Simona Riurean

Searching Algorithm in a nonrelational database 20
Roman Ceresnak, Michal Kvet, Karol Matiasko

System of Automatic Recognition of Video Text Amazigh based on the Random Forest 30
Youssef Rachidi

Investigating Different Social Media Platforms Used by Tourists to Book a Hotel in Greece 38
Olympia Vlachopoulou, Vasileios Paliktzoglou

Risk Disclosure as a Way to Increase the Informative Value of Corporate Reporting for Stakeholders 51
Irina V. Zenkina

Cyber-Security Attacks, Prevention and Malware Detection Application

Darius Moldovan [1][0000-0003-2262-3746], Simona Riurean [2][0000-0002-5283-6374]

[1] Bit Sentinel, București, Romania
[2] University of Petroșani, Petroșani, Romania

https://doi.org/10.33847/2686-8296.4.2_1

Received 28.10.2022/Revised 28.11.2022/Accepted 23.12.2022/Published 28.12.2022

Abstract. The internet has become more or less, for most of us a dangerous place to live, work and relax when no proper measures are taken, and the response to incidents is not very clear and well implemented, both for organizations and individuals. This paper makes a short overview of current types and incidents of cyber-attacks, as well as the current state of threats, and the grade of awareness worldwide. Some methods to prevent cyber-attacks, malware analysis, and threat hunting, are presented, too. The paper also contains an application developed with a series of APIs that link the application to open-source tools and activate them, hence analyzing the content of the possible malicious files.

Keywords: malware, ransomware, social engineering, phishing, crypto-jacking.

1. A Short Overview of Cyber-Attacks

The Internet network improves the efficiency of our daily activities, however, it also brings a lot of drawbacks, one of them being the lack of security, such as the possibility of being attacked or hacked while operating online.

Some of the major threats, that lately bring a great deal of loss (financial, image, time, and so on), both upon persons and companies under attack, are: ransomware, malware, social engineering threats, threats against data, threats against availability (denial of service), crypto-jacking, threats against availability (internet threats), disinformation/misinformation, fake news, and supply-chain attacks [1].

Ransomware cyber-attacks can take a variety of forms, infiltrate victim systems in various ways, but they are all based on the same fundamental principle. After the attack, the victims are not able to access their own data (that are encrypted by the attacker) until they pay the required fee to the attacker. Usually, the attackers do not return data after the payment is made. After the first payment, there is a very high chance that the next time the attackers ask for an even higher amount. Even though a number of security mechanisms, such as firewalls, anti-virus programs, and automated analysis programs, have been developed to fight against this threat, in most of cases the current mechanisms are not able to guard data stored in local or cloud storage resources [2].

Phishing, as part of the social engineering cyberattack, has evolved (since 1995 when the first instance of this technique was reported, when attackers convinced victims to share their AOL account details [3]) into one of the most widespread and malicious forms of cybercrime in the world. The attack occurs when an attacker poses as an official entity to trick their targeted victim to divulging personal information (in an attempt to obtain sensitive information such as usernames, passwords, and credit card details, and money), often for malicious reasons, installing malware, or visiting a website that hosts malware [4].

A DDoS (distributed denial-of-service) attack is an attempt to interrupt the normal traffic of a server, service or network, by overwhelming it or the computing infrastructure, with a flow of traffic. Denial-of-Service (DoS) and DDoS attacks are serious threats both on local and cloud services' availability, due to numerous novel vulnerabilities (especially in cloud, where multi-tenancy and resource sharing are available) [5]. Attackers changed their attack format over the years, damaging operating systems and protocols in an attempt to deny or reduce the quality of the service provided to valid users. Today, attacks are sneakier and impersonate legitimate user in such a way that detection mechanisms of traffic against high – rate DoS attacks are no more appropriate. The LDoS (Low-rate Denial of Service) attack, has the potential to produce more damage than its predecessor due to its advanced nature and the lack of appropriate recognition and protection means. A most recent taxonomy divides DoS attacks in QoS attacks, Slow rate attacks, and Service queue attacks [6].

Crypto-jacking is a hazardous attack because it is silent and well-hidden. The victim has no clue that the malware is installed on his/her own device. Nothing happens and the victim still has access to their own device and data. The malware don't mine personal data, compromise files, ask for rewards, or crash computers. The purpose of crypto-jacking is to use the victim's computer's resources to create virtual currency. The only thing that gives evidence to victims about being under attack is the higher power consumption of the device. There are two main types of cryptojacking attacks; one requires a malicious payload to be installed on the user's computer and the other runs inside the user's browser upon visiting dubious web sites. More advanced methods exploit unpatched vulnerabilities, often zero-days to bypass the user entirely and install the payload. [7].

Powerful organizations spread fake news and a very large volume of disinformation through social networks and organizations-run media outlets, especially since current defenses must keep up with an increasing volume of Zero-Day types attacks [8].

The supply-chain attacks (targeting software or hardware) aim to damage an organization by pointing to less secure elements in the supply chain [9,10]. It can be launched towards any organization from the financial sector, oil industry, to the government sector. The attackers usually interfere within the manufacturing or distribution division of a product by compromising software (build tools or updated infrastructure), by steeling code-sign certificates or signed malicious apps using the identity of dev company, by compromising specialized code shipped into hardware or firmware components, or by pre-installing malware on devices (cameras, USB, phones, etc) [11, 12].

A software supply chain attack occurs when a third-party software dependency used in multiple 'downstream' applications is compromised. By compromising a single open-source package or library, attackers steal confidential data, cause a denial of service, or breach networks at thousands of organizations. This attack vector has become increasingly common, once the "Sunburst" attack in 2020 became widespread [13]. SUNBURST is a massive, fifth-generation cyber-attack, waged against US government agencies and technology companies. The attack compromised systems in over 40 government agencies, including the National Nuclear Security Administration (NNSA - the US agency responsible for nuclear weapons) and additional targets in other countries, including Canada, Belgium, Britain, and Israel, were also hit. The attacker hides a Trojan in a software update of the SolarWinds Orion software, and pushed this update to 18,000 customers, including almost all Fortune 500 companies, government agencies, and contractors including Lockheed Martin. They only discovered the attack in December, 2020, eight months after the original breach [14].

The most frequent attacks in Romania during 2022, have been ransomware, phishing, DDoS and others, as seen in Fig. 1.

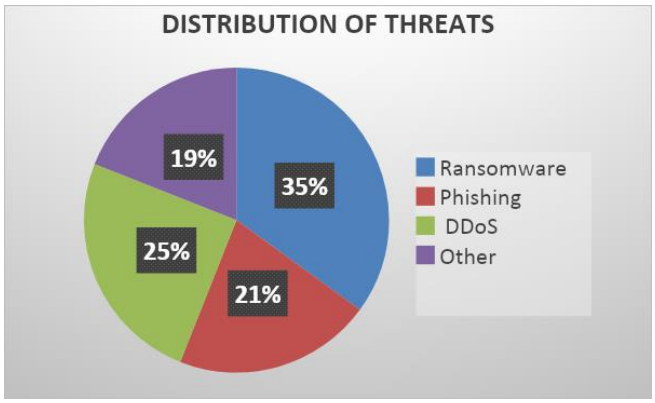

Fig. 1. Distribution of threats in Romania during 2022, by type [15]

1.1. Current Status Regarding Threats and Awareness Worldwide

One of the most distressing personal events that can result from being online is identity theft. According to Norton Cyber Safety Insights Report [16], conducted approximately 12 months (until February 2021), 55 million people have had their identity stolen. The Report that follows a survey conducted by The Harris Poll among 10,030 adults (aged 18 +) in 10 countries, shows that over 55% of internet users do not know what to do if their identity was stolen. Additionally, 60% of those surveyed admit to being "very worried" their identity will be stolen. According to the same report, people in India and the US are more likely to be taking more precautions online, while those in Japan are most likely to struggle with deciphering if the information is from a credible source.

What is more intriguing, the report shows that more than half of adults are more worried than ever about being the victim of cybercrime, but a similar proportion doesn't know how to protect themselves from it. More than 475 million consumers have ever been the victim of a cybercrime, nearly 330 million in 2021 alone [17].

As a response to those alarming results of the report presented above, the Norton Identity Advisor (as well as Norton™ Identity Advisor Plus [18] addresses this concern, by helping consumers each step of the way when they discover they're a victim of identity theft. It has an easy-to-use dashboard to register personal information for monitoring, including Social Media Monitoring that aim to monitor and also notify the account holder if there are signs the account is compromised or if potentially risky links are found. In case that the customer suspects an identity theft, a dedicated Identity Restoration specialist is available [19].

Any of the internet users can be a potential victim of malicious attacks, but recent research shows that small businesses are one of the most popular targets. In fact, small businesses are the target of approximately 43% of all cyber-attacks.

According to the public reports from Orange Business Services [20], the National Security Agency/Central Security Service [21] and the Cyber Security Agency of the European Union (ENISA) [22] there is published a relevant statistic for the year 2022 with the most common cyber-attacks in the online environment, concluding that "The year 2022 we all came under attack".

Hence, ransomware attacks represent approximately 31% of the total and target critical infrastructures in the HIPAA (Health Insurance Portability and Accountability Act) environment, i.e., the medical environment, also the private and public environment where personal data is encrypted or sold in "Black Market" [23].

Identifying cyber-attacks in most cases takes days, if not weeks or even months. As a result, small and medium-sized businesses must overcome these problems, therefore being aware of the most important cyber security dangers and knowing preventive measures that must be followed to reduce the risk of an incident, are compulsory for these companies.

1.2. Some Examples of Recent Cyber-Attacks Worldwide

A well-known case of ransomware attack was the attack on Colonial Pipeline, on May 7, 2021 where a ransom of around $5 million was paid [24] to regain access to files and data that have been encrypted. This company was the target of a group called DarkSide [25] They used this type of attack to steal 100 GB of data in less than 2 hours. This data includes payment data and confidential information held by the company [26].

In January, 2022 Crypto.com was hacked with some 500 wallets targeted. The malicious actors used a 2FA authentication attack to gain access. The hackers stole 33$ Million in cryptocurrency. The total value of the unauthorized withdrawals was 4,836.26 ETH and 443.93 BTC — equivalent to roughly $15.2 million and $18.6 million respectively, at current exchange rates — as well as $66,200 worth of other currencies [27].

In order to exploit the vulnerability regarding a breach (a Multiple Authentication - MFA fatigue), one of the biggest attacks was over the Uber company, in September, 2022. Although the attack on Uber was disruptive to their internal systems, no user data was compromised as part of the hack. The hackers initially gained access to Uber's Slack messaging service and from there moved to the internal databases and then attackers gained access to the Uber Google Cloud account of Uber and the Uber Amazon Web Services (AWS) account. Although Uber got off easy from the attack, it should serve as a stern reminder that the human element is often the weakest link in security defenses. In the Uber attack, the victim was part of the incident response team and likely had some administrative privileges. Through these privileges, the attacker gained access to a file with other credentials giving them essentially the keys to the kingdom [28].

This attack takes advantage of users whose login credentials have been compromised, bombarding them with authorization requests until they give in and approve one. An MFA attack is based on the fact that people have trust in this procedure as being a protective one. When users try to connect to MFA-protected resources, they usually receive a push notification or a code to confirm their credentials. Users respond to these alerts believing that they are being granted permission to access resources. Since this procedure exploits users' trust, they do not expect their own organization's MFA platform to betray them. This caused one of the biggest data leaks and mass layoffs.

One of the major breaches in Thailand's history, records of 39 million patient from Bangkok's Siriraj Hospital have been offered for sale on a dark web forum. The attacker followed up by posting on raidforums.com that goes under the name of "WraithMax" offered to sell the data and supply a sample file via Telegram. The poster claims the data includes names, addresses, Thai IDs, phone numbers, gender details, dates of birth and other information [29].

A „very elaborated" attack compromised The International Committee of the Red Cross servers, attackers compromised data of more than 515.000 "highly

vulnerable persons" including those separated from their families due to conflict, migration and disaster, missing persons and their families, and people in detention [30].

The war in Ukraine (that started on 24th of February, 2022) significantly changed the threat scenery in 2022. There have been noticed important changes in hacktivist activity, cyber actors conducting operations along with kinetic military action, the mobilization of hacktivists, cyber-crime, and aid by nation-state groups during this conflict [31].

The Orange Business Internet Security Report [15] publishes a very interesting report regarding various cyber-attacks worldwide, during 2022, that seriously affected organizations in several regions worldwide and a number of professional areas.

All these attacks brought into focus the importance of insider threats monitoring and security.

At the beginning of the year, in Germany, two oil supply companies said they were victims of the cyberattack since Saturday January 29. Both companies declared force majeure. At the beginning of February, Belgian prosecutors launched an investigation into the hacking of oil facilities in the country's maritime entryways, including Antwerp, Europe's second biggest port after Rotterdam.

In Germany, prosecutors said they were investigating a cyberattack targeting oil facilities in what was described as a possible ransomware strike, in which hackers demand money to reopen hijacked networks [32].

In February 2022, (the 7th), Swissport, an important aviation operations company (that provides services to many airports across Europe) was a victim of ransomware in an attack that caused flight disruptions for at least 22 scheduled flights [32].

In February (the 8th) this year, Vodafone Portugal was the victim of a deliberate and malicious cyberattack intended to cause damage and disruption. The company's 4G and 5G mobile networks, along with fixed voice, television, SMS, and voice/digital answering services were all offline following the attack [33].

One of the world's biggest private banks, Credit Suisse was the victim of a data leak, with confidential information on more than 18.000 bank accounts, gathering some 100$ Billion in wealth, was leaked to a German Newspaper by a whistle-blower [34].

Hackers leaked a large collection of data exfiltrated from Samsung Group's Systems, including source code of applets and system components in use in sensitive environments. On 7th of March, 2022, Samsung today confirmed a breach of its systems, reportedly the work of hacking gang Lapsus$, which saw 190GB of the South Korean electronics company's data, including source code for its Galaxy devices, leaked online. The attack came days after Lapsus$ breached another Big Tech business, chipmaker Nvidia. While both incidents appear to have been mercenary in nature, security researchers believe the gang could be pursuing another agenda too [35].

France's Caisse Nationale D'assurance Maladie (CNAM) health insurance body, made a formal complaint, explained that, in March 17th, 2022, "unauthorized people" had connected to the "Amelipro accounts" of the healthcare workers whose "email addresses had been compromised". It shows that the accounts of 19 healthcare staff had been hacked, causing the details (names, surnames, date of birth, social security numbers, GP details) of at least 510.000 people to be stolen [36].

A brute-force DDoS attack over Finland's Ministry of Defense's website was launched during April 2022. The attack caused minor availability issues for the web portals, and the operators managed to restore their websites in short time [37].

A dataset containing user data for more than 21 million users of several VPN services, was leaked on Telegram. The data contains names, usernames, hashed passwords and e-mail addresses of GeckoVPN, SuperVPN and ChatVPN clients [38].

In May, 2022, a German library was crippled by ransomware. Customers were unable to rent audio books, digital copies of magazines, e-books, nor were able to make requests online or by phone. The attack severely affected Onleihe, a popular app that connects users via EKZ's service to local libraries German Library Service crippled by ransomware, in an attack orchestrated by the Lockbit ransomware group, who then published the data they exfiltrated during the attack, claiming their ransom requests have not been met [39].

The NFT marketplace OpenSea, had a data breach after an employee of the company's e-mail delivery vendor, misused their employee access to download and share email addresses provided by OpenSea users with an unauthorized external party" [40].

The Pegasus Airline company used an unsecured AWS Bucket (bucket is a container for objects stored), and exposed 6.5 TB of data consisting of personal information of flight crews. The EFB bucket was misconfigured to allow open access from the internet [41].

In May 2022, exploiting a misconfiguration, a triple ransomware attacked an automotive supplier, exposing RDP through a border firewall. LockBit, Hive and BlackCat – the three culprits – have encrypted files (some files were encrypted at least 3 times each), the attack lasted for more than 2 weeks [42].

In June, the 1st, 2022, Costa Rica's Public Health services went offline after ransomware attack. The Hive Ransomware was the culprit, with the initial breach happening some 3 days before the report was published. The employees of the Public Health agencies targeted by the ransomware were instructed to "unplug their computers" in order to prevent the malware from spreading through their networks. [43]

Using a Zero-Day Vulnerability in the software stack, Twitter was also breached in July, 2022. That breach allowed the attackers to associate usernames with e-mail addresses and registered phone numbers. The hackers generated a dataset of more than 5.4 million affected user profiles [44].

The virtual pet website Neopets has suffered multiple data breaches since its inception and transfer from Viacom to JumpStart Games in 2014, 2016, and 2020. On 27th of July, 2002, suffered a data breach exposing a database containing personal information (players' names, gender, dates of birth, usernames, email addresses, IPs, countries, and zip codes) of 69 million users. The website reported that the attackers exfiltrated source code of its software products [45].

In August, the 12th, 2022, a hacker obtained the personal information of 48.5 million users of a COVID health mobile app run by the city of Shanghai. This was the second claim of a breach of the Chinese financial hub's data in just over a month. The hacker with the username as "XJP" also posted an offer to sell the data for $4,000 on Breach Forums [46].

According to Romanian National Directorate of Cyber Security and Response to Incidents (DNSC) that released a report regarding cybersecurity attacks in Romania, a series of DDoS attacks targeting the following Romanian government websites [47] have been developed this year:

- gov.ro (the official website of the Government of Romania);
- mapn.ro (the official website of the Ministry of National Defense of Romania);
- politiadefrontiera.ro (the official website of the Border Police);
- cfrcalatori.ro (the official website of the Romanian railways);
- otpbank.ro (the official website of OTP Bank).

Although the examples above are far from being close to the real number of different cyberattacks during 2022 only, it shows that no organization is safe enough online these days and can become the victim of any kind of malicious cyber-attack. So, as many cyber-security measures and personalized procedures are implemented in organizations, the safer their data are.

Hence, no matter the model of business we are involved in, or technology we use [48-50] is essential to be aware of treats and follow the experts' advice and the advanced tools available against cyber- attacks.

1.3. Methods to Prevent Cyber-Attacks

Cyber-attacks are now the fastest growing crime on a global scale. Since the volume and effects of cyber threats continue to accelerate, it has never been more important for individuals and organizations to address emerging threats and to mitigate possible troubles. With a great certainty of the cyber-attacks to come, the need to take a more forward-thinking attitude is compulsory. The Future of Cyber Security aims to help businesses to stay one step ahead of cyber attackers through a number of insightful sessions [51].

Protecting critical assets or networks from disruption or attack is one of the primary concerns in both cyber-security and risk analysis. When these two fields intersect the concept of cyber resilience is born, which can simply be defined as an entity's ability to plan for, absorb, recover from, and successfully adapt in the face of adverse cyber events [52].

According to the standards imposed by NIST (National Institute of Standards and Technology) and ISO/IEC 207001 (International Organization for Standardization and the International Electrotechnical Commission), there are some good practices useful both to people who use an IT system and to organizations to prevent potential cyber-attacks:

- The use of a password manager to create and retain unique and complex passwords for each account;
- To implement two-step authentication for the online accounts whenever possible. This option is offered now by banks, social media platforms, e-commerce platforms and so on. Two-step authentication involves two steps, (as the name suggests) entering the password and then, a unique access code that is received on a different device (the phone, for example;
- To protect the internet browsing activity as every internet action is tracked by businesses and websites. The location, browsing history, and other data are collected by every ad, social button, and website. The data collected reveals more about personal identity than might be expected. Is possible that the websites visited are regularly providing all the data advertisers need to identify the type of person/customer/audience the client is. This is part of how targeted advertising remains one of the Internet's most disturbing innovations;
- To use antivirus software on the computer and keep it up to date. As an organization, it is recommended to perform regular security audits (vulnerability assessment and penetration testing, etc.);
- To use protection technologies such as firewalls, cloud WAF services, periodic backups, and monitoring agents connected to a Security Operation Center (SOC);
- It is recommended for organizations to have well-established procedures regarding "Risk Management" [53].

There are some important procedures and steps to be followed by security responsible in organizations to avoid such disturbing situations (ransomware attacks) that the organizations or individuals can face.

Therefore, a specific procedure for a regular backup of all personal data and data related to the company's activities, is compulsory to be defined and followed, especially to avoid any loss in case of a ransomware event. A periodic audit to prevent attack vectors has to be performed in companies. A cyber risk awareness program has to be developed and employees must be trained accordingly.

The Internal Security Rule Planning of the organization must restrict (and even forbid in some special situations) the use of peripheral devices (or external storage supports) as much as possible. An organization's cyber-security responsible must stay up to date with the latest news in the field and in the event of a cyber-attack report the situation to official entities that can assist and help, before making any payment to the attackers or implementing a protection system.

To protect the organization from phishing attacks, the following actions is compulsory to be implemented:

- One of the most important steps that can be taken to defend an organization against this type of attack, is to train employees and establish certain standards, and develop periodic simulations;
- Employees should be given proper and periodic training so they recognize various phishing patterns and strategies;
- The awareness of the employees/future possible victims, to identify what is malicious in the online environment and what is not, is the greatest resource of prevention to avoid such attacks becoming efficient for attackers. It all depends on how well-prepared and informed are the targeted victims in the online environment and the dangers that exist in this environment.

Mitigating a multi-vector DDoS attack requires a variety of strategies to counter the different trajectories. In order to protect organization against Distributed Denial of Service attacks, the following actions are compulsory:

- The first step, as a compulsory act, is to check that the service provider is prepared for an overload of allocation resources;
- To monitor the DoS and DDoS attacks and test equipment against such an attack (Stress Testing);
- To reduce the attack surface, including the exposure of the ports and protocols used in the Internet, as little as possible;
- To use a cloud service that allows quick access to resources and back-up restoration in record time.

2. Malware Analysis and Threat Hunting

2.1. Malware Analysis Procedure

In order to analyze or identify malware, a study on the infected system is required first. The best method against these attacks is based on "The Cyber Kill Chain" [54] when the Security Engineer needs to think like a bad intruder.

The Cyber Kill Chain contains few steps to be followed:

- Reconnaissance is the first step of investigation as part of a malicious attack when data about the target is gathered, such as what type of technology is used, the e-mail addresses, user IDs, physical locations, software applications, and operating system details, and any other kind of information that might be useful in phishing or spoofing attack;
- Weaponization is the second step that an attacker would follow to create the malware, virus, or worm that can exploit a known vulnerability;
- Delivery is the third step of the process when the intruder launches the attack;

- Exploitation is the phase when the malicious code is executed within the victim's system;
- Installation procedure follow when malware or other attack vectors will be installed on the victim's system like rootkit or backdoor;
- Command and Control is the phase when the attacker is able to use the malware to assume remote control of a device or identity within the target network;
- Actions based on Objective is the final phase when the attacker reaches the objective, naming data theft, destruction, encryption or exfiltration.

Although difficult to accept, every cyber-attack leaves a trace. These traces in technical terms are called "Indicators of Compromise (IOC)" [55]. Based on the procedures mentioned above, an analyst starts from the initial problem and looks for similarities such as: where the malware connects, what files are modified in the system, what new processes have been developed in the system, what applications the malware tries to install, etc.

For a better prevention procedure, a Security Operations Center has well-defined plans to be aware of how to act according to each individual attack's particularities. Under the given conditions, the analyst follows the analysis phases described in PICERL (Prepare, Identify, Contain, Eradicate, Recover, and Lessons Learned) [56].

PICERL (Fig. 2) is a plan recommended by all security training institutions regarding the management of activities in the event of an incident:

- Preparation - refers to the capability of the cyber-security team to respond quickly in case of an incident by knowing all the procedures in case of a cyber security attack and having all the tools for action. They have full access (all the rights) in the network to complete the investigation and ensure that no law interferes with the activity of the analyst to hinder or stop him. Also, the continuous education that analyst takes in case of exceptional events is important and must be provided by the company or by an up-to-date, specialized organization;
- Identification - follows the analysis of the incident, the analyst must find out what happened, when it happened, if the user was compromised, what changes were created on the system, what internet connections were made, what he download-ed, where he made the persistence and what subsequent behavior it may have if it tried to connect to other stations in the network with the aim of compromising them. From this perspective, being only a phishing analysis, the focus is strictly on the analysis of metadata from the mail as well as attachments or added URLs;
- Containment is the phase when steps are taken with the aim to stop the spread of the infection. Domains used in case of infection will be blocked, and hashes of malware samples used to infect systems are also blocked;
- Eradication follows when the analysis has already been completed, and attempts are made to eliminate all systems that have been compromised. In the case of phishing e-mails, the victims are asked to delete them from the mail servers;
- Recovery phase is considered when the incident is over and attempts are made to bring all the stations that were isolated upstate;
- Lessons learned is maybe the most important step of the PICERL procedure. In this state, the incident is resolved and the conclusions drawn must prevent future infections. The discussion here can vary from applied patches, redone configurations, or extra rights for analysts.

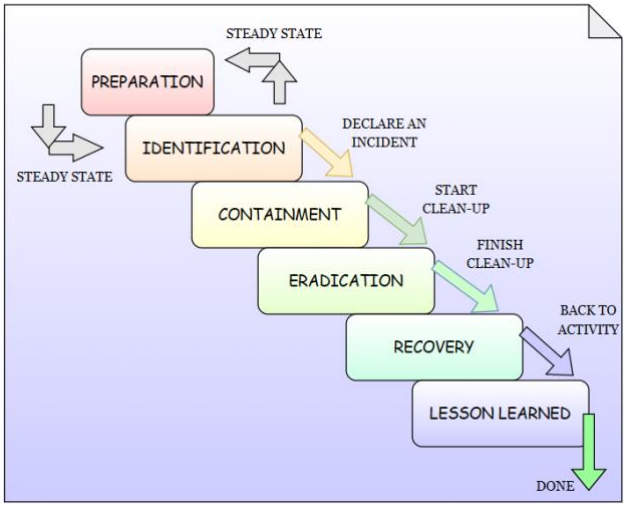

Fig. 2. PICERL Phases of incident response [57]

Malware analysis is extremely important as it helps individuals as well as organizations to identify if a specific file is malicious or not. It provides all information regarding the origin of the file, the processes behind it, and its capabilities and can help identify if the file is legit.

Based on the incident management model, we refer here to all the above-mentioned phases, less the identification step involving e-mail analysis. Regardless of the tools used, the process is the same, following certain steps in extracting information and making a decision if the file or URLs are malicious or not:

- The first step, a look at the body of the e-mail is necessary to analyze the message, to see if the one who sent the e-mail is trying to convince the recipient to take a certain action for his own benefit. Most of the time attackers use social engineering techniques, taking advantage of human weaknesses such as curiosity, listening spirit, etc. The URLs and attachments sent with the e-mail in question will also be extracted, and will later be analyzed.
- The second step is to analyze the email header since it is necessary to look for the following artifacts: Message-ID and Hops. The Message ID is a field that provides a unique identifier of the message that refers to a certain particular version of a message. Hops–represents the route followed by the mail until it reaches the destination mail server.

Thus, it is represented by the 'Received' field. Forefront Antispam Report Header – after an antispam solution has scanned the mail that was sent, inserts the 'Forefront Antispam' field containing additional information about the mail and how it was processed such as country of origin of the message, language in who wrote the message, and other reports on security measures plus scores that allowed the mail to pass. Authentication results–Sender Policy Framework (SPF), Domain Keys Identified Mail (DKIM), Domain-based Message Authentication, Reporting & Conformance (DMARC) [58].

SPF is an authentication protocol that lists IP addresses in a DNS TXT record that are authorized to send e-mails on behalf of domains.

DKIM is an ID or passport that can verify identity. When it is sent from an e-mail server, the server attaches DKIM so the receiving server can verify. Therefore, DKIM authentication provides a method for validating a domain's identity that is associated with a message through cryptographic authentication. It does this by using

an encrypted key pair (one public in DNS and one private) to add a digital signature to every email message. Receiving email servers use this DKIM signature to both validate the authenticity of the sender, and detect if the message was altered/changed during transit. DKIM-signed messages provide Mailbox Providers (MBPs) with trust that the message is authentic and is not being spoofed. MBPs' internal filtering algorithms use SPF and DKIM along with other factors to determine if an email should be placed in the inbox, or spam folder, or be rejected. However, both SPF and DKIM don't allow domain owners to instruct MBPs how to treat a message if the authentication checks can't be validated. To help tell MBPs to know what to do if DKIM and/or SPF fail, senders can implement DMARC. DMARC leverages both SPF and DKIM and provides instructions from the domain owner about what to do with unauthenticated email [59].

DMARC is an email authentication, policy, and reporting protocol. It helps domains address domain spoofing and phishing attacks by preventing unauthorized use of the domain in the Friendly-From address of email messages.

DMARC allows the domain owner to specify how unauthenticated messages should be treated by MBPs. This is accomplished by what is known as a "policy" that is set in the DMARC DNS record. The policy can be set to one of three options: NONE, QUARANTINE, and REJECT.

- Policy = (p=none): no action and message delivered as normal;
- Policy = (p=quarantine): places the message to spam/junk/quarantine folder;
- Policy = (p=reject): the message rejected/bounced.

The R in DMARC is for the Reporting component of the protocol. These reports allow the domain owner to see where all email using their domain in the form address is being sent from [59].

The recipient, the sender, and the subject of the mail are also identified. At this point, a conclusion can be drawn in case there is an inconsistency between the servers the mail originated from and what the message is, the sender, and what the attacker is trying to trick the victim into doing.

This is the step of the procedure when the analysis must be focused on attachments. The analysis can be statically or dynamically according to certain well-defined malware analysis rules. Her free attachment or URL analysis tools can be used. The necessary step to be made is the analysis of the URLs in question, looking for indicators of compromise: HTML 'href' tags or javascript 'iframes'.

At the same time, what the page displays is also important to draw a conclusion in case an URL is in the mail or a malicious attachment.

2.2 Customize Personal Application for Threat Hunting

A "Thread Hunting" application must be dedicated mainly for the purpose of identifying potentially malicious content. The application must be easy to use for those who analyze bad threads. Another requirement that an application must meet is scalability. This is most important to keep the product on the market and to be able to be upgraded with the latest detection methods.

The application presented in our demo example has the possibility to implement open-source tools and execute their detection rules. Each function has been containerized with docker in order to provide both the flexibility to be used on any type of platform and to be easily modified by adding new functions or removing existing detection functions.

Based on a containerization model, the application can be easily changed (scalability) so that it can be used either as a file filter or as a plugin for customers.

The application, regardless of the model chosen for use, receives as input a file with the extension of word *.doc, or *.exe, or *.elf, etc.

The application analyzes the content of the malicious file using the ClamAV Scan [60] utility to get as many suspicious indicators about the nature of the file, from the analysis of the headers to the analysis of Windows_API.

This will give the analyst an overview of what is being analyzed in case there are specific indicators.

In Fig. 3 is presented a part of the code that is incorporated in the software and uses an open-source tool like clamav.

```
using System.Text.RegularExpressions;
using MalwareMultiScan.Backends.Backends.Abstracts;
using MalwareMultiScan.Backends.Services.Interfaces;

namespace MalwareMultiScan.Backends.Backends
{
    /// <inheritdoc />
    public class ClamavScanBackend : AbstractLocalProcessScanBackend
    {
        /// <inheritdoc />
        public ClamavScanBackend(IProcessRunner processRunner) : base(processRunner)
        {
        }

        /// <inheritdoc />
        public override string Id { get; } = "clamav";

        /// <inheritdoc />
        protected override string BackendPath { get; } = "/usr/bin/clamdscan";

        /// <inheritdoc />
        protected override Regex MatchRegex { get; } =
            new Regex(@"(\S+): (?<threat>[\S]+) FOUND", RegexOptions.Compiled | RegexOptions.Multiline);

        /// <inheritdoc />
        protected override bool ThrowOnNonZeroExitCode { get; } = false;

        /// <inheritdoc />
        protected override string GetBackendArguments(string path)
        {
            return $"-m --fdpass --no-summary {path}";
        }
    }
}
```

Fig. 3. Example of code

The application has implemented several scanners such as: Comodo (comodo is a tool based on Endpoint Detection and Response) [61], DrWebScan (similar website like virustotal), DummyScan (based on Yara rules scanner), KesScan (tool by Kaspersky), McAfeeScan (Antivirus) SophosScan and WindowsDefender (solution for detection by Microsoft).

In Fig. 4 is open-source Comodo (EDR) integration.

```
using System.Text.RegularExpressions;
using MalwareMultiScan.Backends.Backends.Abstracts;
using MalwareMultiScan.Backends.Services.Interfaces;

namespace MalwareMultiScan.Backends.Backends
{
    /// <inheritdoc />
    public class ComodoScanBackend : AbstractLocalProcessScanBackend
    {
        /// <inheritdoc />
        public ComodoScanBackend(IProcessRunner processRunner) : base(processRunner)
        {
        }

        /// <inheritdoc />
        public override string Id { get; } = "comodo";

        /// <inheritdoc />
        protected override string BackendPath { get; } = "/opt/COMODO/cmdscan";

        /// <inheritdoc />
        protected override Regex MatchRegex { get; } =
            new Regex(@".* ---> Found Virus, Malware Name is (?<threat>.*)",
                RegexOptions.Compiled | RegexOptions.Multiline);

        /// <inheritdoc />
        protected override string GetBackendArguments(string path)
        {
            return $"-v -s {path}";
        }
    }
}
```

Fig. 4. An open-source Comodo (EDR) integration

Using open-source tools as well as open threat intelligence feeds, we can categorize the file as malicious or not based on the results after the analysis.[62-67]

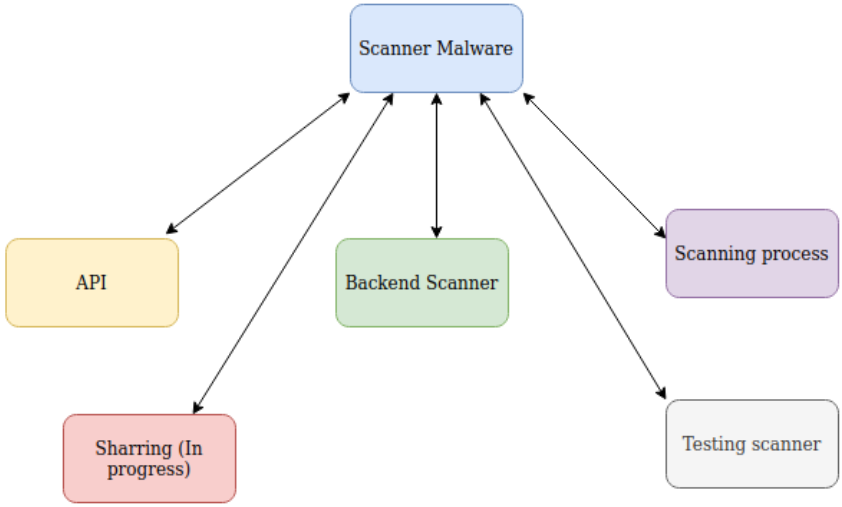

Fig. 5. The diagram of software

To create this application, we used what the mentioned above methodology, "The Cyber Kill Chain" [68]:

1. Reconnaissance: In this stage, the attacker/intruder chooses their target. Then they conduct in-depth research on the target to see specifics and to identify available vulnerabilities that can be exploited.
2. Weaponization: In this step, the intruder creates a malware weapon like a virus, worm or such in order to exploit the vulnerabilities of the target. Depending on the target and the goal of the attacker, this malware can exploit new, undetected vulnerabilities (also known as the zero-day exploits) or it can focus on a combination of different vulnerabilities.
3. Delivery: This step involves sending the weapon to the target. The intruder/attacker can use different methods like USB drives, e-mail attachments and websites for reaching his purpose.
4. Exploitation: In this step, the malware activity starts. The program code of the malware is triggered to exploit the target's vulnerability/vulnerabilities.
5. Installation: In this step, the malware installs an access point for the intruder/attacker. This access point is also known as the backdoor.
6. Command and Control: The malware gives the intruder/attacker access in the network/system.
7. Actions on Objective: Once the attacker/intruder gains persistent access, they finally take action to fulfill their purpose, such as encryption for ransom, data exfiltration or even data destruction.

For better detection/evasion techniques it is important to examine the components of the operating system and how different tools interact with those components.

In case that a detection is too sensitive, then the monitoring procedure takes too much because the team is flooded with false positives and the analysts waste time or the potential burn out.

On the other hand, if the detection rule is too specific, then evasion becomes trivial to achieve for the attacker.

Therefore, an evasion engineer's goal is to conduct their operation while avoiding preventative and potentially detective controls. In order to do this, it is important to understand what aspects of the attack the attacker has control of.

In the following demo. we are uploading a malicious file like malware but keep in mind examples presented here are not to harm any informatic system.

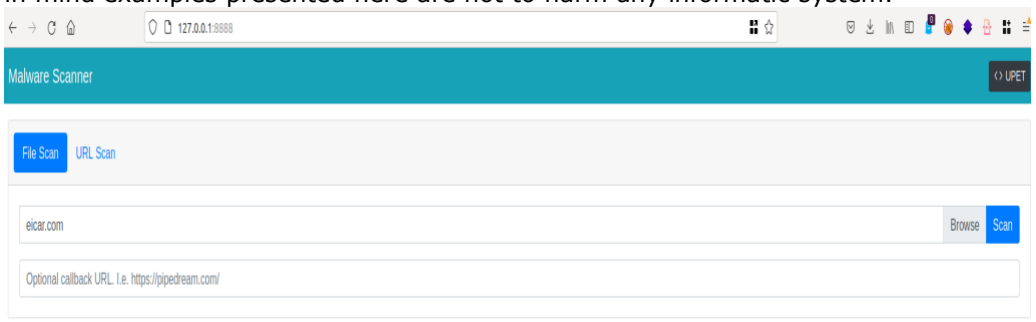

Fig. 6. Upload malicious file

The application works with a series of APIs that link the application to open-source tools and activate them. Therefore, the executable uploaded to the platform is analyzed based on a public database. Once the application hash is verified, it shows if the application is malicious or not.

But that's not all, there's also a sandbox behind it that tries to detonate the executable to see how it responds. In Fig. 7, the result can be seen [60-68].

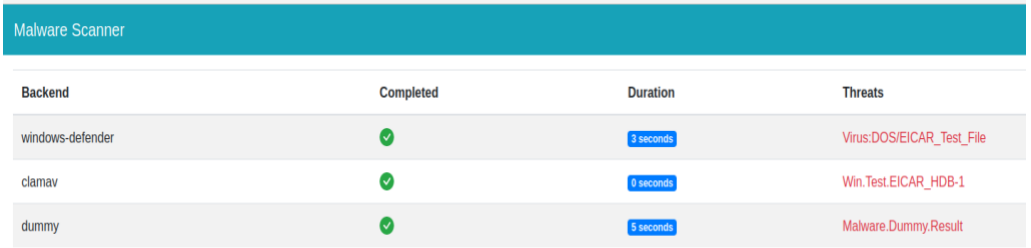

Fig. 7. Result based on the custom scanner

3. Conclusions

In this paper, we took a broad overview of malicious attacks during this year, the results of these attacks, and the preventive measures to avoid unwanted situations and their results following cyber-attacks. We present an application that should assist enthusiasts in malware analysis to reach a quick and efficient conclusion on attachments or on the malware spread vector, or to protect companies that use certain services both locally and remotely. We designed the application to be used from the command line as a stand-alone tool where an executable file of *.exe or *.elf types are given as input. The application developed here solves the compatibility problem of operating systems based on Linux or Windows with the help of the Docker engine. The main purpose of this application is to facilitate the analysis of executable files. The application is also extremely useful because both it and the technologies used are open sources, which allows further improvements in time.

References

1. European Union Agency for Cybersecurity, ENISA Threat Landscape NOVEMBER 2022, ISBN: 978-92-9204-588-3, DOI: 10.2824/764318.
2. Ilker Kara, Murat Aydos., The rise of ransomware: Forensic analysis for windows-based ransomware attacks, Expert Systems with Applications, Volume 190, 2022, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2021.116198.
3. Jakobsson, M.; Myers, S. Phishing and Countermeasures: Understanding the Increasing Problem of Electronic Identity Theft; Wiley: Hoboken, NJ, USA, 2006. 3. Rekouche, K. Early Phishing. arXiv 2011, arXiv:1106.4692
4. Alabdan, R. (2020). Phishing Attacks Survey: Types, Vectors, and Technical Approaches. Future Internet, 12(10), 168. doi:10.3390/fi12100168
5. Bonguet, A., & Bellaiche, M. (2017). A Survey of Denial-of-Service and Distributed Denial of Service Attacks and Defenses in Cloud Computing. Future Internet, 9(3), 43. doi:10.3390/fi9030043
6. V. D. M. Rios, P. R. M. Inácio, D. Magoni and M. M. Freire, "Detection and Mitigation of Low-Rate Denial-of-Service Attacks: A Survey," in IEEE Access, vol. 10, pp. 76648-76668, 2022, doi: 10.1109/ACCESS.2022.3191430.
7. Askarov, A., Hansen, R. R., & Rafnsson, W. (Eds.). (2019). Secure IT Systems. Lecture Notes in Computer Science. doi:10.1007/978-3-030-35055-0
8. W. Shahid et al., "Detecting and Mitigating the Dissemination of Fake News: Challenges and Future Research Opportunities," in IEEE Transactions on Computational Social Systems, doi: 10.1109/TCSS.2022.3177359
9. Maria Kotolov (4 Feb 2021) Supply chain attacks show why you should be wary of third-party providers, https://www.csoonline.com/,
10. [Online] https://learn.microsoft.com/

11. [Online] https://learn.microsoft.com/en-us/microsoft-365/security/intelligence/supply-chain-malware?view=o365-worldwide,
12. Urciuoli, L., Cyber-Resilience: A Strategic Approach for Supply Chain Management, Technology Innovation Management Review; Ottawa Vol. 5, Iss. 4, (Apr 2015): 13-18.
13. [Online] https://portswigger.net/daily-swig/supply-chain-attacks
14. [Online] https://www.cynet.com/attack-techniques-hands-on/sunburst-backdoor-c2-communication-protocol/
15. Orange Business Internet Security Report 5th edition, 2022, https://newsroom.orange.ro/orange-business-services-lanseaza-raportul-business-internet-security-2022/
16. [Online] https://us.norton.com/blog/id-theft
17. [Online] https://now.symassets.com/content/dam/norton/campaign/NortonReport/2021/2021_NortonLifeLock_Cyber_Safety_Insights_Report_Global_Results.pdf
18. [Online] https://uk.norton.com/products/identity-advisor-plus.
19. [Online] https://www.prnewswire.com/news-releases/norton-launches-robust-identity-monitoring-in-the-uk-to-help-consumers-resolve-their-identity-theft-issues-301502907.html
20. [Online] https://www.orange.ro/docs/business/pdf/Business-Internet-Security-Report-5th-edition-2022.pdf
21. [Online] https://www.nsa.gov/Press-Room/Cybersecurity-Advisories-Guidance/
22. [Online] https://www.enisa.europa.eu/publications/enisa-threat-landscape-2022
23. [Online] https://www.hhs.gov/hipaa/for-professionals/security/guidance/cybersecurity/ransomware-fact-sheet/index.html
24. Perlroth, Nicole (May 13, 2021). "Colonial Pipeline paid 75 Bitcoin, or roughly $5 million, to hackers". The New York Times. Retrieved May 13, 2021.
25. Helmore, E. (May 10, 2021). "FBI confirms DarkSide hacking group behind US pipeline shutdown". The Guardian. Archived from the original on May 12, 2021. Retrieved May 10, 2021
26. Walsh, Joe. "Ransomware Attack Shuts Down Massive East Coast Gasoline Pipeline". Forbes. Retrieved February 6, 2022.
27. [Online] https://www.theverge.com/2022/1/20/22892958/crypto-com-exchange-hack-bitcoin-ethereum-security
28. [Online] https://veruscorp.com/mfa-fatigue-leads-to-breach-of-ubers-corporate-systems/
29. [Online] https://informationsecuritybuzz.com/38-9m-health-records-stolen-from-bangkok-hospital/
30. [Online] https://www.orange.ro/docs/business/pdf/Business-Internet-Security-Report-5th-edition-2022.pdf
31. European Union Agency for Cybersecurity, ENISA Threat Landscape NOVEMBER 2022, ISBN: 978-92-9204-588-3, DOI: 10.2824/764318
32. [Online] https://www.securityweek.com
33. https://therecord.media/cyberattack-brings-down-vodafone-portugal-mobile-voice-and-tv-services/
34. [Online] https://www.theguardian.com/news/2022/feb/20/
35. [Online] https://techmonitor.ai/technology/cybersecurity/lapsus-big-tech-samsung-nvidia
36. [Online] https://www.connexionfrance.com/article/French-news/French-health-insurance-data-leak-what-to-do-if-you-are-affected
37. [Online] https://www.infosecurity-magazine.com/news/finland-government-sites-offline/
38. [Online] https://www.spiceworks.com/it-security/data-security/news/data-of-millions-of-vpn-users-leaked/
39. [Online] https://www.itgovernance.eu/blog/en/cyber-attacks-and-data-breaches-in-review-may-2022
40. Hardman C., Important Update on Email Vendor Security Incident, https://opensea.io/blog/articles/important-update-on-email-vendor-security-incident
41. Glover C., Pegasus Airline breach sees 6.5TB of data left in unsecured AWS bucket, https://techmonitor.ai/technology/cybersecurity/pegasus-airline-data-breach-aws-bucket
42. Smith L., Wason R., Zaidi S., Lockbit, Hive, and BlackCat attack automotive supplier in triple ransomware attack, https://news.sophos.com/en-us/2022/08/10/lockbit-hive-and-blackcat-attack-automotive-supplier-in-triple-ransomware-attack/
43. Page C., Costa Rica's public health system hit by Hive ransomware following Conti attacks, https://techcrunch.com/2022/06/01/costa-ricas-public-health-system-hit-by-hive-ransomware-following-conti-attacks

44. Abrahams, L., Twitter confirms zero-day used to expose data of 5.4 million accounts, https://www.bleepingcomputer.com/news/security/twitter-confirms-zero-day-used-to-expose-data-of-54-million-accounts/
45. Hope, A. Data Breach on Virtual Pet Website Neopets Affected 69 million Users and Leaked Source Code, https://www.cpomagazine.com/cyber-security/data-breach-on-virtual-pet-website-neopets-affected-69-million-users-and-leaked-source-code/
46. Baptista, E. Hacker offers to sell data of 48.5 million users of Shanghai's COVID app, https://www.reuters.com/world/china/hacker-offers-sell-data-485-mln-users-shanghais-covid-app-2022-08-12/
47. [Online] https://dnsc.ro/citeste/comunicat-site-uri-ro-afectate-de-un-atact-de-tip-ddos
48. S. Riurean, M. Leba and L. Crivoi, "Enhanced Security Level for Sensitive Medical Data Transmitted through Visible Light," 2021 International Symposium on Networks, Computers and Communications (ISNCC), 2021, pp. 1-6, doi: 10.1109/ISNCC52172.2021.9615732
49. Riurean, S. A study on the VLC security at the physical layer for two indoor scenarios, MATEC Web of Conferences; Les Ulis, Vol. 342, (2021). DOI:10.1051/matecconf/202134205009
50. Riurean Simona, Robert Alexandru Dobre, Alina-Elena Marcu, Security and propagation issues and challenges in VLC and OCC systems, Proceedings Volume 11718, Advanced Topics in Optoelectronics, Microelectronics and Nanotechnologies X; 117182B (2020) https://doi.org/10.1117/12.2572029
51. [Online] https://www.cshub.com/
52. Hausken, K. Cyber resilience in firms, organizations and societies. Internet Things 2020, 11, 100204, doi: 10.1016/j.iot.2020.100204
53. [Online] [https://www.enisa.europa.eu/publications/enisa-threat-landscape-2022
54. [Online] https://www.sentinelone.com/cybersecurity-101/cyber-kill-chain/
55. [Online] https://abnormalsecurity.com/glossary/indicators-of-compromise
56. https://www.sans.org/media/score/504-incident-response-cycle.pdf
57. https://playbooks.flexibleir.com/incident-response-phases-best-practices/
58. [Online] https://www.techtarget.com/searchsecurity/answer/Email-authentication-How-SPF-DKIM-and-DMARC-work-together
59. [Online] https://www.higherlogic.com/blog/spf-dkim-dmarc-email-authentication/
60. [Online] https://www.clamav.net/
61. [Online] https://github.com/ComodoSecurity/openedr
62. [Online] https://vms.drweb.com/online/?lng=en
63. [Online] https://dto.to/group/11539
64. [Online] https://support.kaspersky.com/KES4Linux/11/en-US/177138.htm
65. [Online] https://www.mcafee.com/en-us/antivirus/mcafee-security-scan-plus.html
66. [Online] https://www.sophos.com/en-us/free-tools/virus-removal-tool
67. [Online] https://www.microsoft.com
68. [Online] https://www.lockheedmartin.com/

Searching Algorithm in a nonrelational database

Roman Ceresnak [1],
Michal Kvet [1[0000-0003-3937-7473]],
Karol Matiasko [1[0000-0001-7173-2661]]

[1] University of Zilina, Zilina, Slovakia

https://doi.org/10.33847/2686-8296.4.2_2

Received 28.10.2022/Revised 28.11.2022/Accepted 23.12.2022/Published 28.12.2022

Abstract. The problem of the data growth and it is storing to the nonrelational data-bases is related to their decreasing efficiency of searching. Nowadays, a very popular database in memory will help us with decreasing the efficiency of the operation searching in this paper. This paper examines the data search-ing in applications hosted in cloud service Amazon with using of nonrela-tional database DynamoDB. It develops new procedures to provide faster response to user and to obtain the data using of nonrelational database Dy-namoDB, that will provide the demanded data and subsequently, it will transfer them to the memory. The given procedure is based on two methods. The first method is a recognition of values, to which the user refers and the provision of this data to the database in memory. The second method is re-lated to the automatic storing of the data transferred to the database in memory. We perform various experiments in the paper, which are showing us a border of efficiency respectively inefficiency from a time perspective.

Keywords: selecting data, SQL database, NoSQL database, cloud.

1. Introduction

The population growth caused, that many proved procedures, where to belong also traditional databases, started losing their efficiency gradually. In contrast with relational databases [2], databases NoSQL process and manage the big data, characterized as 3V (volume, variety, velocity) [3]. Databases NoSQL are needed to support various applications, which need various levels of performance consistency, availability, and scalability [4]. Social media such as Twitter and Facebook [5] generate for example exabytes of the data daily, whose exceed options of processing of relational databases. These applications demand high performance, but they do not have to demand strong consistency.

It is impossible to achieve the same efficiency of searching with a huge amount of the data in the nonrelational databases than with searching in the relational databases. Regarding the searching in the nonrelational databases, the data, which do not have necessary to meet the strict structural demands of systems (RDMBS), are stored, because the data for the searching can be texted, semi-structured or unstructured. A search engine database is created to help the users to quickly find the information they need in a highly qualified and cost-effective method. They are optimized for keywords and usually, they offer specialized methods such for example a full-text searching, complicated expressions of the searching.

Databases of the searching engine contain two main parts. The first content is added to an index of the database of the searching engine. When the user performs inquiry, relevant results are quickly returned with the help of the database index of the searching engine. Responses to fast searching are possible because instead of direct text searching, they search for inquiries "searching" according to the index. It

is an equivalent of pages loading in a book, related to the keyword, searching of the index in the back part of the book, in contrast to the searching of individual words on every page in the book. This type of index is called an inverted index because it transfers the data structure-oriented on a page to the data structure oriented at the keywords.

The searching efficiency is possible to make it more effective by using the database in memory. The database in memory (IMDB) is a computer system storing and searching the data records, which are situated in the main computer memory, for example in memory RAM. IDMB has an advantage with the data in RAM unlike traditional databases based on disks, causing an access delay because stored media such as units of hard disk and SSD have a markedly slower time of the access as RAM. It means IDMB is useful when fast reading and data recording are decisive.

IMDB works the way they preserve all the data in memory RAM. It is a medium, where the data are stored in RAM opposite the disks in the databases based on disk access RAM. The disk part remains untouched in some IDMB, but RAM is primary a storing medium. Some IDMB also stores the data on disk as a preventive measure to minimalize the risk of the data loss, because RAM is volatile (for example the data are losing when a computer loses energy).

The majority of IDMB also protects from the data loss in one data center (ability known as "high availability") preserving the copies ("replicas") of all the data records on several computers in a cluster. This data redundancy secures, that with an error of whichever computer any data record will not be lost. Among the most popular databases in memory, which help us to get the data with help of inquiry language, belong databases such as Redis, Memcached, and so on. Artificial intelligence will help us with the purpose of transfer of the data situated in the nonrelational database DynamoDB.

The problem related to the data transfer from the database on disk to the database in memory is the velocity of the data searching and the transfer efficiency. Artificial intelligence was used for these purposes, which secures this transfer and so it also makes the data searching more effective. The main benefits of this paper are as follow:

- It creates a new procedure of how to process the data in the memory,
- It reduces the time needed for the record searching in the nonrelational database,
- It defines the methods of how to automatically adjust the data growth to the database's size in memory.

The rest of the paper is structured as follows. "State of the art" section examines the related work. "Our contribution" part represents the designed searching model and the characteristics of this model. "The data transfer" and "The index creation" parts describe the performance of the operation in DynamoDB. "Experiments" part describes performed tests and subsequently the results of the valuation.

2. State of the art

A comparison between relational data models and nonrelational data models NoSQL was already stated in various papers. For example, between these two database types, they are showed and recorded the times needed to perform basic operations such as data selection, data insert, data update, and data deleted. Several statistics point out the fact the most common operation, which is demanded in the relational and nonrelational database, is data selection operation. Many authors showed in their papers, that the time needed to get the data in the nonrelational database is significantly worse as the time needed to get the data in the relational

database. Because of an acceleration respectively improvement of the time needed to get the data from the nonrelational database, it is possible to store the data to buffer memory and by this to reduce repeated searching in the nonrelational database. Not only the time is reduced by this method, but several accesses to the database, too. The authors performed various comparisons between databases in memory such as Redis, Memcached, and nonrelational databases Mongo, Casandra, and H2. One of the main findings of these works is the verification of data update and delete with an increasing number of the data.

We noticed a work during our research, that focuses on problem-solving with the increasing amount of the data. In [3] the authors created a module by using the library Lontar, which will send the data to the relational database by Hibernate as a framework and a relational mapper, in the case of user's demand. Subsequently, Hibernate accesses to MySQL and maps the relational data to object-oriented, and then it sends the data to the nonrelational database. The searching then works with the help of the mapper, so Lontar could be able to read the data in a relational relationship. According to the authors, some data files got better results in nonrelational database MongoDB with the operation searching as in relational database MySQL. However, in certain situations, as the authors describe further, the relational database got better results than the nonrelational database.

The authors introduced a framework capable to manipulate with the data to overcome the problems related to the decreasing efficiency of the searching in the nonrelational databases opposite the searching in the relational databases, Yet before performing of the basic operations of data selection, data insert, data update, and data delete and edit by the mapper happens. The main role of the mapper is to change the data on the base of rules, to such form, that complies with principles of nonrelational database MongoDB more, regarding the searching. By this module, the situation, when the data are faster searched with a help of nonrelational database MongoDB then relational database MySQL, happens in the majority of cases. Another concept used in this framework is the Cataloging module, which uses JSP (JAVA) as a web programming technology and MySQL as DMBS in the principle [11]. There exist two frameworks supporting it, Struts and Hibernate [11]. Structs are used to set a user's interface and the Hibernate regime is used to map the relational data to the object-oriented data, which will be used by JSP. We design a framework in our work, which also uses two database types. The first database is nonrelational database DynamoDB serving as a primary data storage. In the case, when user demands demanded data, the values will not be directly given to them from the nonrelational database, but the values will be transferred to the database in memory. The main challenge to get this aim is to transfer the data from the nonrelational model to the model in memory.

3. Our Contribution

In this part, we will introduce two modules helping us to make the searching in the nonrelational databases more effective. Data Cached Module and Data Elastic Module. DCM serves as a storage data storehouse, which role is the data transfer to the buffer memory. DEM serves on automatic adjusting to the data size.

3.1. A Subsection Sample

The created module serves on the data processing in the memory. We connected the data we store in the nonrelational database DynamoDB to highly available buffer memory Amazon DynamoDB Accelerator, very well known in short as

DAX, with the help of API interface. This method helps us with 3 accesses: Side-cache, Read-through cache, and Write-through cache.

a) Side-cache

This principle helps us with high overload during the reading of information from the memory. This principle works as follow:

1. An application first tries to load the data from the buffer memory for a given couple of key-value. If the buffer memory was filled up with the data (access to the buffer memory), the value returns. If not, step 2 follows.
2. The application loads the data from the storage of the basic data because the demanded couple key-value was not found.
3. A couple of key-value from step 3 will be written to the buffer memory to make sure the data are present when the application needs to load the data again.

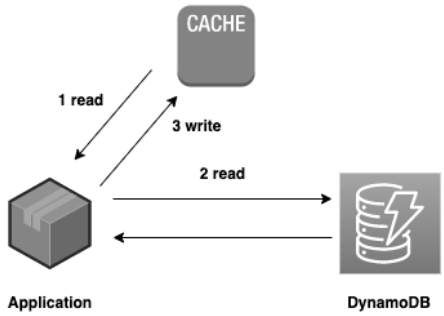

Fig. 1. Side-cache algorithm

b) Read-through cache

DAX is a buffer memory for the reading, because it is compatible with API for the reading of API DynamoDB and stores the results GetItem, BatchGetlem, Scan, and Query to the buffer memory, if they are not currently in DAX. The buffer memory for the reading is effective in difficult working loads. This principle works as follow:

1. Regarding a couple of key-value from the application, it first tries to load the data from DAX. If the buffer memory was filled up with the data (the access to the buffer memory) the value return. If not, step 2 follows.
2. Transparently for the application, if a semi-memory happens, DAX will load a couple of key-value from DynamoDB.
3. To make the data available for every reading that follows, then a couple of key-value will full up in semi-memory DAX.
4. A couple of key-value then will return the value to the application.

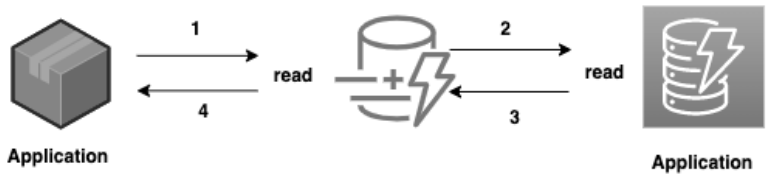

Fig. 2. Read-through cache algorithm

c) Write-through cache

Similarly to the semi-memory for the reading, a semi-memory for the data writing also operates in a line with the database and updates the semi-memory, when

the data are written to the storage of basic data. DAX has also buffered memory for writing, because it stores to the buffer memory (or updates) the items with PutItem, UpdateItem, DeleteItem and BatchWriteItem, API, because the data are written or updated in DynamoDB. At first, DAX is updated (everything is transparent for the application). The following steps indicate a procedure for buffer memory type write-through

1. The application will write itself to endpoint DAX for a given couple of key-value.
2. DAX will catch the writing and then will write a couple of key-value to DynamoDB.
3. After the successful writing, DAX hydrates buffer memory DAX with a new value so whichever following the reading of the same couple key-value results in a finding of the buffer memory. If the writing is unsuccessful an exception will return to the application.
4. Confirmation of successful writing will then return to the application.

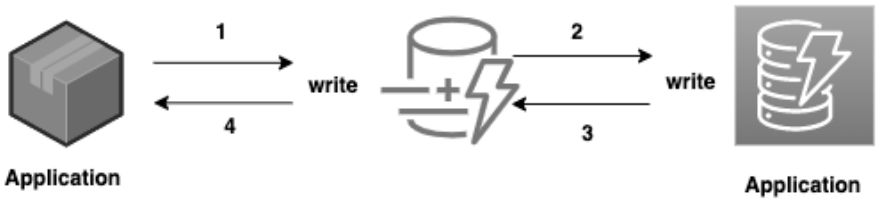

Fig. 3. Write-through cache

3.2. Data Elastic Module

The created module serves with the data increasing in the memory. Nonrelational database DynamoDB fulfills the task of a wide data storage and in the case of the data transfer to the database in memory the data, size can move from several megabytes until gigabytes. This mentioned problem is currently solved by the created module.

We configured monitoring of metrics in cloud service Amazon with the help of service Amazon CloudWatch. The mentioned service makes it possible to edit respectively to add and to remove new calculation units in the case of enabling of horizontal or vertical scaling. An advantageous characteristic of this method is the horizontal scaling, that in the case of a huge number of the data uploads invokes warning of big overload and a script for the reading of information from other replicas in service CloudWatch. The horizontal replica is the part of the script performed automatically during the configuration of the database in memory DAX by the following script.

```
aws dax decrease-replication-factor \
    --cluster-name MyNewCluster  \
    --new-replication-factor 3
```

The monitoring of the metrics in the same way with our method also makes the vertical scaling possible, what is the scaling by addition or removal of the calculation units.

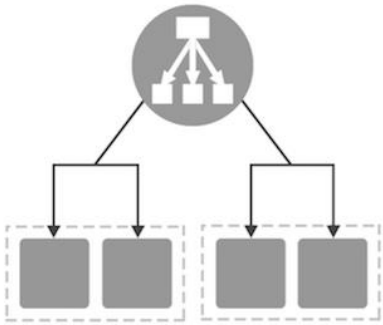

Fig. 4. Application Load Balancer for in-memory database

4. Experiments

In the very first step, we created a simple database model seen in fig 5. This database model is created from two tables, user and comment. These two tables are interconnected by identification relationship type 1:n, which means a 1 user can create various comments and various comments in the table belong right to the 1 user. Subsequently, we compared various orders, whose aim is to get information about the increasing trend of the searching with the increasing number of the data in the relational database Oracle in section A, and then also in nonrelational database DynamoDB in section B. We compared also the times of various operations during data selection in section C because an important aspect of this paper is using of the database in memory.

4.1. Experiments for relational database Oracle

In the very first step, we created a simple database model seen in fig 5. This database model is created from two tables, user and comment. These two tables are interconnected by identification relationship type 1:n, which means ta 1 user can create various comments and various comments in the table belong right to the 1 user.

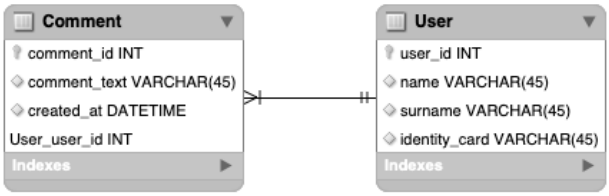

Fig. 5. Data model

We inserted 1000 records into table users and 1000 records into table comments too, based on this defined structure. In the case of this paper, we are interested only in information about the time needed for record searching. We created 3 data selection orders for these purposes looking as follow:

(1) SELECT name, surname FROM user JOIN comment USING (user_id);

*(2) SELECT * FROM user*

```
JOIN comment USING (user_id)
WHERE comment_text LIKE "%today%";

(3)      SELECT  name, surname,
to_char(created_at, 'YYYY-MM-DD HH24:MI:SS') ca FROM user
JOIN comment USING (user_id)
WHERE ca >= TRUNC(current_date)
and ca < TRUNC (current_date) + 1;
```

The first 1000 records served as very first records to us, from which our research can bounce off. The main purpose of the nonrelational databases is to effectively store a huge amount of the data, and that is, why the records for other purposes will be created with the size of 100000 records for table users and 100000 records for table comment. Subsequently, all records are deleted and the records with size 10 000 000 will be inserted to table user and comment. As the last size of the records, we chose a value of 100 000 000 records.

A generator was used for record creation with the size 1 000, 100 000, 10 000 000 and 100 000 000, which is possible to find on this address https://www.generatedata.com/ The generator provides an option to define names and types of attributes and to generate an arbitrary number of the values. After fulfilling the tables by the generated values, we recorder the times needed to perform operations (1), (2), and (3), and they are portrayed in Table 1.

Table 1. Measure time for operation (1) (2) (3) in Oracle

Count of records/operation	1 000	100 000	10 000 000	100 000 000
(1)	0,0020	0,004	0,028	0,44
(2)	0,0021	0,0042	0,031	0,45
(3)	0,0021	0,0044	0,030	0,45

The values needed to get the data from the relational database Oracle are recorded in Table 1. All achieved values for orders (1), (2), and (3) are measured in seconds.

4.2. Experiments for nonrelational database DynamoDB

We used orders (1), (2), and (3) to find out the velocity of demand in the nonrelational database DynamoDB. The values inserted into the database were left the same as in experiments in the relational database. The structure is fully the same as it is recorded in Fig 3.

Table 2. Measure time for operation (1) (2) (3) in DynamoDB

Count of records/operation	1 000	100 000	10 000 000	100 000 000
(1)	0,0035	0,0064	0,047	0,82
(2)	0,0035	0,0067	0,048	0,83
(3)	0,0037	0,0068	0,046	0,82

The values, needed to get the data from the nonrelational database DynamoDB are recorded in table 2. All achieved values for orders (1), (2), and (3) are measured in seconds. During a comparison of the value of the same operations, the values between the results of the relational and nonrelational databases are significantly

different. The nonrelational database in data selection operation is less time effective than relational database Oracle.

4.3. Experiments for the database in memory Redis

The storing in the database in memory is diametrically different than in the relational or nonrelational databases. Except for the mentioned fact, the big problem is also a limitation of the data amount by computer memory. The computer memory about 8 GB was used for testing purposes.

Table 3. Structure of data

ID	Name	Surname	Identity_card	ID
1	John	Harper	12341324	1
2	Joe	Bush	12341234	2
3	George	Obama	23524675	3
.....	.....	.....	.....	.....
.....	.....	....	.....	.....
1000	Alan	Felps	45674866	1000

As is seen in table 3, we created 100, 300, 500, and 10000 records with structure ID, Name, Surname, and Identity_card. The page on this address https://www.generatedata.com/ was used for this purpose against. 3 orders were created for purposes of testing the database in memory's effectiveness, where we were watching the velocity of getting the data. They are the following cases:

(4) MGET Name

(5) MGET Name, Surname

(6) MGET Name, Surname, Identity_card

(7) MGET Name, Surname, Identity_card, Age

We applied the same principle during filling the database as in previous steps. In the actual case, we inserted the generated values to the database, tested operations (4), (5), (6), and (7), and recorded the values. Subsequently, we deleted all the records and inserted the data about 300 records to the database and we continued like this until the size of 1000 records in the database. The recorded values for the operations are recorded in table 4.

Table 4. Measure time for Redis database

Count of records/operation	100	300	500	1 000
(4)	0,00020	0,00022	0,00021	0,00028
(5)	0,00021	0,00023	0,00025	0,00029
(6)	0,00021	0,00022	0,00025	0,00028
(7)	0,00023	0,00024	0,00028	0,00032

All achieved results we got were recorded in table 4 and are in seconds. The seventh operation is influenced by the fact, that value "age" does not exist. As it is seen, the measured values are not diametrically different from the increasing of the values. It is necessary to point out, that with defined growth of the records, it is the logic fact mirroring the efficiency of the searching in the memory.

4.4. Comparison of the final results

The values we got in experimental activity in sections A, B, and C serve right on the comparison with our designed method. The values needed to get the data with operations (1) were measured and recorded in Table 5.

Table 5. Comparison of query performance

Count of records/operation	1 000	1 000 000
Oracle	0,0020	0,44
DynamoDB	0,0035	0,82
Our Approach	0,0033	0,42

As is seen in Table 5, the values measured and got, while using operation (1), do not show any big improvement of the searching in the nonrelational table to us, with a low number of records in the table. This phenomenon is influenced by the data transfer to the memory. A factor of the transfer indicates a necessity to transfer the data from nonrelational database DynamoDB to buffer memory DynamoDAX, which takes a certain time and already when the records are got by the user. It means in operation data selection, the data are physically not got from the nonrelational database DynamoDB, but from the database in memory.

Based on the data transfer, it was possible to compare also the values between the experiments with the database in memory and the data transferred to DynamoDAX with the size of 100 and 500 records during the operation (5).

Table 6. In memory query performance

Count of records/operation	100	500
Redis	0,00020	0,00021
DynamoDAX	0,00015	0,00017

The values recorded in table 6 show us a clear way and efficiency of the data transfer. It is seen, that the values in buffer memory Dynamo DAX are more effective from the time perspective opposite database in-memory Redis.

Whole achieved results related to operation data selection in nonrelational database DynamoDB were not, before the application of our method, timely the same effect than after the application of our method. With using of machine learning and transferring the data to the database in memory, the efficiency of operation data selection in the nonrelational database became more effective after achieving 1000 000 records than with the searching of the data in relational database Oracle. A huge advantage, that results in using of cloud storage Amazon, is related also to the possibility of automatic scaling respectively adding of performance and increasing of the storage not only in nonrelational database DynamoDB, but mostly in the database in memory alternatively, if we do not need as many calculation units, so the reduction of the size of the data storage happens, and so the decreasing of the cost related to running of our designed method happens.

5. Conclusion

NoSQL databases play a significant role in storing and processing huge data and they are used in various wider social applications such as for example Twitter, Facebook, Google, and Yahoo, but they help also with the support of deciding or with creating of advanced analyses. They became the master of high effectiveness and availability of the huge data, but with the loss of the effective searching opposite the traditional databases. This document was devoted to the question of the searching in database NoSQL, concretely Dynamo DB in the cloud background of Amazon to minimalize the impact of this problem.

In this paper, we developed the searching algorithm, that can make the velocity of the searching in a nonrelational database DynamoDB more effective. The designed algorithm is composed of two parts, the first part is based on the principle of caching of the data from the nonrelational database DynamoDB to buffer memory DynamoDAX. The second part is based on the effective data management, that is able, in the case of the huge amount, to automatically create and increase the calculation units of the buffer memory, and by this to adjust the size of the database to the increasing needs of the size of incoming data. This fact relieved us from the limitation of the database size towards the data.

The experiments gave some useful information about the performance and the effectiveness of the created method. It is noted, that the system for processing artificial intelligence demanded higher overhead costs together with automatic creation of the database in memory, but this system was able to make the process of searching in the nonrelational database more effective. On the basis of the experiments, it is clearly seen, the created method is more and more effective with the increasing data amount, which is done by the data transfer to the memory.

Our further work will focus on a generalization of this model and provision of API interface for full use of the created procedure not only for cloud Amazon but also for other databases in memory such as Redis and Memcached. We also plan to evaluate the suggested systems empirically, from a perspective of consistency and performance in other backgrounds, who need a fast response for the data demand.

Acknowledgement

This work was supported by Grant System of University of Zilina No. 1/2020. (8056).

References

1. G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.
2. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
3. S. Jacobs and C. P. Bean, "Fine ppapers, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
4. R. Čerešňák and M. Kvet, "Comparison of query performance in relational a non-relation databases," in *Transportation Research Procedia*, 2019.
5. R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.
6. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
7. M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

System of Automatic Recognition of Video Text Amazigh based on the Random Forest

Youssef Rachidi[0000-0001-9682-5514]

University Ibn Zohr Agadir, Marocco

https://doi.org/10.33847/2686-8296.4.2_3

Received 26.10.2022/Revised 26.11.2022/Accepted 23.12.2022/Published 28.12.2022

Abstract. In this paper; we introduce a system of automatic recognition of Video Text Amazigh based on the Random Forest. After doing some pretreatments on the video and picture, the text is segmented into lines and then into characters. In the stage of characteristics extraction, we are representing the input data into the vector of primitives. These characteristics are linked to pixels' densities and they are extracted on binary pictures. In the classification stage, we examine four classification methods with two different classifiers types namely the convolutional neural network (CNN) and the Random Forest method. We carried out the experiments with a database containing 3300 samples collected from different writers. The experimental results show that our proposed OCR system is very efficient and provides good recognition accuracy rate of handwriting characters images acquired via Video camera phone.

Keywords: Pretreatments, Video Text Amazigh, Mobile phone, OCR, CNN, Random Forest.

1. Introduction

The automatic recognition of handwritten or printed Amazigh characters remains a subject of research and experimentation. The problem is not yet solved despite the fact that results have reached fairly high rates in some applications [1]. Some attempts have been done to improve the current situation [1]. In this context, we have employed a recognition system of Amazigh handwritten characters extracted from video taken by camera phone [2]. Indeed, in the primitives' extraction stage, our approach is based on primitives of the Zoning types [3], of distance profile feature [4], Projection histogram and Gray Level Co-occurrence Matrix (GLCM) technique [5]. These primitives will supply a Convolutional Neural Networks and Random Forest Method in the learning and recognizing phases. Video text handwritten Amazigh, segmented and isolated characters acquired by camera phone, obtained an encouraging results on the majority of this characters.

Habitually, the phases form the structures of handwriting recognition system are: Pre-processing, Segmentation, Feature extraction, Classification and Post-processing [2].

In this paper, our objective is mainly interested in the development of Video Text handwriting Amazigh recognition system and Improvement of the Recognition Rate by CNN in some characteristics extraction, in which the images from video.

The paper is organized as follows. In section 2, the proposed the pre-processing and gives descriptions of the methods that we used throughout the OCR process, which includes the following stages: Binarization, Noise removing, skew detection and correction and Segmentation. The feature extraction procedure adopted in the system is detailed in the section 3. Section 4 describes the classification and recognition using CNN and Random Forest. Section 5 presents the experimental results and comparative analysis. Finally, the paper is concluded in section 6.

2. Pre-Processing

The procedure of preprocessing which refines the scanned input image from video includes several steps: Binarization, for transforming gray-scale images in to black and white images, noises removal, and skew correction performed to align the input paper document with the coordinate system of the scanner and segmentation into isolated characters [1].

2.1 Binarization and Noise Removal

We used the Sauvola method for binarization [6] this method of thresholding is performed as a preprocessing step to remove the background noise from the picture prior to extraction of characters and recognition of text from video.

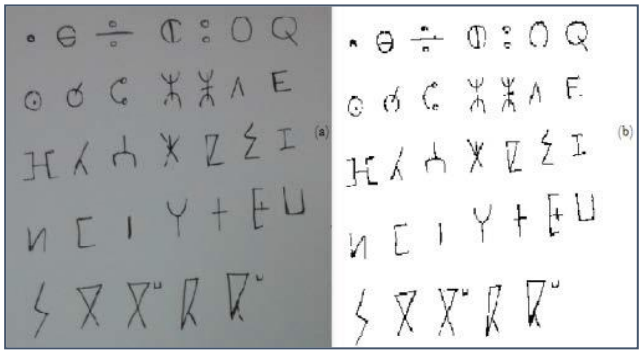

Fig1. (a) Example of an input image, (b) Thresholded image with Sauvola method.

Noise which is in the images is one of the big difficulties in optical character recognition process. The aim of this part is to remove and eliminate this obstacle; there are several methods that allow us to overcome this problem. In this work we decided to use the morphology operations to detect and delete small areas of less than 30 pixels[2].

2.3 Skew detection andcorrection

Skew correction methods are used to align the paper document with the coordinate system of the scanner. Main approaches for skew detection include line correlation [7], projection profiles [8], Hough transform [9], etc. For this purpose two steps are applied. First, the skew angle is estimated. Second, the input image is rotated by the estimated skew angle. In this paper, we use the Hough transform to estimate a skew angle θs and to rotate the image by θs in the opposite direction.

2.4 Segmentation

Next step for OCR is the Segmentation of the image. In This paper we propose a segmentation algorithm, in which text is easily segmented into Lines and Words using the traditional vertical and horizontal projection[10].

2.4.1 ***Line Segmentation***

Once the image of the text cleaned, the text is segmented into lines. This is used to divide text of document into individual lines for further preprocessing. For this, we used analysis techniques of horizontal projection histogram of the pixels in order to distinguish areas of high density (lines) of low-density areas (the spaces between the lines) (see Fig.2). These techniques were often used to extract lines in printed texts[1].

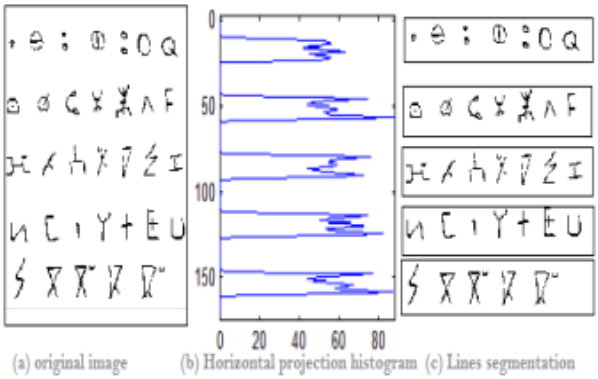

Fig.2 Lines segmentation

2.4.2 ***Characters Segmentation***

We used in this part the vertical projection histogram to segment each text line of characters. Fig.3 shows a text line, the vertical histogram and the result of segmentation into characters [2].

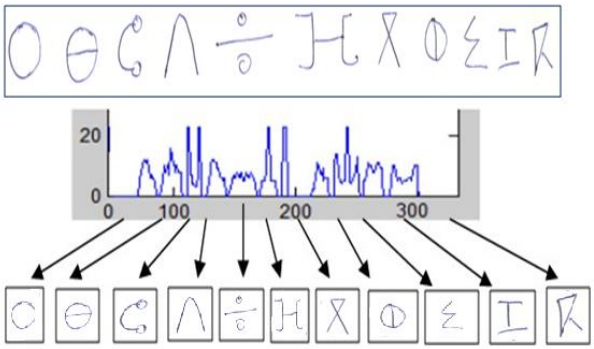

Fig.3 Characters segmentation

3. Feature extraction

In this part we present some feature extraction methods for recognition of segmented (isolated) characters [11]. Selection of a feature extraction method is probably the single most important factor in achieving high recognition performance in character recognition systems. Different feature extraction methods are designed for different representations of the characters, such as solid binary characters, character contours, skeletons (thinned characters) or gray-

level sub-images of each individual character[11], In this paper, we have tested four methods: the Zoning types, Distance profile feature, Projection histogram and Gray Level Co-occurrence Matrix (GLCM) technique.

3.1 Zoning:

The zoning technique [3] is a statistical region-based feature extraction, it aim is to get the local characteristics in lieu of global characteristic. Therefore, according to the size normalized character image (60 x 50 pixels), we divided it into 30 (6 x 5) zones of 10 x 10 pixels size, then we calculated the densities of pixels in each zone, finally we are getting 30 features.

3.2 Projection histogram:

Projection histogram descriptor is a statistical feature; According to this feature we have used two direction of projection horizontal traversing. The horizontal histogram of the character Amazigh computed by counting the number of black pixels in each row. At the last we will have 60 features depending on the direction projection.

3.3 Gray Level Co-occurrence Matrix:

Gray Level Co-occurrence Matrix (GLCM) technique is an approach for extracting statistical texture features that have been proposed by Haralick [5]. The main principle of GLCM is to counts the number of times various combinations of pixel gray levels occur in a given image. Haralick defines 14 statistical features measured from the GLCM. In this work, five important features are used namely energy, contrast, correlation, entropy and homogeneity.

3.4 Distance profile:

In distance profile feature [4] the distance (number of pixels) between the bounding box of image and the first pixel of foreground will be calculated. We have employed two types of profiles sides left and top. Concerning left profile, it is extracted by counting the distance from the left bounding box to the nearest foreground pixels in each row. Then as well, top profile, it is extracted by counting the distance from the top bounding box to the nearest foreground pixels in each column.

Table 1. Combination of the different feature vectors

Feature Method	Contained feature	Size
FM1	Zoning	30
FM2	GLCM	5
FM3	Distance Profile (Left + Top)	110
FM4	Projection Histogram Horizontal	60

4. Claccification

In the complete process of system recognition of forms, the classification plays an important role by pronouncing on the membership of a shape in a class. The main idea of the classification is to attribute an example (A form) not known about one Class predefined from the description in parameters of the form. Several surrounding areas of classification are used in the field of recognition of forms which are more or less good adapted to the recognition of the writing.

In literature, there are many types of classifiers that have been implemented in handwritten optical Amazigh character recognition problems. Among them, in this paper we have used two classifiers: the Convolutif Neural Network (CNN) and Random Forest.

4.1 Convolutif Neural Network

Convolutional networks are derived from perceptron architectures Multi Layer Perceptron (MLP), however they use shared weights, related to the convolution window, which allow them an implicit extraction of local features.

The difference of Convolutional neuron networks compared to conventional networks of MLP type, let us analyze the principle of recognition on the character G (" yag" in Amazigh), Fig.4 A neuron of an MLP is fully connected to all the neurons of the previous layer while for a convolutional network, a neuron is connected to a subset of neurons of the previous layer. Each neuron can be seen as a unit for detecting a local characteristic, a particular structural singularity such as the detection of a vertical or horizontal line,or even a loop. Alongthe trajectory.

The matrix of weights corresponding to the sliding window is identical (notion of shared weights): same detection, same convolution

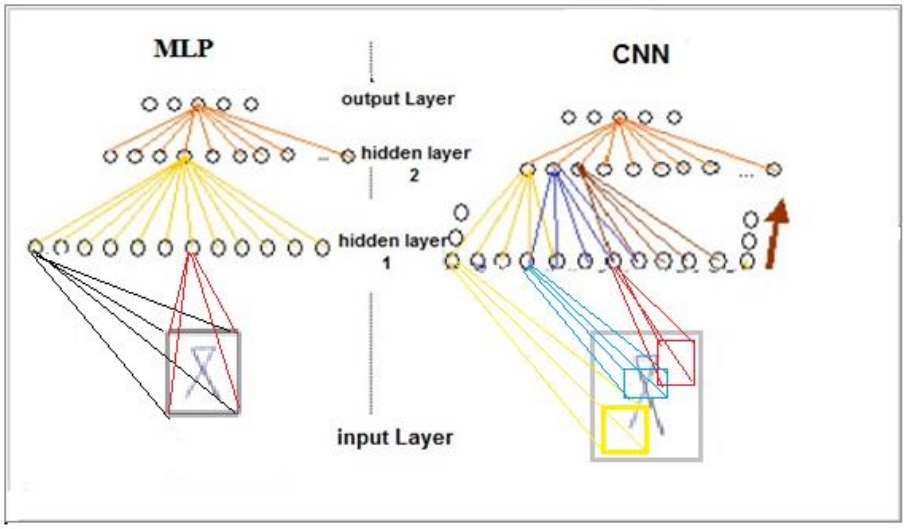

Fig.4 The difference of Convolutional neuron networks compared to conventional networks of MLP type

4.2 ***Random Forest***

Random forest is an ensemble training algorithm that constructs multiple decision trees. It suppresses over- fitting to the training samples by random selection of training samples for tree construction in the same way as is done in bagging(Breiman,1996)[12], (Breiman,1999)[13], resulting in construction of a classifier that is robust against noise. Also, random selection of features to be used at splitting nodes enables fast training, even if the dimensionality of the feature vector is large [1].

- Algorithm

z={(x1, y1), … (xn, yn)} learning sample, xi describes nominal variables p explanatory

:

1. for b=1 to B (B number of trees)

(a) Draw a boostrap sample zb of size N from the training data

(b) Grow a random-forest tree Tb to the bootstrapped data, by recursively repeating the following steps for each terminal node of the tree, until the minimum node size nmin is reached.

i. Select m variables at random from the pvariable

ii. Pick the variable/split-point among them

iii. Split the node into two daughter nodes

2. Output the ensemble of tree $\{T_b\}_1^B$

To make a prediction at a new point x: Regression:

$$\hat{f}_{\mathrm{rf}}^{B}(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x) \tag{4}$$

Classification: let $\hat{C}_b(x)$be the class prediction of thebth random-forest tree.Then

$$\hat{C}_{\mathrm{rf}}^{B}(x) = majority\ vote\ \{\hat{C}_b(x)\}_1^B \tag{5}$$

5. Experimental results

Due to the absence of standard database of handwritten Amazigh characters acquired by camera phone, we have constructed our own database of upper-case Amazigh character images obtained by Video camera phone. The database contains 100 samples of 33 classes, collected from 10 different writers. As a result the database consists of 3300 samples.

The samples are divided randomly into two set, one for training stage, we have used 85 % (2805 samples) and the other for testing stage, we have used 15 % (495 samples). We have tested the proposed system on database of handwritten Amazigh characters acquired by camera phone the HUAWEI Y9 of this characteristics; 16Megapixels.

For classification stage we have used two classifiers: the Convolutif Neural Network (CNN) and the Random Forest for each classifier we employed a set of different features extraction methods.

The Zoning feature extraction (FM1) provides higher recognition and learning rate by Random Forest with the achievement of a Recognition rate of 94.02 %. Also FM 4 and FM 3 give some encouraging results. According to the results of Convolutif Neural Network the hybrid method FM 2 achieves a very good recognition and training rate: 96, 19 % of Learning rate and 94, 01 % of Recognition rate.

Table 2. Results of different single feature vectors using Convolutif Neural Network and Random Forest classifiers

Classifier Feature Vector	Convolutif Neural Network		Random Forest (N=600)	
	Learning R.	Recognition R.	Learning R.	Recognition R.
FM1	95.66 %	93.36 %	95.26 %	94.02 %
FM2	96.19 %	94.01 %	94.54 %	92.64 %
FM3	93.41 %	91.59 %	93.94 %	93.22 %
FM4	95.10 %	93.70 %	93.62 %	93.13 %

Out of the 495 Read-only characters, 392 were recognized, representing a recognition rate of 94.02%. With respect to the rate obtained for each letter, the best result achieved with this approach was 98.89%, for the character (‘Ha’ in Amazigh)). Table 3 below shows the recognition rate obtained on certain characters.

Table 3. Recognition rate

Characters	Recognition Rate
	94.39 %
	98.89 %
	93.72 %

The recognition errors are high for the letter (‘Rr’ in Amazigh), which is explained in particular by the insufficiency of the characteristics used to better describe each character during phase of extraction the primitives, and to the initial data used during the learning step. A good estimate of this data can reduce the error rate of our system.

6. Conclusion

In this paper, we have presented a system of handwriting Video Text Amazigh recognition based on the method Random Forest and Convolutif Neural Network. Several features have been studied and compared; as a result we've chosen Sauvola [10] method duo its ability to remove the noise. The experiments carried out in database were performed on a database obtained by Video camera phone with applying different classifiers and for each classifier we have tested a set of single feature methods.

The results obtained in this paper that has been compared and analyzed have shown that Random Forest with Zoning feature is the best in terms of recognition accuracy rate and GLCM technique provide higher recognition rate by CNN.

In future work, we will add other features methods that improve the results for some characters for example, minimize the length of execution of program which to calculate the recognition rate.

References

[1] W.Chen ,Xiaoshen X, J.Wang, B. Pradhan ,Haoyuan H, D. Tien Bui , Z. Duan, Jianquan M, 'A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility', Vol 151, pp 147-160 , April 2017.

[2] H. El Bahi, Z.Mahani and A. Zatni "A robust system for printed and handwritten character recognition of images obtained by camera phone" .Published in WSEAS Transactions on Signal Processing, Volume 11, 2015, pp.9-22

[3] Elima H, Abdul H, Kishore K, "A Zoning based Feature Extraction method for Recognition of Handwritten Assamese Characters"
IJCST Vol. 6, Issue 2, April - June 2015

[4] D. Guan ; D.Xiang ; X. Tang ; Li Wang ; G. Kuang "Covariance of Textural features: A New Feature descriptor for SAR image Classification", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 12,issue 10, pp. 3932-3942,Oct 2019.

[5] J. Sauvola and M. Pietikainen, "Adaptive Document Image Binarization," Pattern Recognition 33(2), pp. 225–236,2000

[6] SAB Haji, A James, S Chandran, "A Novel Segmentation and Skew Correction Approach for Handwritten Malayalam Documents", Vol 24, pp 1341-1348 , 2016

[7] E. Grilli, F. Menna, F. Remondino, "A REVIEW OF POINT CLOUDS SEGMENTATION AND CLASSIFICATION ALGORITHMS", Volume XLII-2/W3, March 2017 Nafplio, Greece.

[8] Y.Boukharouba, 'A New algorithm for skew correction and baseline detection based on the randomized Hough Transform, JKSU-CIS, Vol 29, issue 1, pp 29-38, January 2017

[9] Archana A. Shinde, D.G.Chougule, "Text Pre- processing and Text Segmentation for OCR" IJCSET
|January 2012| Vol 2, Issue 1,810-812

[10] Y.rachidi, Z.mahani ; "Recognition image Obtained by camera phone of character Arabic", IJDR, Vol 08, March 2018.

[11] Breiman, L. ;Using adaptive bagging to debias regressions. In Technical Report. Statistics Dept.UCB.

[12] Saad Albawi , Tareq Abed Mohamed, Saad Al-zawi ,"Understanding of a convolution neural network", ICET, 21-23 Aug 2017.

Investigating Different Social Media Platforms Used by Tourists to Book a Hotel in Greece

Olympia Vlachopoulou[1][0000-0003-1207-4130], Vasileios Paliktzoglou[2]

[1] London Metropolitan University, UK, Chandigarh University, India
[2] Bahrain Polytechnic, Bahrain

https://doi.org/10.33847/2686-8296.4.2_4

Received 28.10.2022/Revised 28.11.2022/Accepted 23.12.2022/Published 28.12.2022

Abstract. The tourism industry has been recognized as one of the largest economic sector in Greece. The expansion of social media has contributed to introducing new digital marketing tools and changed the way tourist acquire and digest information in the decision-making process to book a hotel. The aim of this quantitative research is to investigate the different types of social media platforms used by tourists to book a hotel in Greece. Descriptive analysis was employed to analyze the data (N= 171) from tourists in Greece. The findings revealed that the majority of participants used TripAdvisor, Instagram and Nikana.gr to book a hotel in Greece, followed by grecia.directbooking.ro and Booking.com receiving lower percentages. Moreover, the guests' reviews of the hotel, the photos and shots of the hotel on the social media platforms and special offers and discounts on hotels' social media were the three main participants' criteria for choosing social media to book a hotel. This study provides an insight for all the relevant stakeholders involved with social media in the travel and hospitality sector more specifically in the hotel field in Greece.

Keywords: Social media, tourism, hotels, booking, decision-making.

1. Introduction

Tourism as phenomenon comprises of social, cultural, and economic aspects that involve people traveling to nations or locations for personal, recreational or professional reasons [1]. Social media is a type of electronic communication with variety of different social media platforms that allows people to produce, share and even exchange information, ideas, experiences, photographs, and videos [2]. The use of social media is one of the most popular online activity worldwide, forecasting that the number will reach to almost 4.41 billion in 2025 [3].

According to Azazi and Shaed [4] the majority of the platforms used by the tourists are social media. In the context of tourism, social media provide platforms for travelers to find information as well as share their expertise and experiences with other tourists [5], [6]. Commonly social media used by tourists are Facebook, Instagram, Twitter, TripAdvisor, Trivago, Booking.com, Instagram, TikTok and a lot more [7]–[10]. The expansion of social media has contributed to introducing new digital marketing tools and changed the way tourist acquire and digest information while making a purchasing decision [11]. Research outlines that tourists nowadays are very demanding expecting personalized experiences from hotels that lead to their increased or decreased booking intention [12]. Thus, social media as a marketing tool plays an important role in all aspects of tourism especially hotels and accommodation and it is becoming an increasingly emerging research topic.

1.1 Background and rationale

The tourism industry has been recognized as one of the largest economic sector, generating 10.3% of global gross domestic product (GDP) and supported the livelihoods of 330 million people in 2019 [13]. Considering the nature of tourism as a service, social media are appropriate platforms for travelers to obtain information [14]. Hotels with the use of social media can engage and interact with customers while also providing accurate, real-time and personalized information long before they make a purchasing decision [15]. Prior to making a purchase choice, tourists begin to gather and analyze information from a variety of sources in order to meet their purchasing needs. They then evaluate several options depending on the information they have gathered [16]. The change from traditional marketing strategies to online is a challenge for tourism and hospitality experts who want to adapt their businesses to travelers' needs for more information and engagement [17]. This reveals additional challenges in the overall decision-making process of booking accommodation [18] and thus, more investigation is required.

Social media has emerged as strategy and it is significant element in the travel industry in recent years, with organizations rapidly investing more time and money in developing shareable articles, videos, and interactive media across all channels [19], [20]. In the context of hospitality, social media is having a holistic perspective of all digital information and striving to have to meet tourists' needs rather than simply presenting them with information. Social media also plays an important role in conveying effective information to the consumers thus, attract them to keep engaging with the brands [19]. There is relatively limited academic literature on the importance of the different types of social media platforms used by tourists to book a hotel in Greece, despite the increased interest in the field.

1.2 . Research aim

The goal of the social media platforms for hotels is to increase customer awareness of their brand and product as well as learn more about their needs and requirements and it is crucial to have an online presence so that they can connect with customers, get to know them, and tailor the services they offer [21]. Social media influence on tourists' decision-making in booking a hotel is an emerging research topic, with a diverse range of travel and hospitality studies in the literature [4], [5], [7], [10], [11], [22]. However, the different types of social media platforms affecting tourists' decision-making process to book a hotel in Greece is yet to be investigated in depth. The aim of this research is to investigate the different types of social media platforms used by tourists to book a hotel in Greece. Thus, to achieve the aim of the research the following research question was derived:

What are the different types of social media platforms used by tourists to book a hotel in Greece?

In order to address the research question, the type of social media platforms, and criteria for choosing them by the tourists, were investigated.

The study consists of five sections: the introduction section includes the background and rationale, and the research aim. The literature review section consists

of the importance of social media and tourism in Greece, the social media's role in tourism and its influence on the hotel industry. The third section represents the research methodology as well as the ethical considerations for the research. The result section represents the descriptive analysis of the research, and final the conclusion section represents the key findings of research, discussion of the implications of the results, limitation of the study and the concluding and future remarks.

2. Literature Review

2.1 Social media

Social media is described by Kotler and Keller as a tool or method used by customers to exchange information with other individuals and businesses (or vice versa) in the form of text, photographs, audio, and video [23], [24]. The Table 1 represents the different types of social media commonly used.

Table 1: Different types of social media platforms

Social media platforms	Examples
Social networking sites	Facebook, LinkedIn
Social blogging	Tumblr
Microblogging	Twitter
Social review sites	TripAdvisor, Yelp, our Square
Image sharing sites	Instagram, Pinterest
Video hosting sites	YouTube, Vimeo
Discussion sites	Reddit, Quora
Virtual social worlds	Second Life
Social bookmarking and voting	Delicious, Digg
Sharing economy platforms	Airbnb, Pantheon, Kickstarter
Social knowledge sharing sites	Wikipedia, Wikitravel

2.2. Tourism in Greece

Greece is a popular tourist destination, partly due to the ancient cultural history and numerous archaeological sites. The vast majority of tourist expenditure in Greece (96%) comes from leisure tourists rather than business travelers, which has a substantial impact on the Greek economy. Before the coronavirus (COVID-19) pandemic, Greece's total contribution to gross domestic product (GDP) from travel and tourism was around 38 billion euros, but due to the effects of the health crisis, this dropped by more than half in 2020. Tourism significantly improves employment in the country, with approximately 759 (19.8%) thousand jobs available in the Greek travel and tourism industry by 2020 [25]. In 2019, around 34.2 million visitors stayed in Greek accommodations, with two-thirds of visitors being international tourists [26].

2.3 The role of social media in tourism

Several scholars have emphasized how social media's development has transformed the travel and tourism sector as well as effect tourist's behavior [21], [27]–[30]. In the current business climate, social media provides an affordable digital platform for attracting new clients and promoting travel-related goods and services by interacting directly with existing ones [31]. Through consumer review, social networking sites, blogs, and media sharing sites among others, social media provide a platform for travelers to share their experiences and opinions online in the form of text, photographs, and videos; these are increasingly turning into a significant source of travel information for many travelers [9], [32]–[38].

2.4 Influence of social media on the hotel industry

It can be argued that the growing influence of social media in the tourism industry has a positive impact on how hotel guests can share their experiences, creating new content that is easily available and influential to their peers [39], [40]. Decision-making in the hotel industry has undergone a major change as a result of the use of social media at every level of the decision-making process for customers [41]. Social media has an impact on potential tourists since the information shared by previous travelers can mold, direct, and change their initial decisions [21].

However, this review procedure reveals new challenges in the hotel decision-making process [18]. Due to this tendency, social media platforms are crucial for the hotel sector, particularly given that more than half of modern hotel customers check reviews before making a reservation [42]. In order to please clienteles, hotels should improve the content on their social media pages to make them more engaging, educational, interactive, and customer-focused (Leung et al., 2015). Hotel companies need to utilize social media as a tool to improve customer experience and, as a result, engage with customers, attempting to understand their deep desires and needs, offering various services, building brand value and gaining a competitive advantage over rivals [22].

3. Data and Methodology

3.1. Research design and strategy

Descriptive research design is used to conduct the research. More specifically, the descriptive research was used to analyze the participants' demographic characteristics and to gather information regarding the different types of social media platforms used by tourists (the participants) to book a hotel in Greece and the criteria of choosing them.

This study follows a quantitative approach using a survey questionnaire to collect the data. In order to explore, show and describe quantitative data can be used [44], [45]. Quantitative data may be easily generalized, and this is a just another benefit of the approach [46]. A survey design strategy was chosen for the study that is well-suited with research that employs quantitative methodologies [45]. The

measurement is a benefit of this type of approach and this is due to the construction of trustworthy and accurate data replicable by other researchers [46]. Moreover, the benefits of using surveys include the ability to gather data from multiple respondents and the ease with which the results can be compared [47].

A non-probability sampling technique was used for the purposes of this study [46]. The convenience sampling was chosen for this study this one is the least expensive and time-consuming [46]. The sample size of this research is N=187 responses.

3.2 Data collection method

The strategy for administering the survey, during the summer months of 2022, was using a personally administered questionnaire in Greek ports and airports. This study follows the most frequently used design in marketing research, the cross-sectional study that involves information from any given sample of population collected only once [48].

To gather the information needed to respond to the research questions and attain the study's aim a questionnaire was designed [45]. The personally administered questionnaire is a suitable data collection strategy when the survey is limited to a local area. The key benefit of this strategy is that allows the researcher to quickly collect completed responses. On-the-spot clarification is available for any questions that the respondents may have. The researcher can also explain the study's topic and encourage participants to be open and honest in their responses, where also anonymity of responders is high. A drawback of personally administered questionnaires is that the researcher could create bias by explaining the questions [49].

The questionnaire consisted of a letter of consent that was used to inform the participants about the purposes of the research and seek permission from them to take part in the research and the survey questions. The questionnaire included the demographics characteristics of the participants such as the gender and age, and questions relevant the different types of social media platforms and criteria choosing them from the participants to book a hotel in Greece.

In order to investigate the different types of social media platforms used by tourists to book a hotel in Greece, and to address the research question, descriptive analysis was performed using SPSS 25.0 software.

3.3 Ethical aspects of the research

It is crucial to consider and tackle ethical issues when using humans in research [50]. To receive consent, the researcher informed potential respondents and provided them with a thorough explanation of the study. The types of data that will be gathered as well as the research's purpose and aim were specified [51].

It was emphasized that involvement was voluntary, as this is a crucial idea in research [52]. Participants were informed that there was no possible harm to taking part in the research and that they could stop at any time. A cover letter explaining the purpose of the study and the types of questions was given to participants before

they completed the questionnaire [52]. The freedom to leave the research at any moment was emphasized and communicated to the participants. It was underlined that they should be honest in their responses to maintain the integrity of the study [52]. Participants were given the assurance that all information would be handled anonymously and in confidence, preventing them from any potential harm and that they could be updates on the study's findings by emailing the researchers [51], [52].

4. Results

A total of 187 tourist participated in the research providing information about the social media platforms used by tourists in booking hotels in Greece. The participants' frequency was 100 males (53.48%) and 87 females (46.52%) as indicated in Fig. 1.

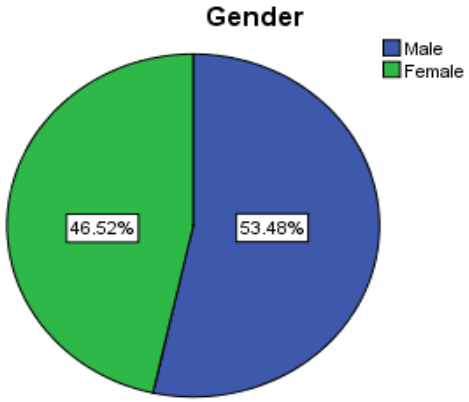

Fig. 1. Gender of the research' participants

The sample was separated into six age groups: (1) 18-24 years; (2) 25-34 years; (3) 35-44 years; (4) 45-54; (5) 55-64 and (6) more than 65. From the findings the majority of the participants (N= 52, 27.81%) aged between 35 and 44 years old, as presented in Fig. 2.

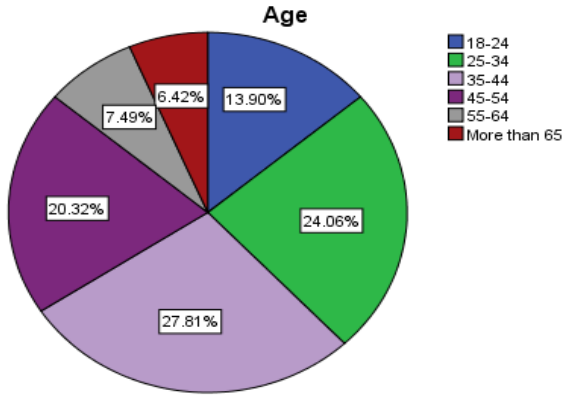

Fig. 2: Age groups of the research' participants

Participants were asked if they have used social media platform to book a hotel in Greece. The majority of the responses (91.40%) replied positive, however 8.60% of the participants they did not use any social media platform to book a hotel in Greece. Thus, 16 participants' replies removed from the initial data (N=187) leaving the data with N=171. From the 16 participants' replies removed 43.80% aged more than 65 years old followed with 31.50% of the participants being aged between 55 and 64 years old.

Moreover, the participants were asked to select from a number of specific social media platforms they used to book a hotel in Greece. As presented in the Figure 3 the majority of the participants used TripAdvisor (26.32%) to book a hotel in Greece, followed by Instagram (21.05%). Nikana.gr was the third from participants prefers with percentage 14.04%. The social media platforms grecia.directbooking.ro and Booking.com have the same selection' percentage (12.28%) from the participants according to the findings. The last two preferences of participants to book a hotel in Greece are the Facebook and grckainfo.com social media platforms with percentages 8.77% and 5.26% respectably.

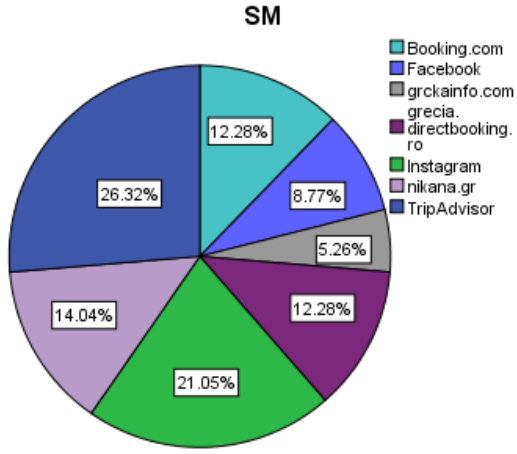

Fig. 3. Social media platforms used from the research' participants

The 5 top destinations according to the findings are Halkidiki (17.00%), Thassos (14.60%), as represented in Table 2.

Table 2. Destinations of the research participants

	Destination	Frequency	Percentage
1	Halkidiki	29	17.0
2	Thassos	25	14.6
3	Lefkada	14	8.2
4	Skiathos	14	8.2
5	Asprovalta	13	7.6
6	Kefalonia	11	6.4

7	Corfu	10	5.8
8	Parga	9	5.3
9	Volos	9	5.3
10	Preveza	8	4.7
11	Kavala	7	4.1
12	Sivota	7	4.1
13	Others	15	8.7
	Total	171	100.0

The sixth survey questions aimed to discover what criteria attracted participants most from the hotels' social media platforms to book the hotel. The "guests' reviews of the hotel" were the first criterion that attracted participants the most from the hotels' social media platforms to book the hotel (68.40%), followed by the "photos and shots of the hotel on the social media platforms" (57.90), as presented in the Figure 4.

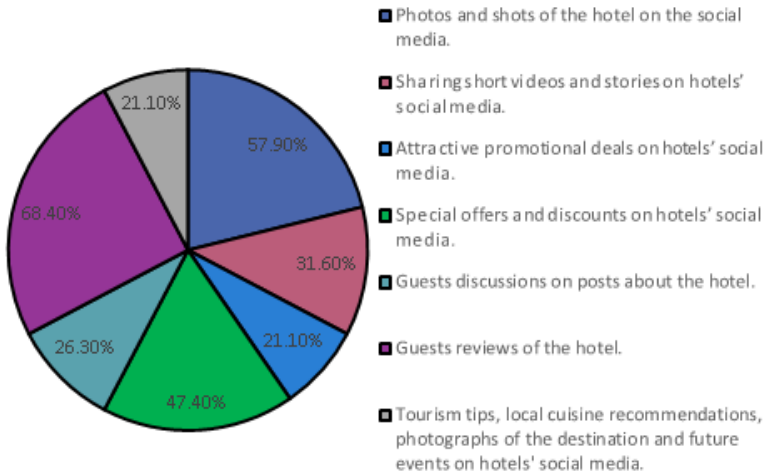

Fig. 4. Social media criteria attracted participants the most from the hotels' social media platforms

5. Conclusion

5.1 Summary of the key findings

This study aims to address the gap in the literature related to the different types of social media platforms used by tourists to book a hotel in Greece. In order to address the research question, the type of social media platforms, and criteria for choosing them by the tourists were investigated.

According to the findings 16 participants (8.60%) did not use any social media platform to book a hotel in Greece with the majority aged more than 65 years (43.80%). Moreover, the majority of participants used TripAdvisor (26.32%), Instagram (21.05%) and Nikana.gr to book a hotel in Greece, followed by

grecia.directbooking.ro and Booking.com receiving lower percentages (12.28%). It was found that Facebook (8.77%) and grckainfo.com (5.26%) were used by fewer participants. The "guests' reviews of the hotel" with percentage 68.40%, "photos and shots of the hotel on the social media platforms" with percentage 57.90% and "Special offers and discounts on hotels' social media" with percentage 47.40% were the three main participants' criteria for choosing social media to book a hotel.

5.2 Key findings relevant to the literature

From the findings, it was identified that 16 participants did not use any social media platform to book a hotel in Greece with the majority aged above than 65 years old. These are somehow in line with previous research that supports that tourist above the age of 40 are less interested in using social media and have a preference toward using a conventional way for booking a hotel [7], [9].

From the findings, it was found that the majority of participants used TripAdvisor and Instagram to book a hotel in Greece which is in line with the current literature mentioning that TripAdvisor and Instagram are the more popular social media and plays an important role in tourists' decision-making process of booking a hotel [7]–[10], [53]. Facebook used by fewer participants and this finding is not in line with previous research highlighting that Facebook is notable social media platform used for hotel selection [7]. From the findings, it is noteworthy to mention the identified new entrants in the tourism arena such us Nikana.gr, grecia.directbooking.ro, grckainfo.com. These new competitors maybe the reason why Booking.com and Facebook received a lower preference from the tourists.

The "guests reviews of the hotel" was the first participants' criterion for choosing social media to book a hotel and this finding is in line with previous research states that in the decision-making process tourists give priority to hotels with a lot of positive reviews [9], [35]. Moreover, "photos and shots of the hotel" on the social media platforms is an important factor in the tourists' decision-making process to book a hotel and this finding is in line with previous research that highlights that tourists' decisions-making process is greatly influenced by online visual content, particularly photos [34], [36], [38]. "Special offers and discounts on hotels' social media" was also an important criterion for tourists to book a hotel and this finding is in line with previous research that supports that promotion marketing strategies increase hotels' bookings [32], [37].

5.3 Recommendations

The study contributes to literature by unveiling criteria that influence tourist's decision-making process using social media to book a hotel in Greece. This study provides an insight for all the relevant stakeholders involved with social media in the travel and hospitality sector more specifically in the hotel field in Greece.

With the rapid change of technology, the user's preferences are constantly changing thus, all the stakeholders should take into consideration which social media platform to include in their marketing strategy to increase brand awareness and brand loyalty. For example, Facebook is one of the commonly used in tourism, however this

study revealed the appearances of competitors such: Nikana.gr, grecia.directbooking.ro and grckainfo.com. Thus, the relevant stakeholders should also consider these new players in the market in order to promote their hotels and to attract new customers. Moreover, guests' reviews and photos play an important role in tourists' decision-making process to book a hotel, thus, all the stakeholders should carefully examine any suggestions for improvement from tourists.

For all the involved stakeholders such as hotels' owners, managers and social media specialists and all the rest interested in the topic the above recommendations should be considered in order to take advantage of social media for the decision-making process to book a hotel in Greece.

5.4 Limitations and future work

Specific limitations of this study are the convenience sampling technique, the cross-sectional study and the bias in the personally administered questionnaire. Convenience sampling was utilized in the study, which limits the generalizability of the results. To lessen the convenience sampling bias, following the recommendation of the literature in the study the data was collected in various carefully chosen times and days of the week aiming toward more balanced sample size [54], that is also large increasing the likelihood to include people from many different subsets of the population [55].

In this study a cross-sectional study was performed to collect the data at a specific timeframe and the choice of this approach can be justified given the amount of time needed to complete this research. A pilot test was performed before the major research to review and improve data collection processes and lessen bias in the personally administered questionnaire.

It is hope that this study with provide practical implications to the stakeholder to take advantage of social media for attracting new customers and promote travel-related goods and services by interacting directly with existing ones. The links between the analyzed criteria were examined in the current study using a quantitative methodology. Future studies can use different methodologies, such as a qualitative or mixed-method approach to investigate criteria that affects tourist's decision-making process using social media to book a hotel in Greece. The scope of future research can be also broadened by conducing future longitudinal studies examining different target groups to examine the tourist's decision-making process using social media to book a hotel in different countries.

References

[1] UNWTO, "Glossary of tourism terms," 2022. https://www.unwto.org/glossary-tourism-terms last accessed 2022/12/07.
[2] V. Paliktzoglou and J. Suhonen, "Facebook as an assisted learning tool in problem-based learning: the Bahrain case," *International Journal of Social Media and Interactive Learning Environments*, vol. 2, no. 1, pp. 85–100, 2014.
[3] "Statista, Number of social media users 2025," 2022. https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/ last accessed 2022/12/06).

[4] N. A. N. Azazi and M. M. Shaed, "Social Media and Decision-Making Process among Tourist: A Systematic Review," *Jurnal Komunikasi: Malaysian Journal of Communication; Universiti Kebangsaan Malaysia Press: Bangi, Malaysia*, vol. 36, no. 4, pp. 395–409, 2020.
[5] Z. Xiang and U. Gretzel, "Role of social media in online travel information search," *Tourism management*, vol. 31, no. 2, pp. 179–188, 2010.
[6] B. Zeng and R. Gerritsen, "What do we know about social media in tourism? A review," *Tourism management perspectives*, vol. 10, pp. 27–36, 2014.
[7] V. Gupta, "The influencing role of social media in the consumer's hotel decision-making process," *Worldwide hospitality and tourism themes*, no. 4, pp. 378–391, 2019.
[8] A. Kavoura and A. Stavrianeas, "The importance of social media on holiday visitors' choices–the case of Athens, Greece," *EuroMed Journal of Business*, vol. 10, no. 3, pp. 360–374, 2015.
[9] E. Varkaris and B. Neuhofer, "The influence of social media on the consumers' hotel decision journey," *JHTT*, vol. 8, no. 1, pp. 101–118, Mar. 2017, doi: 10.1108/JHTT-09-2016-0058.
[10] X. Xu and S. Pratt, "Social media influencers as endorsers to promote travel destinations: an application of self-congruence theory to the Chinese Generation Y," *Journal of travel & tourism marketing*, vol. 35, no. 7, pp. 958–972, 2018.
[11] S. K. Sarkar and B. George, "Social media technologies in the tourism industry: An analysis with special reference to their role in sustainable tourism development," *International Journal of Tourism Sciences*, vol. 18, no. 4, pp. 269–278, 2018.
[12] B. A. Sparks and V. Browning, "The impact of online reviews on hotel booking intentions and perception of trust," *Tourism management*, vol. 32, no. 6, pp. 1310–1323, 2011.
[13] World Travel & Tourism Council, "Travel & Tourism Economic Impact," 2022. https://wttc.org/Research/Economic-Impact last accessed 2022/12/06.
[14] G. W.-H. Tan, V.-H. Lee, J.-J. Hew, K.-B. Ooi, and L.-W. Wong, "The interactive mobile social media advertising: an imminent approach to advertise tourism products and services?," *Telematics and Informatics*, vol. 35, no. 8, pp. 2270–2288, 2018.
[15] D. Buhalis and R. Law, "Progress in information technology and tourism management: 20 years on and 10 years after the Internet—The state of eTourism research," *Tourism management*, vol. 29, no. 4, pp. 609–623, 2008.
[16] J. K. Ayeh, D. Leung, N. Au, and R. Law, "Perceptions and Strategies of Hospitality and Tourism Practitioners on Social Media: An Exploratory Study," in *Information and Communication Technologies in Tourism 2012*, Vienna, 2012, pp. 1–12. doi: 10.1007/978-3-7091-1142-0_1.
[17] M. Spita, E. Peitzika, and S. Chatzi, "Social media adoption among small and medium-sized Greek hotels: a survey about its antecedents and its impact on performance outcomes," *International Journal of Decision Sciences, Risk and Management*, vol. 9, no. 1–2, pp. 23–54, May 2020, doi: 10.1504/IJDSRM.2020.10032367.
[18] P. Z. Baruca and Ž. Civre, "How do guests choose a hotel?," *Academica Turistica - Tourism and Innovation Journal*, vol. 5, no. 1, pp. 75–84, 2012.
[19] N. S. Ahmad, R. Musa, and M. H. M. Harun, "The impact of social media content marketing (SMCM) towards brand health," *Procedia Economics and Finance*, vol. 37, pp. 331–336, 2016.
[20] X. Chen, X. Shen, X. Huang, and Y. Li, "Research on Social Media Content Marketing: An Empirical Analysis Based on China's 10 Metropolis for Korean Brands," *SAGE Open*, vol. 11, no. 4, pp. 1–18, Oct. 2021, doi: 10.1177/21582440211052951.
[21] J. N. Fotis, D. Buhalis, and N. Rossides, in *Social media use and impact during the holiday travel planning process*, in Fuchs, M. Ricci, F. and Cantoni, L (Eds.), Information and Communication Technologies in Tourism 2012, Springer, Vienna, pp. 13-24., 2012.
[22] M. Veríssimo and N. Menezes, "Social media as a tool to enhance customer experience in hospitality industry.," *Portuguese Journal of Marketing*, vol. 38, no. 34, pp. 23–30, 2015.
[23] P. Kotler and K. L. Keller, *Marketing Management*, NY: Pearson Education, Ltd. New York: A Pearson Education Company, 2012.
[24] E. Marinakou, C. Giousmpasoglou, and V. Paliktzoglou, "The impact of social media on cultural tourism," in *Implications of Social Media Use in Personal and Professional Settings*, IGI Global, 2015, pp. 231–248.
[25] "Statista, Travel and tourism in Greece - statistics & facts," 2022. https://www.statista.com/topics/8595/travel-and-tourism-in-greece/ last accessed 2022/12/06.

[26] "Statista - Number of arrivals in tourist accommodation in Greece from 2006 to 2019," 2022. https://www.statista.com/statistics/413222/number-of-arrivals-spent-in-short-stay-accommodation-in-greece/ last accessed 2022/12/06.
[27] T. H. Jung, E. M. Ineson, and E. Green, "Online social networking: Relationship marketing in UK hotels," *Journal of Marketing Management*, vol. 29, no. 3–4, pp. 393–420, Feb. 2013, doi: 10.1080/0267257X.2012.732597.
[28] J. Matloka and D. Buhalis, "Destination Marketing through User Personalised Content (UPC)," in *Information and Communication Technologies in Tourism 2010*, Vienna, 2010, pp. 519–530. doi: 10.1007/978-3-211-99407-8_43.
[29] Assoc. Prof. Dr. R. Yazdanifard and L. Yee, "Impact of Social Networking Sites on Hospitality and Tourism Industries Impact of Social Networking Sites on Hospitality and Tourism Industries," *Global Journal of Human Social Science (E)*, vol. Volume 14, no. 8, pp. 1–6, Jan. 2014.
[30] B. Zeng, "Social Media in Tourism," *Journal of Tourism & Hospitality*, vol. 2, no. 2, pp. 1–2, Jan. 2013, doi: 10.4172/2167-0269.1000e125.
[31] M. E. Styvén and Å. Wallström, "Benefits and barriers for the use of digital channels among small tourism companies," *Scandinavian Journal of Hospitality and Tourism*, vol. 19, no. 1, pp. 27–46, Jan. 2019, doi: 10.1080/15022250.2017.1379434.
[32] A. Ampountolas, G. Shaw, and S. James, "The role of social media as a distribution channel for promoting pricing strategies," *Journal of Hospitality and Tourism Insights*, vol. 2, no. 1, pp. 75–91, 2019.
[33] J. K. Ayeh, N. Au, and R. Law, "Predicting the intention to use consumer-generated media for travel planning," *Tourism Management*, vol. 35, no. C, pp. 132–143, 2013.
[34] S.-C. Chuang, "The Effects of Emotions on the Purchase of Tour Commodities," *Journal of Travel & Tourism Marketing*, vol. 22, no. 1, pp. 1–13, Sep. 2007, doi: 10.1300/J073v22n01_01.
[35] M. Constantoglou and N. Trihas, "The Influence of Social Media on the Travel Behavior of Greek Millennials (Gen Y)," Journal of Tourism and Hospitality Management, vol. 8, no. 2, pp. 10–18, Dec. 2020, doi: 10.15640/jthm.v8n2a2.
[36] E. Ert, A. Fleischer, and N. Magen, "Trust and reputation in the sharing economy: The role of personal photos in Airbnb," *Tourism Management*, vol. 55, pp. 62–73, Aug. 2016, doi: 10.1016/j.tourman.2016.01.013.
[37] R. W. Hamilton, R. T. Rust, and C. S. Dev, "Which features increase customer retention," *MIT Sloan Management Review*, vol. 58, no. 2, pp. 79–84, 2017.
[38] M. Ren, H. Vu, G. Li, and R. Law, "Large-scale comparative analyses of hotel photo content posted by managers and customers to review platforms based on deep learning: implications for hospitality marketers," *Journal of Hospitality Marketing & Management*, vol. 30, no. 3, pp. 1–24, May 2020, doi: 10.1080/19368623.2020.1765226.
[39] A. T. Attila, "The impact of the hotel industry on the competitiveness of tourism destinations in Hungary," *Journal of Competitiveness*, vol. 8, no. 4, pp. 85–104, 2016.
[40] T. H. Jung, M. C. T. Dieck, and N. Chung, "Determinants of hotel social media continued usage," *International Journal of Contemporary Hospitality Management*, vol. 30, no. 2, pp. 1152–1171, 2018.
[41] S. Hudson and K. Thal, "The Impact of Social Media on the Consumer Decision Process: Implications for Tourism Marketing," *Journal of Travel & Tourism Marketing*, vol. 30, no. 1–2, pp. 156–160, Jan. 2013, doi: 10.1080/10548408.2013.751276.
[42] D. C. Taylor, N. A. Barber, and C. Deale, "To tweet or not to tweet: that is the question for hoteliers: a preliminary study.," *Information Technology and Tourism*, vol. 15, no. 1, pp. 71–99, 2015.
[43] X. Y. Leung, B. Bai, and K. A. Stahura, "The marketing effectiveness of social media in the hotel industry: a comparison of Facebook and Twitter.," *Journal of Hospitality & Tourism Research*, vol. 39, no. 2, pp. 147–169, 2015.
[44] M. Saunders, *Research Methods for Business Students*, 6th edition. Harlow, England ; New York: Pearson Custom Publishing, 2012.
[45] M. Saunders, P. Lewis, and A. Thornhill, *Research Methods for Business Students*. Prentice Hall, 2009.
[46] A. Bryman, *Social Research Methods*, 4th edition. Oxford ; New York: Oxford University Press, 2012.

[47] R. L. Miller and J. D. Brewer, Eds., *The A-Z of Social Research: A Dictionary of Key Social Science Research Concepts*, 1 edition. London ; Thousand Oaks, Calif: SAGE Publications Ltd, 2003.
[48] J. W. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 3rd edition. Thousand Oaks, Calif: SAGE Publications, Inc, 2008.
[49] U. Sekaran and R. Bougie, *Research methods for business: A skill building approach*. New York: John Wiley & Sons, 2016.
[50] H. Lune and B. L. Berg, *Qualitative research methods for the social sciences*, 9 edition. Harlow, England Munich: Pearson, 2017.
[51] H. Noble and J. Smith, "Issues of validity and reliability in qualitative research," *Evidence-based nursing*, vol. 18, no. 2, pp. 34–35, 2015.
[52] J. Fleming and K. E. Zegwaard, "Methodologies, Methods and Ethical Considerations for Conducting Research in Work-Integrated Learning.," *International Journal of Work-Integrated Learning*, vol. 19, no. 3, pp. 205–213, 2018.
[53] I. B. Hubner, C. Carina, E. Ennelis, V. Natalia, and J. Juliana, "The Use of Social Media on Tourist Decision Making in Determining Hotel Selection," *International Journal of Social and Management Studies*, vol. 2, no. 3, pp. 161–171, 2021.
[54] A. Galloway, "Non-Probability Sampling," in *Encyclopedia of Social Measurement*, K. Kempf-Leonard, Ed. New York: Elsevier, 2005, pp. 859–864. doi: 10.1016/B0-12-369398-5/00382-0.
[55] B. Formplus, "Convenience Sampling: Definition, Applications, Examples," 2022. https://www.formpl.us/blog/https//www.formpl.us/blog/convenience-sampling last accessed 2022/12/04.

Risk Disclosure as a Way to Increase the Informative Value of Corporate Reporting for Stakeholders

Irina V. Zenkina[1][0000-0003-1020-4050]

[1] Financial University under the Government of the Russian Federation, Russia

https://doi.org/10.33847/2686-8296.4.2_5

Received 21.10.2022/Revised 21.11.2022/Accepted 23.12.2022/Published 28.12.2022

Abstract. The article is devoted to the study of risk as a category of accounting and reporting and substantiation of directions for comprehensive disclosure of risks in order to increase the informative value of corporate reporting for stakeholders. The article shows the development of approaches to the definition of risk and provides an updated definition of risk in accordance with modern concepts. A classification of risks is proposed in the context of the concept of multiple capitals and the concept of sustainable development, which is relevant to the task of adequate disclosure of information about risks. It is demonstrated that the modern legal regulation of accounting, standards and guidelines in the field of corporate reporting assigns an important role to risks. Based on an empirical study, the recommended limits for the disclosure of information about risks by organizations in the framework of ensuring the transparency of reporting are determined. The factors stimulating economic entities to disclose information on the risks of sustainable development in corporate reporting are considered. An assessment is made of the impact of digital tools and technologies on the ability to predict, assess and disclose risks in corporate reporting, as well as on the effectiveness of decisions of organizations' stakeholders.

Keywords: risk classification, ESG risks, risk disclosure, corporate reporting.

1. Introduction

The process of making business decisions by economic entities and their stakeholders is closely related to risk. This is especially true for strategic decisions, since they are characterized by high uncertainty and a long-term planning horizon, require significant investments of various resources, and therefore are associated with large-scale potential consequences of the implementation of such decisions.

Modern realities cause the actualization of known and the emergence of new risks that require study, comprehensive assessment and regular monitoring in terms of consequences for both a particular organization and its stakeholders. Along with the aggravation of economic risks, including market, financial, operational risks, there is an increase in the importance of environmental, social risks and governance risks. Climate risks are coming to the fore and are becoming a powerful factor in business decision-making at all levels.

In the context of increasing the role of risk management in cost management, there is a need for transparency in corporate reporting regarding the reflection of information about risks. However, the question of the essence of risk remains debatable to date, a generally accepted classification of risks of organizations' activities has not been developed, and unified approaches to their disclosure in corporate reporting have not been formed. In turn, this is reflected in the weak integration of risks into the decision-making process by stakeholders. In this regard, it becomes important to develop approaches to the disclosure of risks in corporate

reporting, assessment and integration of risks into the decision-making process in order to ensure their effectiveness.

The purpose of this article is to study risk as a category of accounting and reporting and substantiate the directions for comprehensive disclosure of risks as a factor in increasing the informative value of corporate reporting for stakeholders.

2. Literature Review

The issue of risk disclosure in corporate reporting is in the focus of attention of the scientific community, practitioners, corporations, regulators and public organizations.

Aspects of risk assessment and risk management of organizations are currently one of the most relevant and widely discussed in the professional environment.

Various approaches to the classification of business risks are presented in the works of Damodaran A. [1], Crouhy M. [2], Pike R. [3], Blank I.A. [4], Kogdenko V.G. [5].

The publications by Efimova O.V. [6], Rozhnova O.V. [7], Zenkina I.V. [8] are devoted to the issues of risk disclosure in the financial and non-financial reporting of organizations in the interests of stakeholders in order to increase the validity of their decisions.

A wide range of researchers such as Bauer R., Khan D. [9], Goss A., Roberts G. S. [10], Attig N., Gul S. E., Gedami O., Su D. [11], Chava S. [12], Giraporn P., Giraporn N., Beprasert A., Chang K. [13], Zenkina I.V. [14] examines sustainability risks and demonstrates the correlation between ESG risks and financial performance of an organization.

3. Methodology

The concept of risk in modern economic literature is characterized by a wide range of definitions.

According to ISO 31000 Risk Management (2018) [15], risk is "the effect of uncertainty on objectives". Impact refers to the deviation of an outcome from an expected outcome and can be "positive, negative, or both, and may relate to, create, or lead to opportunities and threats".

The Institute of Internal Auditors, the world's leading standards body for internal auditing, defines risk as "the uncertainty of an ongoing event that could affect the achievement of objectives".

According to the Business Analysis Body of Knowledge (BABOK), recognized worldwide as a professional standard, risk is the impact of uncertainty on the value of changes implemented in an organization, decisions being implemented, or on the organization itself [16].

The COSO Enterprise Risk Management (ERM) Framework defines risk as the likelihood of occurrence of events that can have an impact on the achievement of strategic and business objectives [17].

It is important to note the ongoing gradual transformation of the traditional approach to defining risk as "the likelihood that an event will occur and adversely affect the achievement of the mission and business goals of the organization". Its reasonable alternative is the approach based on the understanding of risk as "a measure of an organization's exposure to losses at the level of uncertainty" [18]. The main difference between these approaches is that from the traditional point of view, the maximum risk occurs with the highest probability of losses, and from the modern point of view, the maximum risk corresponds to the most significant degree of negative consequences, even with an insignificant probability of their occurrence.

Based on advanced approaches in the field of risk management and taking into account the high relevance of the sustainable development agenda, the concept of risk can be clarified as the degree of an organization's exposure to the consequences of the occurrence of events that can affect the implementation of the strategy, the achievement of business goals and sustainable development goals, the ability to create value over time.

Risk management is usually associated with the management function and the organizational units that implement it. However, its principles are applied in the implementation of all functions and at all levels of management, and the content is not limited to the identification of relevant risks of the organization's activities and involves the active management of them.

According to the COSO ERM Framework, risk management is an integrated system that includes the culture, competencies and practices associated with the strategy setting and performance management process that an organization relies on to create, maintain and increase value.

The COSO ERM Framework contains 5 components and 20 principles grouped according to them (Table 1).

Table 1. Risk Management Principles

Governance & Culture	**Strategy & Performance Objective-Setting**	**Performance**	**Review & Revision**	**Information, communication and reporting**
1. Exercises Board Risk Oversight 2. Establishes Operating Structures 3. Defines Desired Culture 4. Demonstrates Commitment to Core Values 5. Attracts, Develops, and Retains Capable Individuals	6. Analyzes Business Context 7. Defines Risk Appetite 8. Evaluates Alternative Strategies 9. Formulates Business Objectives	10. Identifies Risk 11. Assesses Severity of Risk 12. Prioritizes Risks 13. Implements Risk Responses 14. Develops Portfolio View	15. Assesses Substantial Change 16. Reviews Risk and Performance 17. Pursues improvement in Enterprise Risk Management	18. Leverages Information and Technology 19. Communicates Risk Information 20. Reports on Risk, Culture, and Performance

Source: Authoring based on COSO ERM [17]

The principles related to the "Governance and Culture" component form the overall framework for risk management. The principles associated with the components "Strategy and Performance Objective-Setting", "Performance", "Review and Revision" directly determine the procedure for analyzing, assessing risks and making management decisions based on them. The principles relating to the "Information, Communication and Reporting" component indicate the importance of ensuring the transparency of information about risks and reflect the main directions of relevant activities within the framework of risk management.

Most of the principles of the risk management concept directly affect the processes of identifying, analyzing, assessing risks and disclosing information about risks in corporate reporting in the interests of stakeholders (Table 2).

Table 2. The content of the key principles of risk management, regulating the analysis and disclosure of information about risks

Principle	Key characteristics
Analyzes Business Context	An organization should consider the external and internal business environment in the process of developing a strategy to realize its mission, vision and core values. The impact of business conditions on the risk profile can be assessed in the context of past, current and future events.
Defines Risk Appetite	Decisions related to the choice of strategy and the justification of risk appetite are not connected by linear relationships, when one of them precedes the other. Approaches to identifying risk appetite are chosen by organizations based on the characteristics of their activities. The best approach is to define risk appetite in the context of risk profile and risk capacity.
Evaluates Alternative Strategies	An organization should analyze alternative strategies, assess the risks and opportunities of each option. When choosing a development strategy, the management of an economic entity should take into account the risk profile and risk appetite.
Formulates Business Objectives	Objectives are defined at various levels, but should correlate with the overall strategy of the organization. At the same time, the degree of efficiency in achieving goals is determined by the boundaries of acceptable deviations from the goals set.
Identifies Risk	An organization should identify the risks associated with the implementation of the strategy, the achievement of business and sustainability goals. In this regard, it is necessary to form a consolidated list of risks and subsequently determine which risks are relevant. Risks can be structured into separate categories, which the organization determines at its discretion. At the same time, risks should be determined at the level of all business processes and business functions.
Assesses Severity of Risk	Risks are assessed in terms of impact and likelihood. The probability of a risk can be expressed by an expert assessment, a quantitative assessment and the frequency of the risk. The description of the risk includes the analysis of its factors and consequences based on quantitative and qualitative approaches. A risk map is used as a tool for graphical visualization of risk materiality.
Prioritizes Risks	An organization should identify the most significant risks in order to select an adequate strategy and rational allocation of resources within the framework of risk management. Risk prioritization is determined according to certain criteria, such as risk adaptability, risk complexity, risk impact strength, risk impact sustainability, and recovery speed after risk realization.
Implements Risk	Justification of the strategy involves taking into account business conditions, established goals, risk prioritization, risk appetite and risk materiality. The choice is based on five main risk management strategies, including:

	acceptance, avoidance, transference, mitigation and exploitation. In cases where the level of risk is too high and the risk response strategy is unacceptable, the organization should review its strategic and operational objectives.
Develops Portfolio View	An organization should determine how the organization's residual risk profile matches its risk appetite. In this regard, there are several approaches to portfolio risk analysis depending on the degree of integration of risk management with business management: - minimal integration (focus on significant risks) - the organization identifies and evaluates discrete risks. Focus on significant risk events; - limited integration (focus on risk categories) - the organization forms a structured risk database. The risk portfolio represents a list of risks grouped into categories; - partial integration (focus on the risk profile) - the organization shifts the focus to business goals and the risks associated with their achievement; - full integration (comprehensive view of risks) - the organization focuses on strategy, business goals and sustainable development goals. Risks are identified at all levels of decision making.
Assesses Substantial Change	An organization needs to monitor changes in ongoing operations and assess their impact on business strategy and risk profile.
Reviews Risk and Performance	The principle lies in the need to conduct a risk analysis within the framework of business analysis and its system integration into the management of the organization's activities. The key issues for risk and performance analysis are: - What are the risks and how they can affect the efficiency of the organization? - What is the level of risks and how are they assessed? - How adequate are the levels of key risks to achieve the objectives of the organization? - Has the organization achieved its objectives with the expected performance?
Pursues improvement in Enterprise Risk Management	An organization should strive to continually improve the effectiveness of risk management at all levels.
Reports on Risk, Culture, and Performance	An organization needs to disclose information about risks in corporate reporting.

Source: Authoring based on COSO ERM [17]

Of particular note is the "Reports on Risk, Culture, and Performance" principle, according to which economic entities need to disclose information about risks in corporate financial and non-financial reporting.

This principle closely correlates with the Global Management Accounting Principles, which include the principle of communication, the principle of using relevant information, the principle of analyzing the influence of various factors on the value of the company, and the principle of management based on trust. This is due to the critical role of management accounting in generating data, including information on business risks, for the preparation of integrated reporting in order to meet the information needs of stakeholders.

The modern legal regulation of accounting, existing standards and guidelines in the field of corporate reporting assign a significant role to risks.

In particular, the procedure for disclosing information about risks in the financial statements of economic entities is given in IFRS 7 "Financial Instruments: Information Disclosure", document of the Ministry of Finance of Russia No. PZ-9/2021 "On Disclosure of Risks of an Organization's Business Activities in Annual Financial Statements".

It should also be noted that the Bank of Russia has developed recommendations on the disclosure by public joint-stock companies of non-financial information related to the activities of such companies [19]. These recommendations provide a framework for boards to consider ESG factors and related risks and opportunities, as well as sustainability issues, to create long-term value. They can be used depending on the scale and specifics of the organization's activities, the implemented corporate practices and business processes, the speed and depth of ongoing changes related to sustainable development issues in various industries.

In this regard, it is difficult not to agree with the opinion of O.V. Efimova, who believes that reporting can become a tool for creating a reputation as an informationally transparent company only if these reports provide complete and reliable information not only about its financial position, results obtained, ability to generate cash flows, but also about existing risks of activity [6].

4. Results

For high-quality disclosure of information about risks in corporate reporting, first of all, a classification of the risks of the organization's activities that meets modern ideas is needed.

In this respect, on the basis of the classifications presented in the works of Damodaran A. [1], Crouhy M. [2], Pike R. [3], Blank I.A. [4], Kogdenko V.G. [5], and an empirical study of the practice of disclosing information about risks in corporate reporting, it seems informative and analytically valuable to disclose the following list of risks (Fig. 1).

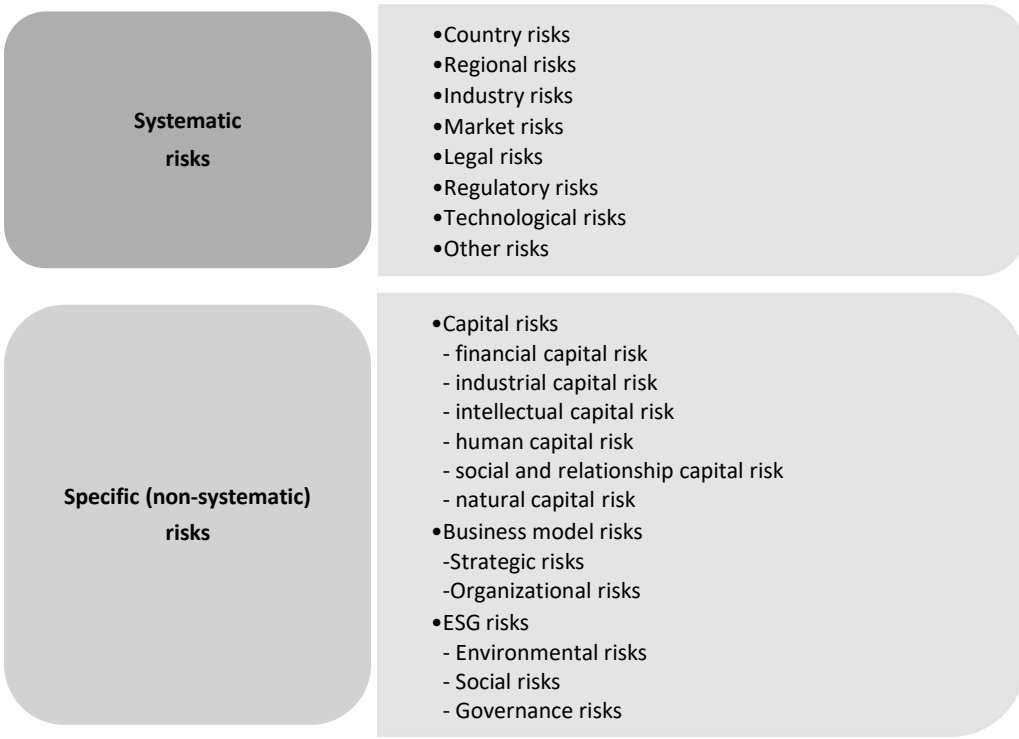

Fig. 1. Classification of risks to be disclosed in corporate reporting
Source: Compiled by the author

1. **Systematic or external risks** affecting all economic entities:
- *Country risks* - risks caused by the economic, political and social conditions of the countries in which the organization operates;
- *Regional risks* - risks associated with the economic, political and social conditions of individual administrative or geographical regions;
- *Industry risks* - risks specific to a certain type of activity, reflecting the probability of losses of the organization due to changes in the economic state of the industry, depending on the degree of changes both within the industry and in comparison with other industries;
- *Market risks* - risks of unfavorable changes in the price environment and demand in the goods and services market, unfavorable dynamics of exchange rates and interest rates in the financial market;
- *Legal risks* - risks associated with changes in currency and customs regulations, tax legislation;
- *Regulatory risks* - risks related to non-compliance by organizations with existing rules and regulations, in particular, aggravated in a pandemic amid quarantine restrictions;
- *Technological risks* - risks that are caused by the development of technologies and the possibility of losing consumers due to a decrease in competitiveness.
- *Other risks* - the risk of developing an unfavorable epidemiological situation, military operations, etc.

Systematic risks associated with the external business environment of organizations are undoubtedly important for assessing the long-term sustainability of an economic entity in the framework of stakeholder decision-making. However, the

priorities for stakeholders in terms of determining the prospects for interaction are specific risks or risks of the internal business environment that affect a particular organization. According to A. Damodaran, specific risk factors are the source of 75-80% of the risk to which a publicly traded company is exposed [1].

In this connection, special attention should be paid to the identification, analysis, assessment, monitoring and reporting of risks specific to a particular economic entity.

2. **Specific risks or risks of the internal business environment**, cover:

1) Risks associated with capital, that is, the business resources that the organization has ***(Capital risks).***

2) Risks due to the organization's business model (Business model risks).

3) Risks of sustainable development (ESG risks).

Capital risks are risks due to the availability, quality and accessibility of financial, industrial, intellectual, human, social and relationship, natural capitals, determined by the probability of loss or decrease in the value of capital in the short, medium and long term, as well as characterized by indicators of the organization's impact on capital. Due to the fact that financial capital is a source of formation of all other types of tangible and intangible capital of an organization, financial risks occupy a central place in this group of risks.

According to the fair statement of V.G. Kogdenko, "the company's risk management system should be aimed at protecting the most valuable capital, the list of which is determined by the International Framework for Integrated Reporting" [5].

Financial risks are risks of inefficient use of financial capital, decrease in financial stability, performance and investment attractiveness of the organization, including:

- *The risk of limited access to capital* - the risk associated with the company's inability to accumulate equity and debt capital in sufficient volume and on acceptable terms;
- *Credit risk* – the risk of counterparties failing to fulfill their debt obligations in full and on time;
- *Liquidity risk* – the risk that an organization will not have sufficient funds to meet its financial obligations.

Financial risks may also include *investment risks*, as well as *tax risks*.

Business model risks cover the following groups of risks depending on the decision-making area:

- *Strategic risks* - risks associated with the development and justification of the organization's strategy, which are decisive for other risks, since the organization's activities and the risks arising from its implementation depend on the choice of business strategy.
- *Operational risks* - the risks of losses due to inconsistencies or errors in business processes, employee actions, systems or as a result of external events, including:
  - *Risks of internal business processes* - risks of reducing the efficiency of *operating activities;*
  - *Supply chain risk* - the risk of reducing the efficiency of supply chains, in particular, reducing the reliability, response and flexibility of the supply chain, increasing costs and reducing the efficiency of asset management in the supply chain, resulting in financial losses for the organization;
  - *Personnel risks* - risks associated with the lack of required personnel, insufficient qualifications of personnel, low involvement and loyalty of employees, as well as exceeding official powers by personnel, including corporate fraud, corruption and other offenses.

Sustainability risks (ESG risks) are risks caused by organizations' non-compliance with existing rules, rules and principles regarding ESG efficiency, which

may affect the performance of the company and its stakeholders, lead to negative legal, environmental, social and economic consequences, including:

- *Environmental risks* – risks associated with the impact of the organization's activities on the environment, including physical and transitional climatic risks;
- *Social risks* - risks related to the impact of the organization's activities on the social sphere, including the observance of human rights;
- *Governance risks* - risks of deterioration in the perception by stakeholders of the reliability and business attractiveness of the organization [8].

Given the importance of a comprehensive reflection of risks, sustainable development risks should occupy a special place in corporate reporting, which is due to the recognition at the international level of the crucial role of taking into account sustainable development factors in assessing the activities of economic entities and making investment decisions [14]. This was the reason for including this group of risks in the proposed classification.

As Sergey Shvetsov, First Deputy Chairman of the Bank of Russia, notes "an assessment of these risks is necessary, among other things, because over time they can transform into financial risks, and their disclosure may become a requirement in the near future, as soon as uniform international standards for non-financial indicators appear" [20].

In accordance with modern concepts, economic entities should strive to achieve a balance of economic efficiency, environmental responsibility, social responsibility and management efficiency, as well as ensure the transparency of corporate reporting regarding the disclosure of information on sustainable development and the risks of implementing the organization's strategy.

The current trend is the integration of sustainable development risks into the process of substantiating and making investment decisions, called ESG integration. Investors are interested in minimizing ESG risks as indicators of responsible business practices due to their significant impact on financial performance and, in general, on the viability of the business and the effectiveness of investment projects.

At the same time, it should be emphasized that it was the urgent need of the investment community for an expanded framework for disclosing information about financial and non-financial drivers of value creation and the risks associated with them that served as the main incentive for the creation and global promotion of the concept of integrated reporting.

According to the International Integrated Reporting Framework, the main elements of the content of an integrated report that are fundamentally related to each other and are not mutually exclusive include:

1. Organizational overview and external environment.
2. Governance.
3. Business model.
4. Risks and opportunities.
5. Strategy and resource allocation.
6. Performance.
7. Outlook.
8. Basis of preparation and presentation.

The <IR> Framework give an important place to risks in the process of value creation, preservation or erosion.

The "Risks and Opportunities" element is prioritized for the informative and analytical value of integrated reporting, especially as in the revised <IR> Framework (January 2021) value creation, which is the central concept of this reporting format, includes cases of preservation and erosion of value.

The "Risks and Opportunities" element requires disclosure in the integrated report of information about the risks and opportunities that affect the organization's

ability to create value in the short, medium and long term, as well as how the organization manages them. In particular, the report may include the following information:

- key risks and opportunities inherent in the organization related to the impact of the organization on the financial, industrial, intellectual, human, social and relationship, natural capitals and with the further availability, quality and accessibility of relevant capitals;
- a specific source of risks and opportunities (internal, external or combined);
- assessment of the probability of occurrence of risks or the extent of their impact in case of realization;
- actions taken to mitigate or manage key risks and to create value from key opportunities, including an indication of their associated strategic objectives, strategies, policies, targets and key performance indicators;
- the organization's approach to any real risks in the short, medium and long term that are of paramount importance to maintaining the organization's ability to create value and that can have very serious consequences, even if the probability of their occurrence can be considered quite low.

However, in addition to an entity that reports in accordance with financial reporting requirements, risks, opportunities, and performance outcomes arising from or associated with other entities/stakeholders that are not part of the reporting entity but that have a significant impact on its ability to generate value, relate to key aspects that define the perimeter of the integrated report, that is, the boundaries within which factors are considered relevant for their inclusion in the organization's integrated report.

Thus, although there are opportunities for disclosing information about risks in financial and non-financial reporting, today it is the integrated report as an advanced corporate reporting format that provides the most adequate level of risk disclosure and, therefore, has the greatest informative value for interested users.

Despite the revision of the <IR> Framework carried out in January 2021, their updating is superficial and did not introduce significant changes into the previous version of the document, adopted in December 2013. At the same time, economic entities require clearer guidance regarding the preparation of an integrated report and the disclosure of material information in it, including information about risks.

The lack of detailed, understandable and specific recommendations regarding information on risks to be reflected in corporate reporting causes a widespread problem of insufficient completeness and quality of disclosure of such information by organizations. According to O.V. Rozhnova, information about risks in the reporting space is usually difficult to trace, it is almost impossible to understand what happened to the risks noted earlier, whether the measures taken to level them were effective, how the risks of this company correlate with the risks of competitors, partners, creditors and investors [7].

An empirical study of the best practices of more than 30 Russian organizations that are leaders in the field of transparency of corporate reporting, such as Gazprom, Norilsk Nickel, Rosatom, using data from the Russian Union of Industrialists and Entrepreneurs [21], allows us to recommend the wide dissemination of best practices in corporations and the inclusion of the following risk data in their reporting.

1. Processes of the risk management system:

- risk identification;
- risk assessment;
- risk management;
- risk monitoring.

2. Organizational model of the risk management system:

- strategic, tactical, operational levels of risk management;

- organizational and methodological support of the risk management system at the level of the corporation and departments;
- risk owners - management bodies responsible for risk management;
- internal control bodies of the risk management system.

3. Key risk indicators:

- a list of systematic and specific risks relevant to the activities of the organization;
- assessment of the level of key risks.

A risk radar can serve as an analytically valuable tool for reflecting information about the key indicators of corporate risks in an integrated report.

4. Correlation of key risks with the strategic goals of the organization - a reflection of the strategic goals that are affected by the corresponding risks.

5. Risk management practices:

- approaches to risk assessment and risk management;
- results of risk management;
- dynamics of risks, including options for increasing, decreasing and maintaining a stable level.

The perimeter of the integrated report, necessary for the adoption of balanced and comprehensively justified management decisions by stakeholders, involves the disclosure of information about the organization's risk management practices.

Along with the above aspects of the risk management system (RMS), corporations are encouraged to disclose in an integrated report the use of digital risk assessment and management technologies as an indicator of the digitalization of the RMS.

The current digital transformation of business provides new ways to solve problems, create a unique experience of interaction with customers and employees, and accelerate business efficiency. The digitalization of risk management has a positive impact on the quality of decisions made by stakeholders.

The high level of digital transformation of risk management, including the functions of accounting, reporting and analysis, implies, first of all, that the employees of the organization have advanced digital skills and competencies, as well as the orientation of the risk management process towards the active use of big data, digital tools and technologies. Due to the digitalization of risk management, forecasting capabilities are expanding, the accuracy of risk assessment and the quality of risk disclosure in corporate reporting are increasing, and, consequently, the effectiveness of strategic and tactical decisions made by stakeholders of organizations is achieved.

5. Conclusion

Thus, being a traditional economic and statistical category, risk is currently being actively developed as an accounting category, becoming a full-fledged object of corporate reporting, in which investors as key stakeholders and other stakeholders of the organization are interested.

In this regard, modern corporate reporting, the advanced format of which is integrated reporting, demonstrates the relationship between the strategy, the business goals defined by it and the goals in the field of sustainable development and should comprehensively reflect the risks and their impact on the implementation of the strategy and the performance of the organization.

The article clarifies the definition of risk and proposes a classification of the risks of the organization's activities, taking into account the provisions of the concept of six capitals and the concept of sustainable development. Proposals are given for disclosing risk information in an integrated report, including the processes and organizational model of the risk management system, key risk indicators, their correlation with the strategic goals of the organization, risk management practices

and the use of digital technologies. It is argued that the integration of ESG risks into the process of substantiating business decisions positively affects their quality and contributes to the achievement of the organization's strategic goals and its effective interaction with stakeholders.

Comprehensive identification and assessment of risks, their consideration in decision-making are in the sphere of mutual interests of economic entities and stakeholders. In turn, this serves as an impetus for improving corporate reporting, causing an increase in its informative and analytical value at the present stage.

Acknowledgments

The author is grateful to the reviewers and the editor for their contribution to preparing the paper for publication.

References

1. Damodaran A. Strategic Risk Management: Principles and Methods. Moscow, Vil'yams Publ (2017).
2. Crouhy M., Galai D., Minasyan V.B., Mark R.M. The Essentials of Risk Management. Moscow, Yurait Publ (2023).
3. Pike R., Neale B., Linsley Ph. Corporate Finance and Investment: Decisions and Strategies. St. Petersburg, Piter Publ (2006).
4. Blank I.A. Financial Security Management of the Enterprise. Kiev, El'ga Publ (2013).
5. Kogdenko V.G. Investigating Company Risks within the Framework of the Stakeholder Approach to Analysis. Economic Analysis: Theory and Practice, vol. 17, iss. 6, pp. 1051–1072 (2018). URL: https://doi.org/10.24891/ea.17.6.1051
6. Efimova O.V. Analytical Aspects of Disclosure of Financial Statements Explanatory Information. Auditor's Journal, no. 7, pp. 38–50 (2015).
7. Analytical Capabilities of Integrated Reporting and their Use for Strategic Decisions. Moscow: KnoRus (2020).
8. Zenkina I.V. Methodical approaches and tools of company's sustainable development analysis. Economic Analysis: Theory and Practice, vol. 18, no. 9, pp. 1667–1686 (2019). URL: https://doi.org/10.24891/ ea. 18. 9.1667
9. Bauer R., Khan D. Environmental Management and Credit Risks. ECCE Working Paper, University Maastricht, The European Centre for Corporate Engagement (2010).
10. Goss A., Roberts G.S. Impact of Corporate Social Responsibility on the Cost of Bank Loans. Journal of Banking and Finance, No. 35, pp. 1794 -1810 (2011).
11. Attig N., Gul S.E., Gedami O., Su D. Corporate Social Responsibility and Credit Ratings. Business Ethics Journal, No. 117, pp. 679 – 694 (2013).
12. Chava S. External Environmental Factors and the Cost of Capital. Management Science, No. 60 (9), p. 2223 – 2247 (2014).
13. Giraporn P., Giraporn N., Beprasert A., Chang K. Does Social Responsibility Improve Credit Ratings? Place of origin of goods. Financial management, No. 43 (3), p. 505 – 531 (2014).
14. Zenkina I.V. The Impact of Regulatory Risks of ESG Integration on the Sustainable Development of Power Companies. National Interests: Priorities and Security, vol. 17, iss. 4, pp. 624–648 (2021). URL: https://doi.org/10.24891/ ni.17.4.624
15. ISO 31000:2018 – Risk management — Guidelines. URL: https://www.iso.org/standard/65694.html
16. Business Analysis Body of Knowledge (BABOK v.3).
17. COSO Enterprise Risk Management – Integrating with Strategy and Performance (2017).
18. Krepysheva A.M., Sergievskaya A.A., Storchevoy M.A. Definition and Measurement of Risk in Compliance Management. Strategic Decisions and Risk Management, no. 2, pp.150-159 (2020). URL: https://doi.org/10.17747/2618-947X-2020-2-150-159
19. Bank of Russia Letter No. IN-06-28/49, dated July 12, 2021, "On Recommendations for Disclosing Non-Financial Information by Public Joint Stock Companies Related to the Activities of Such Companies". URL: https://www.cbr.ru/StaticHtml/File/117620/20210712_in-06-28_49.pdf

20. ESG disclosure: theory and practice of implementing new recommendations of the Central Bank. Joint stock company (2021). URL: https://group.interfax.ru/interfax/about/smi/esg-raskrytie-teoriya-i-praktika-vypolneniya-novykh-rekomendatsiy-tsb/
21. RUIE Indexes in the Field of Sustainable Development, Corporate Responsibilities and Reporting (ESG-indexes) "Responsibility and Openness" and "Vector of Sustainable Development".
URL:https://rspp.ru/upload/iblock/e07/efghhr37sx3rkf35uznh3pd9t1ihq4tz/Prezentatsiya_ESG_indeksy-RSPP-2022.pdf

Aims and Objectives

Published online by ICS two times a year, Journal of Digital Science (JDS) is an international peer-reviewed journal which aims at the latest ideas, innovations, trends, experiences and concerns in the field of digital science covering all areas of the scholarly literature of the sciences, social sciences and arts & humanities. The main topics currently covered include: Artificial Intelligence Research; Digital Economics, Education, Engineering, Finance, Health Care.

The main goal of the journal is the effective dissemination of original incites/results generated by the human brain and presented/reflected in articles using modern information/digital technology.

Editorial Board

Contact information

Website: https://ics.events

Email: conf@ics.events