Tech Science Press

An Advanced Analysis of Cloud Computing Concepts Based on the Computer Science Ontology

Paweł Lula[1], Octavian Dospinescu[2,*], Daniel Homocianu[2] and Napoleon-Alexandru Sireteanu[2]

[1]Krakow University of Economics, Krakow, Poland
[2]Alexandru Ioan Cuza University, Iasi, 700706, Romania
*Corresponding Author: Octavian Dospinescu. Email: doctav@uaic.ro
Received: 20 August 2020; Accepted: 29 September 2020

Abstract: Our primary research hypothesis stands on a simple idea: The evolution of top-rated publications on a particular theme depends heavily on the progress and maturity of related topics. And this even when there are no clear relations or some concepts appear to cease to exist and leave place for newer ones starting many years ago. We implemented our model based on Computer Science Ontology (CSO) and analyzed 44 years of publications. Then we derived the most important concepts related to Cloud Computing (CC) from the scientific collection offered by Clarivate Analytics. Our methodology includes data extraction using advanced web crawling techniques, data preparation, statistical data analysis, and graphical representations. We obtained related concepts after aggregating the scores using the Jaccard coefficient and CSO Ontology. Our article reveals the contribution of Cloud Computing topics in research papers in leading scientific journals and the relationships between the field of Cloud Computing and the interdependent subdivisions identified in the broader framework of Computer Science.

Keywords: Cloud computing scientific literature; cloud related concepts; CSO ontology

1 Introduction

In-depth scientific studies of cloud computing have a relatively recent history. Thus, the research carried out by Chiregi et al. [1] and Ibrahim et al. [2] highlights that journals published by Elsevier, Springer, IEEE, Emerald, Taylor, and Wiley have been concerned with this field since 2010.

Cloud Computing has developed as a critical innovation in the field of ICT that can revolutionize the way information resources are consumed and delivered. Thus, according to Yu et al. [3], in developing economies, this innovation is considered a new way that can generate a new information infrastructure with a real potential for future economic growth. The authors pointed out that, in the case of China, the development of the cloud computing industry was achieved from an early stage (2008) through the co-evolution of technological and institutional infrastructures, leading to a preliminary cloud ecosystem. This process involved a wide range of different actors, from the government to business. The interaction between these actors influenced the development of the cloud computing industry based on each

participant's interests. It was the case of the period between 2008 and 2016. This study concludes that the development of cloud computing technology can contribute to economic progress only through partnerships between the government and the business environment in order to identify market requirements and manage related risks.

Regarding the expansion of cloud computing players nationally and globally, Kshetri et al. [4] analyzes the determinants of such an evolution. The conclusion reached is that the modeling of the cloud computing industry and market was possible through the action of contradictory, conflicting, and paradoxical forces. Therefore, the following facilitators and inhibitors resulted: Standards and standardization institutions, regulatory ones, and legal regulations on cyber-control.

Ali et al. [5] shows that in developing countries, there is a tendency to reform e-Government in an attempt to provide easily accessible and high-quality services to citizens. Although the intention is commendable, there are still many challenges as the cost growth rate that is difficult to estimate and control. Managing the data, information, knowledge, and hardware infrastructure is an expensive component and creates other difficulties. The main obstacles and challenges regarding the e-Government cloud are lack of data control [6], security, and privacy [7], access authorization, data leakage, and system failure [8]. These challenges can lead to e-government project failures. Therefore, a solution is needed to overcome them, and Cloud Computing plays a vital role in solving these problems.

Nowadays, the cloud computing sector is a growing field of many providers engaged in a "digital revolution" that will make classic IT models obsolete in the next ten years. Although still evolving, many circumstances can generate anti-competitive or monopolistic behavior in the cloud industry market [9]. Vendors may arrange peculiar or exclusive negotiations and may refuse to share technical information on compatible products. Innovation can also be restricted by pricing and monopolistic behavior, ultimately leading to a reduction in competition. In addition to competition law, other rules have a powerful impact when competing in the cloud computing services industry. Concentration regulations can have a direct influence on the process of controlling market concentration in the CC industry. In terms of mergers and competition law, one of the main issues to be considered concerns the concept of interoperability. This concept is particularly important in the field of cloud computing, as it has an immediate impact on openness and competition, with an instant effect on standardization and intellectual property rights. Taking into account the studies carried out by Song [9] and Walden et al. [10], it appears that, although the legislative framework somehow lags behind the technological progress, competition law still plays an important role so that dominant market players cannot abuse of their position. The ongoing use of competition law usually means the number of analyses and investigations related to software and hardware platform monopolies. These laws may extend to points of sale in cloud computing infrastructures.

Novais et al. [11] have studied the impact that Cloud Computing and its technologies have on the supply chain. The analysis of specialized literature shows that there was a relationship of influence between the adoption of cloud computing and the technological integration of partners and business processes in the supply chain. Also, the use of Cloud Computing in the supply chain has positive effects on the integration of information and financial flows. Topics and lines of research that have crystallized in recent times include the relationship between cloud computing and logistics, commercial integration, and manufacturing process integration. Research results [12–16] show that Cloud Computing supports the integration of supply chain processes and activities because it considerably improves scalability, flexibility, agility, adaptation to change, and supply chain planning. D'Arcy et al. [17–19] show that commercial aspects and trends go beyond the classic limits of the supply chain by switching to mobile cloud computing.

Cloud Computing is also to consider from the perspective of intra-organizational and inter-organizational integration. Thus, in terms of intra-organizational integration, Cloud Computing can be

connected with technologies and systems such as ERP (Enterprise Resource Planning) [20,21], and Radio Frequency Identification [22,23]. Research results show that Cloud Computing, together with intra-organizational technologies, can reduce information distortions within organizations and increase the efficiency of internal procurement processes. Regarding the inter-organizational integration, Chen et al. [24] and Singh et al. [25] have focused mainly on the relationship between Cloud Computing and web technologies. The results of the studies show that the efficiency and competitiveness of the supply chain can meliorate by integrating web 2.0 technologies with Cloud Computing. In the same direction, Camara et al. [26] show that Cloud Computing can improve the way resources are shared and distributed among members of the supply chain, leading to an increase in the dynamics of collaborative systems.

Battleson et al. [27] and Liu et al. [28] indicate that the flexibility of Cloud Infrastructure can improve the ability of a company to adapt and its skill to quickly integrate new IT applications, which fundamentally changes an organization's IT framework and the way IT resources are installed and used. In many industries, scalability is a fundamental factor for a company to respond quickly to market changes.

Liu et al. [28] centralized the literature and concluded that the main features/dimensions of the IT infrastructure fall into two types: flexibility and integration. Most studies highlight the flexibility of IT infrastructure and its importance for business. Thus, flexibility [29–31] refers to concrete issues such as rapid development and development of significant applications, hardware and software modularity, scalability, and compatibility of infrastructure components, connectivity, and standardization of networks and platforms in organizations. On the other hand, integration refers to issues such as the exchange of information between different locations, products, or services, exploiting synergistic opportunities between the components of a business, data consistency, functional integration of applications, adaptability, and connectivity.

Jeyaraj [32] define Cloud Computing as an archetype that allows access to a usual pool of cloud computing resources in an on-demand or pay-per-use model. Cloud computing offers more benefits to users and organizations in terms of capital expenditures and operating expenses savings. According to Noor et al. [19], mobile cloud computing promises several benefits, such as increased battery life, scalability, and reliability. However, there are still challenges to face to enable ubiquitous deployment and adoption of cloud computing. Some of these challenges include security, confidentiality and trust, bandwidth and data transfer, data management and synchronization, energy efficiency, and heterogeneity. Despite the benefits, some barriers restrict the use of cloud computing. Security is an important issue that always matters. The lack of this vital feature leads to the negative impact of computational archetype, leading to personal, ethical, and financial damage. Security challenges are analyzed on three levels: Computational [33], communication, and data [34]. The security of cloud computing environments is becoming increasingly important in the context of the Internet of Things and the need for integration [35]. With the evolution of ubiquitous computing, everything connects everywhere, so these concepts have been studied extensively in the literature [36]. However, intrusions and vulnerabilities will be more frequent due to the complexity of the systems and the difficulty of controlling each access attempt.

Specialized studies [37] show that human society is facing an unprecedented technological evolution that will powerfully change the way we interact with the world around us and the way we program applications. Mobile computers and related applications have had a significant impact. Another potential area of research is the Internet of Things (IoT) that aims to develop a smart network of interconnected devices. There have been numerous emerging research paradigms regarding their respective fields of research and their intersections. These include Mobile Cloud Computing (MCC), cloud computing, fog computing, IoT cloud computing, Mobile Edge Computing (MEC), WoT, and SWoT (Semantic WoT). It happens quite often that a concept refers to several paradigms or to a single paradigm that is defined by several terms. As a result, we can say that these paradigms' definitions are not standardized.

According to a systematic study conducted by Androcec et al. [38], there are four major categories of interest for cloud computing ontologies, which have emerged in connection with the literature. Thus, the proportions are as follows: Cloud resource and service description—25%, Cloud security, and privacy—8%, Cloud interoperability—13%, and Cloud service discovery—54%. On the other hand, Al-Sayed et al. [39] consider that a standardized ontology is still missing today.

From a technological point of view, the cloud computing approach branches in several directions. In this regard, Boukerche et al. [40] distinguishes between concepts such as Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). These concepts serve to develop new areas as cloud networks and services. There are already computing implementations for automotive industry services. Some examples are network as a service, storage as a service, and cooperation as a service. The implications are straightforward in terms of vehicle management in cloud computing: data centers, traffic management, internet vehicles, urban surveillance, security, and infotainment.

Cloud computing applications also tend to develop for the chemical industry [41], contributing decisively to the professional interpretation of data and information. Other applications and areas of the Internet of Things category, on which Cloud Computing has a particular impact, are the following [35]: Smart transportation solutions, remote patient monitoring, home sensors and sensors in airports, sensors to monitor the problems that may occur in engine operation, and smart grids. Applications of interest are also in areas such as storage over the internet, internet overhead, internet applications, and energy efficiency.

Du et al. [42] believe that cloud computing is shaping the world of cyber technologies while evolving as the principal computing infrastructure for sharing resources like services, applications, and platforms. This approach is called "X as a service" and brings important functionalities and current economic benefits. But in cyberspace, cloud computing is limited because its services can only be accessed remotely. Still, it may be necessary to access them closer to the physical location of the actual activity. However, there are many situations related to service requests, in which cloud computing only helps to a small extent due to "cyber-limitation". Since SOA has become more and more popular, this new architecture has served in the development of applications in the field of robotics. Chen [43] specified the use of SOA concepts to generate new composite embedded systems and robotic applications; they mentioned that they even built a prototype system. The SOA robotic architecture relies on the expansion of cloud computing—RaaS (Robot as a Service). In a RaaS system, many robotics units provide various services to consumers, playing the roles of the service provider, service broker, and service client [44].

As Cloud Computing developed, so did the issue of CC governance. Thus, according to Bounagui et al. [45], through CC governance, organizations can have real control over the services provided by the CC infrastructure. Currently, there are several different approaches to CC governance. Thus, the He [46] model supports organizations that want to manage all their IT services using cloud computing. The model takes into account the business objectives and aligns the CC governance with them and the need to manage CC assets and services. The model's purpose is to provide a benchmark for cloud service providers that best meet the needs of end-users. The author divided CC governance into five main areas: strategic planning, organizational alignment, service lifecycle management, policy management, and service level management. For each of these, the processes and activities needed to ensure effective governance of the CC are very clearly detailed.

In terms of future directions and trends in cloud computing, Varghese et al. [47] identifies several directions for cloud computing research on information ecosystem management strategies, the development of distributed architectures, improving the reliability of cloud systems, the impact of system development on sustainability, and advanced security. The new architectures and facilities will have to serve the stated requirements of the Internet of Things philosophy, in line with the challenges of

processing large volumes of data. At the same time, it is becoming increasingly clear that current systems developed by human programmers will evolve into a new generation of self-learning systems.

Hakak et al. [48] shows that gamification has gained considerable interest in educational circles due to its ability to enhance learning among students. In the future, they expect gamification to go beyond traditional learning, resulting in issues such as scalability and modernization of learning modules. A viable solution to these problems would be to combine gamification and cloud computing. However, the capacity of cloud computing is still in the early stages of development. Potential applications for cloud gamification are: courses related to Natural Language Processing, virtual reality, distributed learning system, mobile learning, and real-time learning skills.

One of the challenges is scientifically explained by Alles [49]. Thus, this author concludes about Cloud Computing stating that AIS (Association for Information Systems) is unable to establish a distinct role in the clear separation between cloud computing and other fields. The AIS community fails to make a clear distinction between cloud computing as a subject of proprietary research and cloud computing as a method of information sharing. An approach that extends this topic may add value to the research, but at the same time, there are risks associated with dissipating the original concept. This conclusion can be extended even beyond the area of interest of the AIS.

The main objective of this paper is to implement a methodology for identifying the most relevant concepts related to a particular key topic of interest, considering mainly academic writings as a source of evaluation and determination. In our study, we started with the Cloud Computing topic. Then, we applied the Computer Science Ontology on ISI Web of Science (WOS) data (https://apps.webofknowledge.com) in the form of high-quality publications, as well as the evolution over time of the frequency of publications both on the topic of interest and on the related ones.

2 Research Methodology

2.1 Goals

The authors decided to perform the analysis of:

- The contribution of Cloud Computing topics from research papers in top scientific journals;
- The relationships between the Cloud Computing domain and the correlated fields defined in Computer Science.

2.2 Main Assumptions

The authors formulated the following research hypotheses:

- The general analysis should stand on the exploratory text analysis of the abstracts of the papers published in the Web of Science database. The latter contains only peer-reviewed articles of exceptional academic quality;
- The in-depth analysis should rely on ontologies; the authors decided to use the Computer Science Ontology[1] to support the analysis process.

2.3 Analysis Process

The analysis process included the following steps:

1. Data retrieval using a web scraping technique;
2. Building a model for representing the Computer Science Ontology;

---

[1] https://cso.kmi.open.ac.uk/home

3. The ontology-based annotation of abstracts;
4. The analysis of the importance of topics related to the concept of Cloud Computing;
5. The ontology-based analysis of the relationships between the concepts identified in the abstracts.

2.3.1 Data Retrieval

With the development of Big Data computing technology, most documents in several fields became digital, and we now have new methods and approaches to obtain quantitative research results. According to Kim et al. [50], text mining is the technology used to classify, group, extract, search, and analyze data to find patterns or features in a set of unstructured or structured documents written in natural language. We propose a method for extracting information on the subject of Cloud Computing using text mining through web scraping from the ISI WOS website and analyze concepts extracted using the Computer Science Ontology.

As a synthesis of the methodological stages, we highlight:

- The download from ISI WOS of almost 10,000 complete records with abstracts, titles, and keywords;
- The development of a custom Node.Js crawler to use this data in our research and obtain up to eight key concepts and corresponding scores for each record;
- The aggregation of all scores and identification of those concepts being the most strongly related in terms of the score using the SQL language;
- The download from ISI WOS of the frequency time-series for topics corresponding to all the relevant concepts identified previously.

Puppeteer is a Node library API that allows us to control Chrome heedlessly. Headless Chrome is a way to run the Chrome/Chromium browser without actually running Chrome/Chromium, and we can automate anything we do on these browsers, such as emulating a keypress, a click, and so on. With the Puppeteer library, we can crawl Clarivate Analytics and extract the relevant articles we need for our analysis. Access to Clarivate was due to the E-information portal (Fig. 1).

```
10    const page = await browser.newPage(); // create a new page
11
12    await page.goto ("https://www.e-nformation.ro/", {waitUntil: 'networkidle0'}); // go to target URL
13    await page.setViewport({ // set viewport size
14        width: 1000,
15        height: 700
16    })
17    await page.waitFor('a[href="javascript:void(0);"]');
18    await page.click('a[href="javascript:void(0);"]');
19    await page.waitFor(3000);
20
21    await page.waitFor('input[name="login_user_name"]'); // automatically login to e-nformation
22    await page.type('input[name="login_user_name"]', 'alexandru.sireteanu@feaa.uaic.ro');
23    await page.type('input[name="login_pwd"]', '*****');
24    await page.click('input[name="doLogin"]');
25    await page.waitFor(3000);
26
27    await page.goto('http://am.e-nformation.ro/login?url=http://www.webofknowledge.com', {waitUntil: 'networkidle0'});
      // go to target URL
28    await page.setViewport({
29        width: 1000,
30        height: 700
31    })
32    await page.waitFor(3000);
```

Figure 1: Custom crawler to access the E-information portal

To retrieve the content of each item from the article page, we used the $eval() method from the Cherrio library and the page.evaluate() method. The latter allows us to extract the desired results and catch errors, as seen in Fig. 2.

```
 96       const artName = await page.$eval('#records_form > div > div > div > div.l-column-content >
          div > div.title > value', value => value.innerText.trim()).catch((e)=>{console.error(e)});
          // extract the name of article
 97
 98       const artAbstract = await page.evaluate(() => { // extract the abstract of article
 99           try {
100               var data = $("div:contains('Abstract')").next('.FR_field').text();
101               return data;
102           } catch(err) {
103               console.error(err);
104               reject(err.toString());
105           }
106       });
```

Figure 2: Method for retrieving the abstracts

Finally, we scraped all the details of the articles on each page and returned them in .csv format, as seen in Fig. 3.
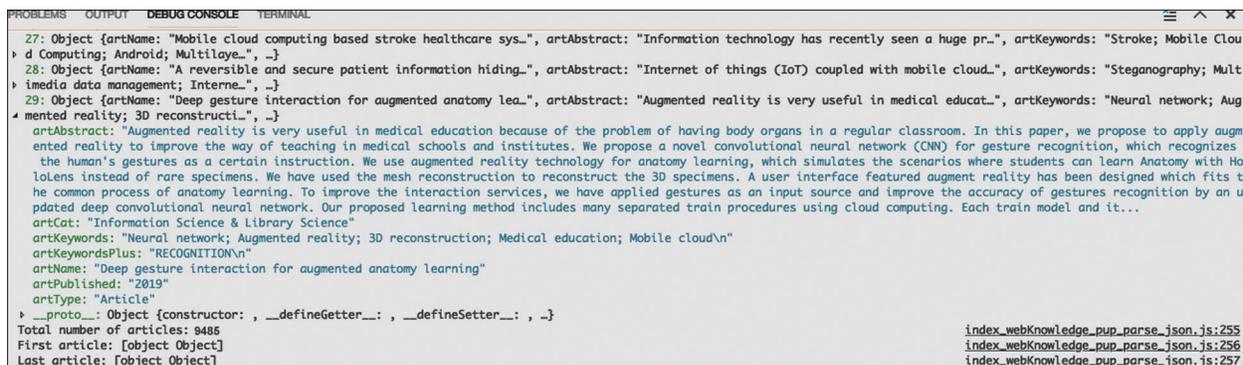
Figure 3: Scraping details of articles in .csv format

We queried WOS for the exact phrase of "Cloud Computing" (search by topic). Then we filtered the SCIE and SSCI WOS categories meaning publication only in journals that currently have IF and AIS > 0 (see Fig. 4), according to the results of journal-focused searches provided by the Journal Citation Reports—JCR online application at https://jcr.incites.thomsonreuters.com.
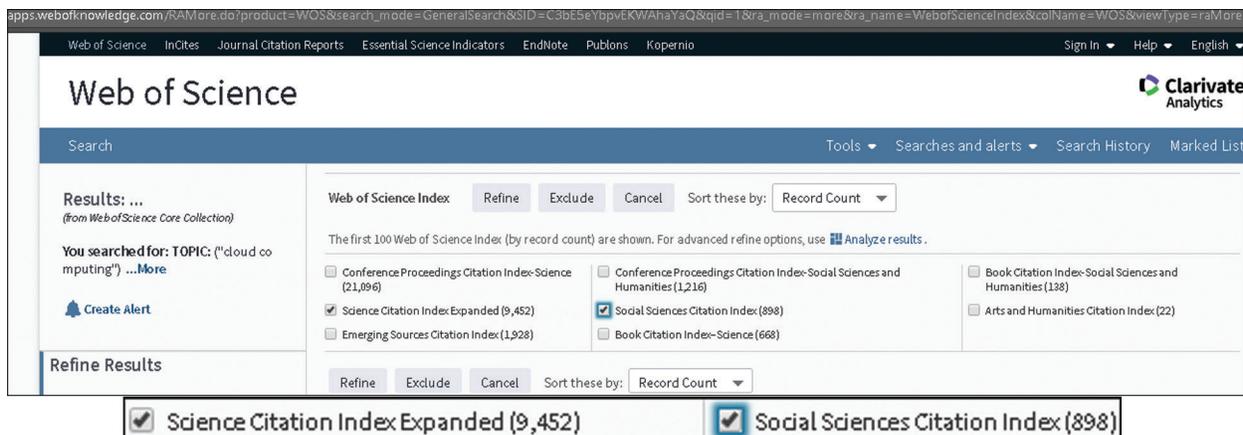
Figure 4: Manually filtering (ISI WOS) for the most relevant articles published in journals with IF and AIS > 0 to verify the results obtained automatically, using the API

In the next step, we filtered only consistent contributions (see Fig. 5—without reprints, retractions, editorial materials, etc.). We finally exported almost 9500 records from the ISI WOS online platform (full record format as Windows tab-delimited) in a .txt file, which can be read by any spreadsheet program. These records were extracted in 19 volumes of 500 lines each (almost 50 MB of text data) using the same online platform above.

Figure 5: Additional filters for article type

Then we concatenated for each record four parts consisting of title, authors' keywords, additional journal keywords, and abstract and then we removed the copyright texts from the results using a combination of text-oriented functions in Excel for each resulting text block. We took the copyright texts out because they generated false results. Such raw blocks served for analysis with the Computer Science Ontology.

2.3.2 Building a Model Representation of the Computer Science Ontology

The Computer Science Ontology [51] is an automatically built ontology that covers the field of computer science with about 14 thousand topics related to 163 thousand semantic relationships. The license of the CSO ontology is Creative Commons Attribution 4.0 International License (CC BY 4.0).

For the project described here, we considered two types of semantic relationships:

- The *superTopicOf* relationship that connects two different topics and indicates that the first is a parent (direct ancestor) of the other;
- The *preferentialEquivalent* relationship that defines alternative concepts and uses one of them as the primary label; this relationship allows the unification of different terms that refer to the same notion.

For describing the relationships between concepts corresponding to the *superTopicOf* connection, we used graph models with:

- Vertexes corresponding to all those concepts that appear in the *superTopicOf* relationships in the CSO ontology (the names of the latter used as vertexes' identifiers);
- Edges that correspond to the *superTopicOf* relations (these connections lead from ancestors to descendants defined by the *superTopicOf* predicate).

For building the graph model, we used the igraph package in the R language [52]. Then, for every vertex in the graph above, a list of alternatives was defined. A list of alternative concepts for a given vertex was created by identifying all those concepts related to a given one using the *preferentialEquivalent* relationship.

In the third step of the analysis, for each vertex, a list of patterns was created. These patterns were in the form of phrases (word sequences). We generated them to facilitate the identification of concepts. In the CSO ontology, we constructed the concept names by combining words that describe a particular term using a "_" sign as a separator (for example, a "software engineering" concept represented by a "software_engineering"). Non-alphabetic characters were decoded in the names using hexadecimal codes. To build patterns, the name

of a given such concept and the names of alternative ones were transformed by replacing all appearances of the "_" signs with spaces and replacing the hexadecimal values with the corresponding characters.

In the next stage of the process described here, each word in each pattern received the "#" symbol on the front of it. The purpose was to indicate that these words are mandatory (meaning that all words in a pattern must appear in a given part of an abstract to annotate it with the name of the corresponding concept). Details of the notation used to identify the patterns are available in Section 2.3.3.

Finally, the concepts together with the corresponding patterns were transformed into a text file in *yaml* format and saved to a file on disk.

All the knowledge used in the consecutive stages of analysis was stored in the graph describing the CSO ontology and in the *yaml* file containing all the concepts taken from the CSO ontology and the patterns that allow their identification in the textual documents.

2.3.3 Abstracts' Annotation

We used an annotation technique proposed in Lula et al. [53]. The data stored in the *yaml* file served for performing the annotation process. The content of the *yaml* file acts as an associated table in which the concept name serves as the key, and a list of patterns forms a value connected to a given key. For example, for an "ontology_alignment" concept, a list of patterns has the following form:

["#ontology #alignment", "#ontology #matching", "#ontology #mapping"]

Next, for each pattern, we built its alternative version. In it, we used all the words in their elementary form. This lemmatization process relied on the use of the Apache OpenOffice dictionaries and the *hunspell* package for the R language [54]. Then, we performed an analysis of abstracts. During this step, we executed the following sequence of operations for each abstract:

1. The division of the text of a given abstract into phrases considering the positions of the punctuation marks.

2. For similarity calculation considering each pattern, for each phrase obtained in step 1, we used the following algorithm: First, we checked the presence of all mandatory words. In the absence of such a word or words, the measure of similarity was zero. If all required words were present, then a Jaccard coefficient between a set of words in a phrase (T) and another one in a pattern (P) was calculated using Eq. (1):

$$s_O = \frac{|T \cap P|}{|T \cup P|} \tag{1}$$

3. Then, we transformed all the words in a phrase taken from an abstract into their primary form (T'). Next, we used Eq. (2) to calculate a Jaccard coefficient between the set T' and a lemmatized version of a pattern (P'):

$$s_L = \frac{|T' \cap P'|}{|T' \cup P'|} \tag{2}$$

4. As a final version of a similarity measure, a maximum calculated for the values defined in the two steps presented above was defined using Eq. (3):

$$s = \max(s_O, s_L) \tag{3}$$

As a result of the above process for each abstract, we identified a set of patterns for which the *s* measure was more than 0.

Having all the patterns identified for a given abstract, the last stage of the annotation process will take place. At this stage, we assigned them to the appropriate concepts. And for each one, a measure of its contribution to a given abstract was calculated using Eq. (4):

$$c_{i,j} = \max_k \left(s_{i,j,1}, s_{i,j,2}, \ldots, s_{i,j,n_j} \right), \ k = 1, \ldots, n_j \tag{4}$$

where:

$c_{i,j}$—contribution of the jth concept in the ith abstract;

$s_{i,j,k}$—contribution on the kth pattern assigned to the jth concept in the ith abstract;

n_j—number of patterns we allot to the jth concept.

Elements $c_{i,j}$ form a contribution matrix (Eq. (5)).

$$C = \left[c_{i,j} \right] \tag{5}$$

This matrix reveals the contribution of each concept (column of the matrix) in each abstract (row of the matrix).

2.3.4 Analysis of the Importance of Topics Related to the Concept of Cloud Computing

In the CSO ontology, the representation of the cloud computing area uses the concept of "cloud_computing" and its descendants.

First, we found all the direct descendants (children) of the "cloud_computing" concept. According to the CSO ontology, the latter has 34 direct descendants. They form a list L (Eq. (6)).

$$L = [l_1, l_2, \ldots, l_n] \tag{6}$$

where l_i is the ith direct descendant (child) of the "cloud_computing" concept.

For each element of the L list, we built a set that contains a given concept and all its descendants. Let's assume that D_i is a set containing the l_i concept and all its descendants. Also, we considered that, for the kth abstract, the contribution of the l_i concept results from Eq. (7).

$$p_{k,i} = \max(C[k, D_i]) \tag{7}$$

where $C[k, D_i]$ is a vector composed of elements located in the matrix C, from row k and columns identified by elements of the D_i set.

Elements $p_{k,i}$ form a matrix P in which rows represent abstracts, and columns represent concepts from the L list.

Next, a vector v (Eq. (8)) was defined.

$$v_k = \sum_i p_{k,i} \tag{8}$$

The value $v_k > 0$ means that the kth abstract contains references to topics related to the cloud computing area. Finally, the contribution of each concept from the L list in the whole *corpus* was defined (Eq. (9)).

$$P_i = \frac{\sum_k \mathrm{sgn}(p_{k,i})}{\sum_k \mathrm{sgn}(v_k)} \tag{9}$$

where sgn(.) is the signum function defined as indicated in Eq. (10).

$$\mathrm{sgn}(x) = \begin{cases} -1 & if\ x < 0 \\ 0 & if\ x = 0 \\ +1 & if\ x > 1 \end{cases} \tag{10}$$

This measure P_i informs about the significance of the l_i concept in the whole *corpus*.

2.3.5 Analysis of Relationships between Concepts Appearing in Abstracts

Within the project's framework, we also proposed a method for analyzing the relationship between two concepts in the CSO ontology.

Let's assume that for a given kth abstract, a relationship between two concepts, l_i and l_j, should be calculated. To achieve this goal, we used the following algorithm:

1. For the concept l_i, we created a set D_i containing the concept l_i and all its descendants. By analogy, a set D_j for the concept l_j.

2. Two vectors (Eqs. (11) and (12))

$$v_{k,i} = C[k, D_i] \tag{11}$$

$$v_{k,j} = C[k, D_j] \tag{12}$$

resulted, where $C[k, D_i]$ is a vector composed of elements located in the matrix C, from row k and columns identified by elements of the D_i set.

3. The set D is defined as indicated in Eq. (13).

$$D = D_i \cup D_j \tag{13}$$

4. Then, the vector $w_{k,i}$ with $|D|$ elements is defined (Eq. (14)). In this vector, we identified successive components considering the structure of the D set. We copied the elements from the $v_{k,i}$ vector to the $w_{k,i}$ one at positions resulting from the same parts of the D set. The remaining ones, identified by labels appearing in the D set and not appearing in D_i, are completed by values equal to 0 (Eq. (15)). Formally it may be expressed as:

$$w_{k,i}[D_i] = v_{k,i}[D_i] \tag{14}$$

$$w_{k,i}[D_i'] = 0 \tag{15}$$

where D_i' is a complement of the D_i set relative to the D set.

Similarly, the vector $w_{k,j}$ with $|D|$ elements is defined (Eqs. (16) and (17)).

$$w_{k,j}[D_j] = v_{k,j}[D_j] \tag{16}$$

$$w_{k,j}[D_j'] = 0 \tag{17}$$

5. Then, we calculated the Jaccard coefficient for the vectors $w_{k,i}$ and $w_{k,j}$. The latter served as a measure of similarity between the l_i and l_j concepts as observed in the kth abstract (Eq. (18)):

$$sim_k(l_i, l_j) = \frac{\sum_a \min(w_{k,i}[a], w_{k,j}[a])}{\sum_a \max(w_{k,i}[a], w_{k,j}[a])} \tag{18}$$

Having the similarity measures $sim_k(l_i, l_j)$ calculated for every abstract, the aggregated one, for the whole *corpus*, can be defined (Eq. (19)):

$$sim(l_i, l_j) = \frac{\sum_{k=1}^{K} \text{sgn}(sim_k(l_i, l_j))}{K} \tag{19}$$

where K represents the number of abstracts in the given *corpus* and means the ratio of abstracts in which concepts related to l_i and l_j ones, were identified simultaneously.

3 Results

3.1 Analysis of the Importance of Topics Related to the Cloud Computing Concept

We used the method presented in Section 2.3.4.

In the CSO ontology, the "cloud_computing" concept has 34 direct descendants. They form the L list:

L = [application_execution, autonomic_computing, cloud_service_providers, security_and _privacy_issues, mobile_cloud_computing, high_availability, cloud_data, multi-tier_applications, cloud_storage, storage_services, storage_resources, map-reduce, virtual_machines, virtualizations, resource_ provisioning, service_level_agreements, security_challenges, cluster_computing, job_execution, cloud_infrastructures, utility_computing, computing_paradigm, distributed_computing_environment, data-intensive_application, computing_resource, computing_environments, cloud_computing_ environments, cloud_environments, computing_services, computing_technology, cloud_computing_ services, software_as_a_service, it_infrastructures, cloud_security]

For the set of abstracts studied during the analysis, we presented the contribution of the direct descendants of the Cloud Computing concept in Fig. 6.
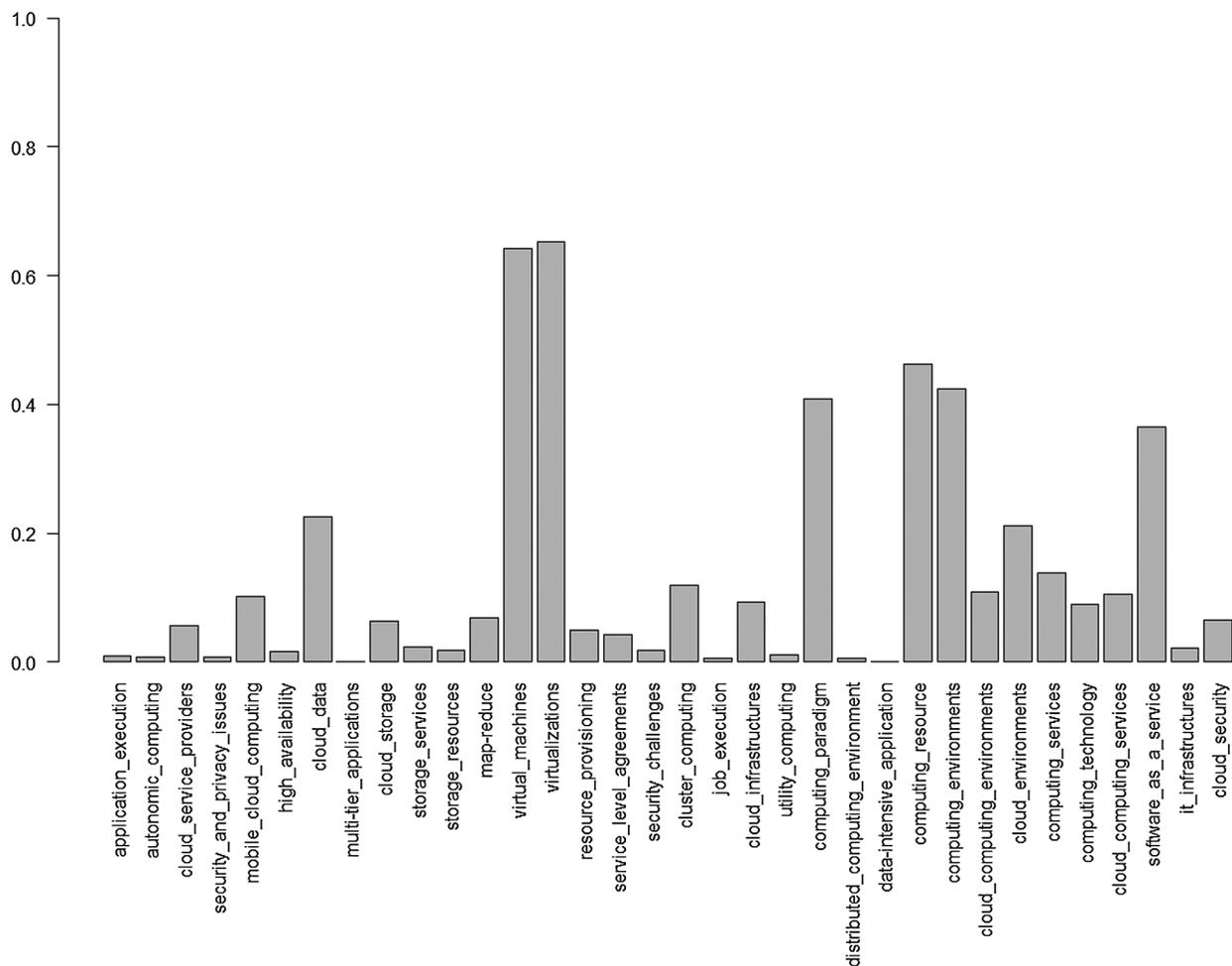
Figure 6: The contribution of the direct descendants of the cloud computing concept

3.2 Analysis of the Relationships between the Concept of Cloud Computing and Others Related to Computer Science

We performed the analysis using the method presented in Section 2.3.5. First, the relations between the Cloud Computing concept and the main fields within Computer Science were analyzed. The values obtained during the analysis indicate a proportion of papers that have simultaneous references to Cloud Computing or its descendants and the selected fields of computer science. We have shown the results in Fig. 7.
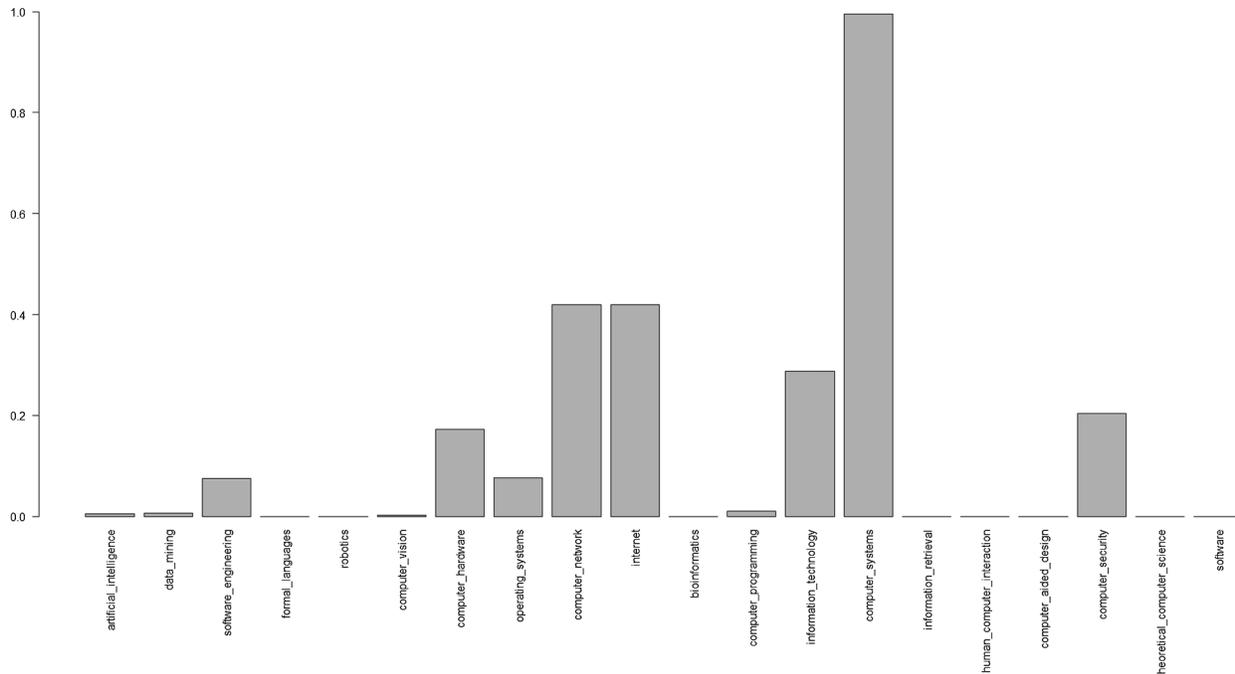
Figure 7: The Significance of relationships between cloud computing and main fields of the computer science concept

In the next part, we present relevant relationships between Cloud Computing and the first five concepts with the highest rates (Fig. 7).

The first category of emphasized relationships refers to Cloud Computing and the main areas under the umbrella of the Computer Network concept. We emphasize the proportion of papers with simultaneous references to Cloud Computing and the subfields of the Computer Network in Fig. 8.

The second category of emphasized relationships focuses on Cloud Computing and the main areas within the Internet concept. We indicated the proportion of papers with simultaneous references to Cloud Computing and Internet subfields in Fig. 9.

The third category of emphasized relationships refers to Cloud Computing and the main areas under the umbrella of the concept of Information Technology. We synthesized the proportion of papers with simultaneous references to Cloud Computing and Information Technology subfields in Fig. 10.

The 4th category of emphasized relationships is about Cloud Computing and the main areas within the Computer Systems concept. We indicated the proportion of papers with simultaneous references to Cloud Computing and the subfields of the Computer System in Fig. 11.
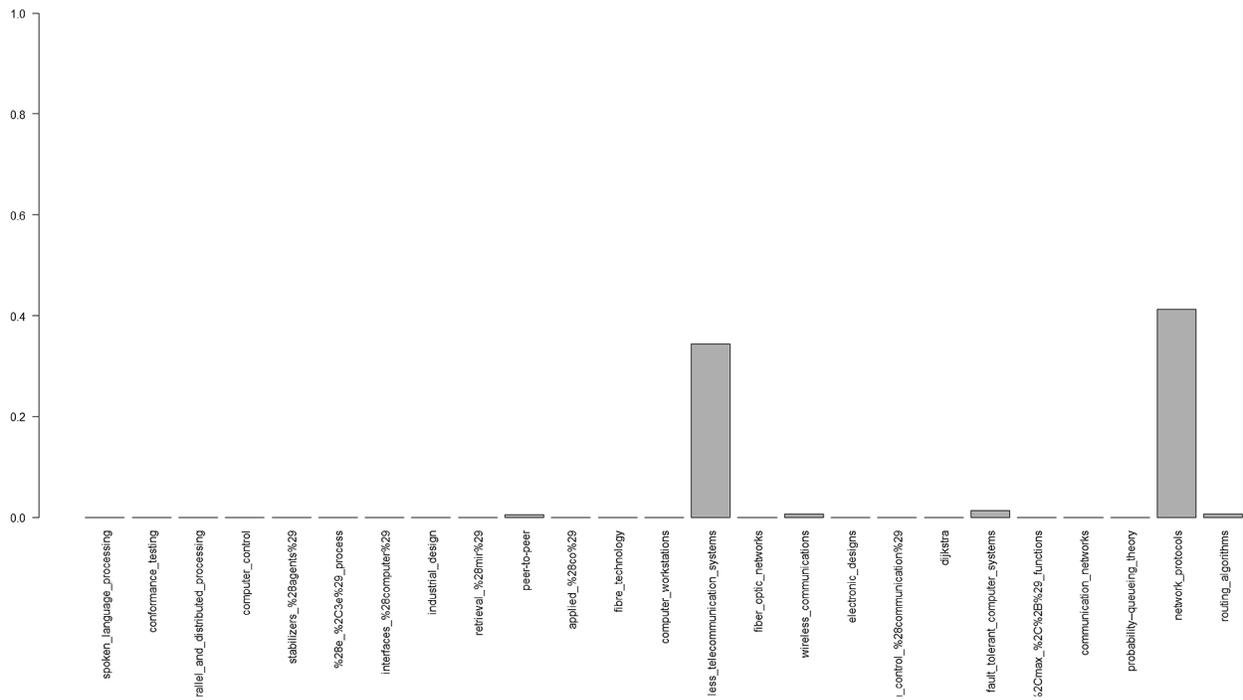
Figure 8: The significance of relationships between cloud computing and central subfields of the computer network concept
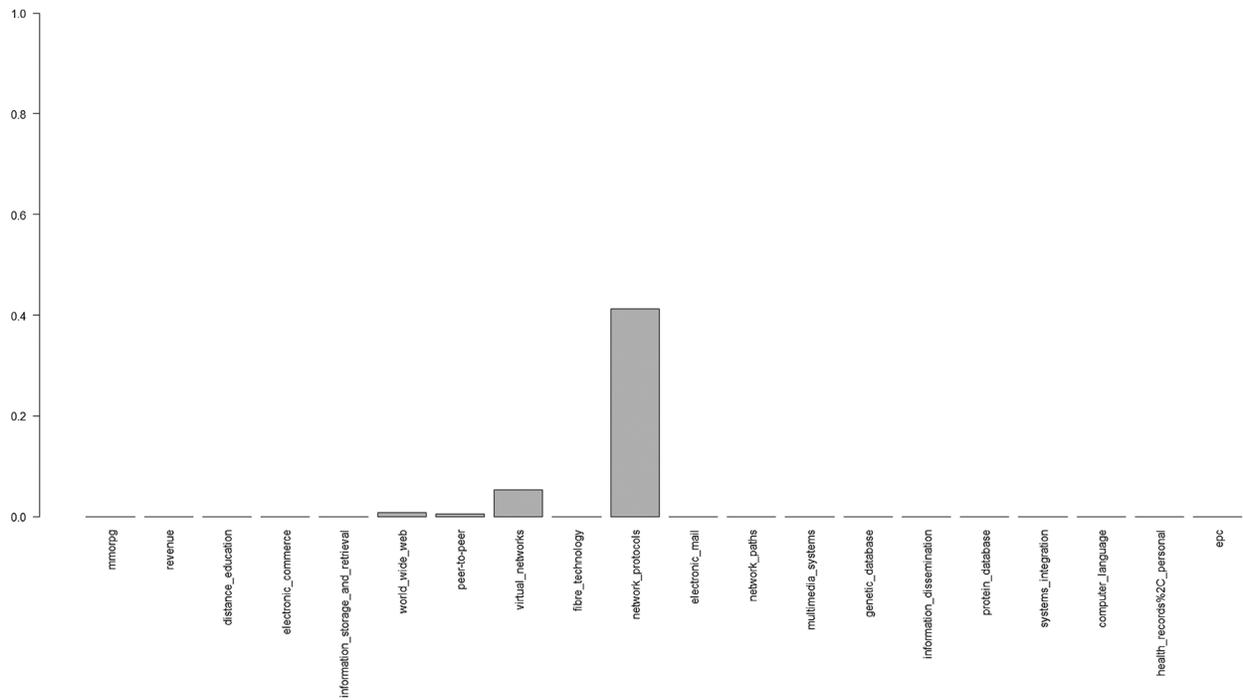
Figure 9: The significance of relationships between cloud computing and the principal subfields of the internet concept
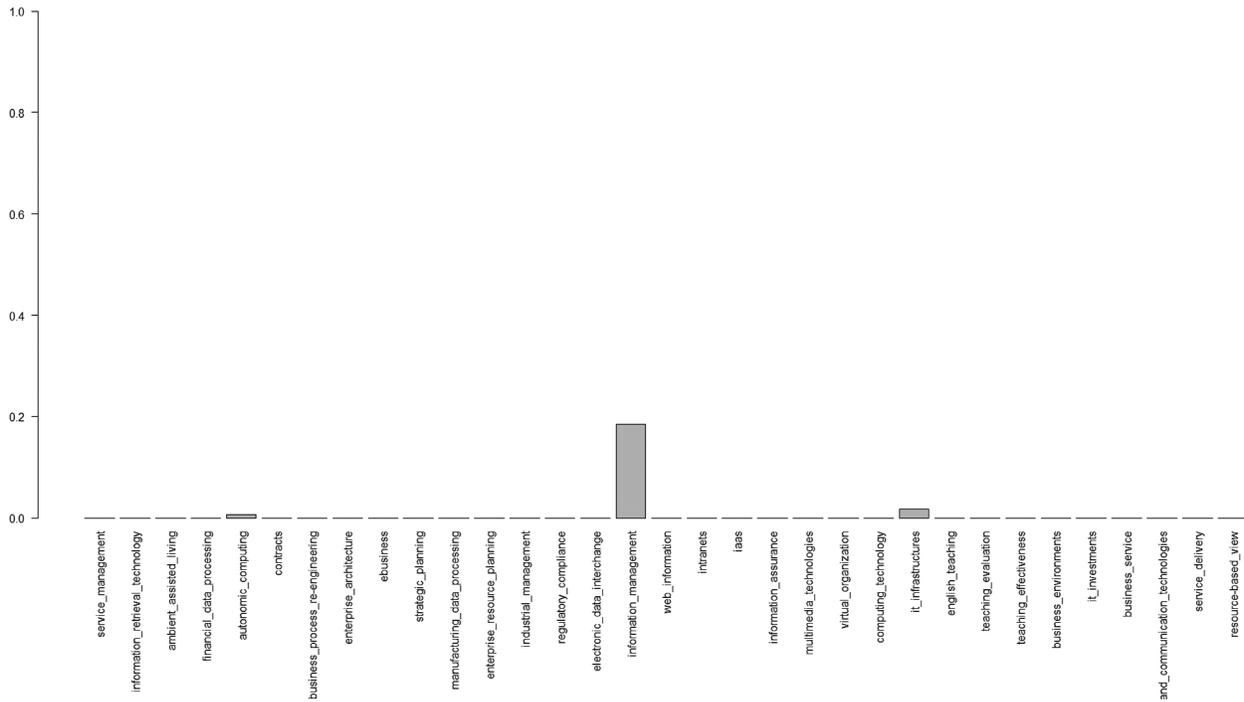
Figure 10: The significance of relationships between cloud computing and the foremost subfields of the information technology concept
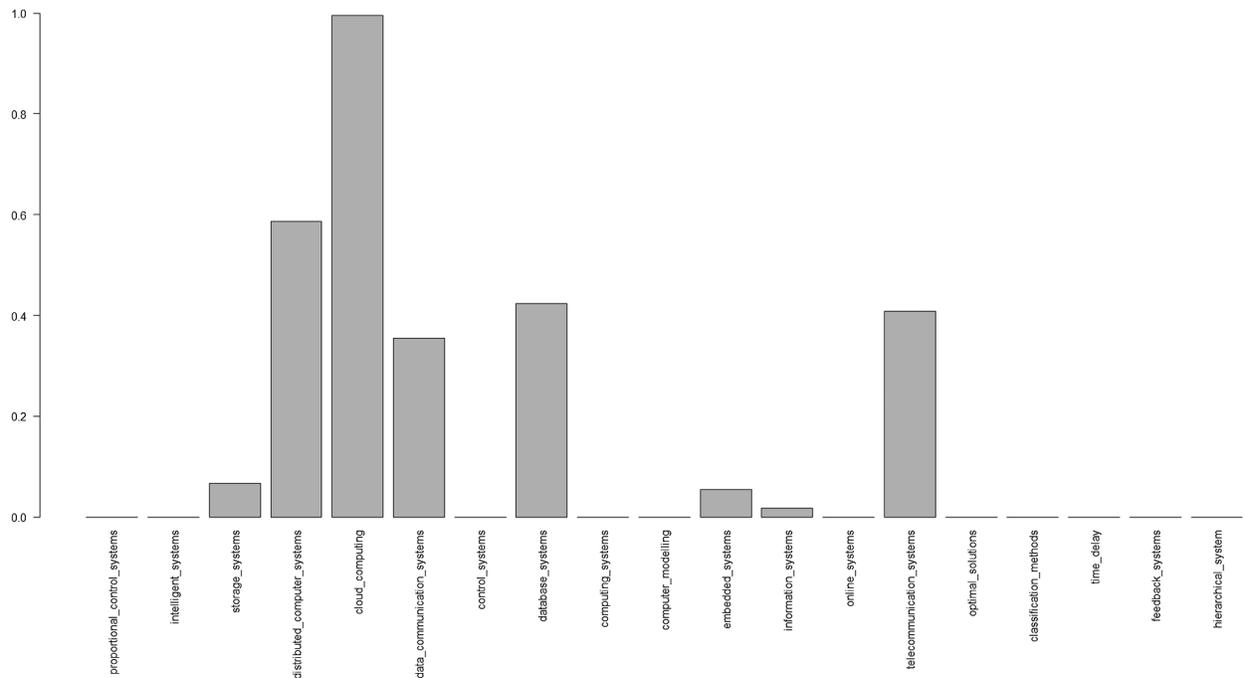
Figure 11: The Significance of Relationships between cloud computing and central subfields of the computer system concept

The last category of emphasized relationships focuses on Cloud Computing and the main areas under the umbrella of the Computer Security concept. We showed the proportion of papers with simultaneous references to Cloud Computing and Computer Security subfields in Fig. 12.
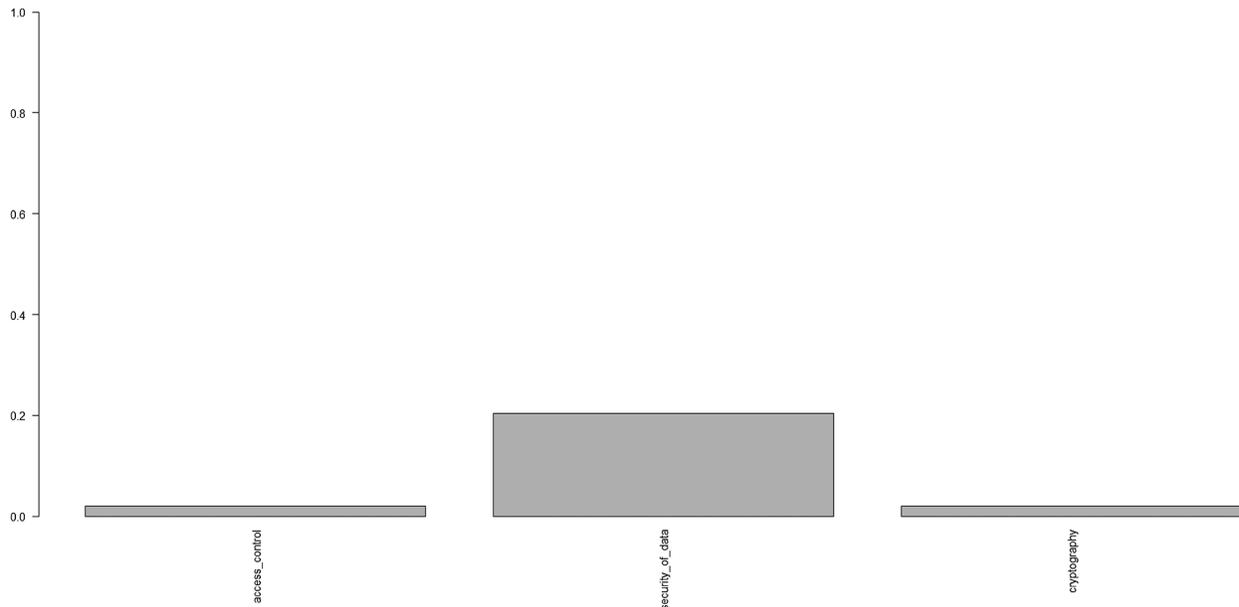
Figure 12: The significance of relationships between cloud computing and the principal subfields of the computer security concept

From all the five categories of suggestive relationships above (Figs. 8–12) we selected practically the most significant relationships, namely between Cloud Computing and: Telecommunication Systems and Network Protocols (under the umbrella of the Computer Network concept), Network Protocols, and Virtual Networks (for the more general notion of the Internet), Information Management and IT Infrastructures (for the one of Information Technology), Distributed Computer Systems, Data Communication Systems, Database Systems and Telecommunication Systems (for the Computer System concept) and finally Security of Data (under the umbrella of Computer Security).

4 Conclusions

In this paper, we studied the topic of Cloud Computing based on related ISI Web of Science data, mainly as abstracts of high-quality publications (SCIE and SSCI categories) for scientific papers published in the last 44 years.

We started with a custom data retrieval tool based on web scraping techniques. We relied on many other approaches: preparing data using custom filters, splits, joins, and text extraction functions, a score aggregation based on the Jaccard-coefficient, the analysis of the frequency of time-series of results using statistical tools and, a peculiar Computer Science Ontology model representation together with the construction of relationships between graph-based concepts, ontology-based annotations of abstracts, analysis of the importance of related topics, and also suggestive graphical representations.

In this way, we were able to identify robust relationships supported by high scores between Cloud Computing and two not necessarily exhaustive lists of primary and related concepts. The first includes Computer Networks, Internet, Information Technology, Computer Systems; and Computer Security, in

this particular order of importance. Regarding the afore-mentioned order of parent concepts, the second related list includes Telecommunication Systems, Network Protocols, Virtual Networks, Information Management, IT Infrastructures, Distributed Computer Systems, Data Communication Systems, Database Systems and Security of Data.

In addition to the discovery of patterns hidden behind a chosen topic of interest considered in this article, namely Cloud Computing and many others related with Computer Science, the importance of the study is mainly a methodological one that allows the objective identification of relationships and limits when dealing with concepts and related scientific domains, fields, and subfields.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding this study.

References

[1] M. Chiregi and N. Jafari Navimipour, "Cloud computing and trust evaluation: A systematic literature review of the state-of-the-art mechanisms," *Journal of Electrical Systems and Information Technology*, vol. 5, no. 3, pp. 608–622, 2018.

[2] F. Ibrahim and E. Hemayed, "Trusted cloud computing architectures for infrastructure as a service: Survey and systematic literature review," *Computers & Security*, vol. 82, pp. 196–226, 2019.

[3] J. Yu, X. Xiao and Y. Zhang, "From concept to implementation: The development of the emerging cloud computing industry in China," *Telecommunications Policy*, vol. 40, no. 2–3, pp. 130–146, 2016.

[4] N. Kshetri, "Institutional and economic factors affecting the development of the Chinese cloud computing industry and market," *Telecommunications Policy*, vol. 40, no. 2–3, pp. 116–129, 2016.

[5] K. Ali, S. Mazen and E. Hassanein, "A proposed hybrid model for adopting cloud computing in e-government," *Future Computing and Informatics Journal*, vol. 3, no. 2, pp. 286–295, 2018.

[6] M. N. Almunawar, "Benefits and issues of cloud computing for e-government," *Review of Public Administration and Management*, vol. 3, no. 1, pp. 1–2, 2015.

[7] S. Alshomrani and S. Qamar, "Cloud based e-government: Benefits and challenges," *International Journal of Multidisciplinary Sciences and Engineering*, vol. 4, no. 6, pp. 15–19, 2013.

[8] A. Tripathi and B. Parihar, "E-governance challenges and cloud benefits," in *Proc. of the IEEE Int. Conf. on Computer Science and Automation Engineering*, Shanghai, 2011.

[9] S. Song, "Competition law and interoperability in cloud computing," *Computer Law & Security Review*, vol. 33, no. 5, pp. 659–671, 2017.

[10] I. Walden and L. Luciano, "Ensuring competition in the clouds: The role of the competition law?," *ERA Forum*, vol. 12, no. 2, pp. 265, 2011.

[11] L. Novais, J. M. Maqueira and A. Ortiz-Bas, "A systematic literature review of cloud computing use in supply chain integration," *Computers & Industrial Engineering*, vol. 129, pp. 296–314, 2019.

[12] A. Brant and M. Sundaram, "A novel system for cloud-based micro additive manufacturing of metal structures," *Journal of Manufacturing Process*, vol. 20, no. 3, pp. 478–484, 2015.

[13] R. Oliveira, F. Noguez, C. Costa, J. Barbosa and M. Prado, "SWTRACK: An intelligent model for cargo tracking based on off-the-shelf mobile devices," *Expert Systems with Applications*, vol. 40, no. 6, pp. 2023–2031, 2013.

[14] G. Andreadis, G. Fourtounis and K. D. Bouzakis, "Collaborative design in the era of cloud computing," *Advances in Engineering Software*, vol. 81, pp. 66–72, 2015.

[15] J. H. Yang and P. Y. Lin, "A mobile payment mechanism with anonymity for cloud computing," *Journal of Systems and Software*, vol. 116, pp. 69–74, 2016.

[16] C. Christauskas and R. Miseviciene, "Cloud-computing based accounting for small to medium sized business," *Engineering Economics*, vol. 23, no. 1, pp. 14–21, 2012.

[17] P. D'Arcy, "CIO strategies for consumerization: The future of enterprise mobile computing," *Dell CIO Insight Series.* 2011.

[18] B. Anastasiei and N. Dospinescu, "Electronic word-of-mouth for online retailers: Predictors of volume and valence," *Sustainability*, vol. 11, no. 3, pp. 1–18, 2019.

[19] T. Noor, S. Zeadally, A. Alfazi and Q. Sheng, "Mobile cloud computing: Challenges and future research directions," *Journal of Network and Computer Applications*, vol. 115, pp. 70–85, 2018.

[20] P. Helo, M. Soursa, Y. Hao and P. Anussornnitisarn, "Toward a cloud-based manufacturing execution system for distributed manufacturing," *Computers in Industry*, vol. 65, no. 4, pp. 646–656, 2014.

[21] C. S. Chen, W. Y. Liang and H. Y. Hsu, "A cloud computing platform for ERP applications," *Applied Soft Computing*, vol. 27, pp. 127–136, 2015.

[22] J. Mo and W. Lorchirachoonkul, "Design of RFID cloud services in a low bandwidth network environment," *International Journal of Engineering Business Management*, vol. 3, no. 1, pp. 38–43, 2011.

[23] P. Golding and V. Tennant, "Performance review of RFID in the supply chain," in *Proc. of the 1st Int. Workshop on RFID Technology—Concepts, Applications, Challenges IWRT 2007; In Conjunction with ICEIS 2007*, 2007.

[24] F. Chen, R. Dou, M. Li and H. Wu, "A flexible QoS-aware web service composition method by multi-objective optimization in cloud manufacturing," *Computers & Industrial Engineering*, vol. 99, pp. 423–431, 2016.

[25] A. Singh, N. Mishra, S. I. Ali, N. Shukla and R. Shankar, "Cloud computing technology: Reducing carbon footprint in beef supply chain," *International Journal of Production Economics*, vol. 164, pp. 462–471, 2015.

[26] S. B. Camara, J. M. Fuentes and J. M. M. Marin, "Cloud computing, Web 2. 0, and operational performance: The mediating role of supply chain integration," *International Journal of Logistics Management*, vol. 26, no. 3, pp. 426–458, 2015.

[27] D. A. Battleson, B. C. West, J. Kim, B. Ramesh and P. Robinson, "Achieving dynamic capabilities with cloud computing: An emipirical investigation," *European Journal of Information Systems*, vol. 25, no. 3, pp. 209–230, 2016.

[28] S. Liu, F. Chan, J. Yang and B. Niu, "Understanding the effect of cloud computing on organizational agility: An empirical examination," *International Journal of Information Management*, vol. 43, pp. 98–111, 2018.

[29] T. Ravichandran and C. Lertwongsatien, "Effects of information systems resources and capabilities on firm performance: A resource-based perspective," *Journal of Management Information Systems*, vol. 21, no. 4, pp. 237–276, 2005.

[30] L. Fink and S. Neumann, "Exploring the perceived business value of the flexibility enabled by information technology infrastructure," *Information & Management*, vol. 46, no. 2, pp. 90–99, 2009.

[31] P. P. Tallon and A. Pinsonneault, "Competing perspectives on the link between strategic information technology alignment and organizational agility: Insights from a mediation model," *MIS Quartely*, vol. 35, no. 2, pp. 463–486, 2011.

[32] A. Jeyaraj, "Recent security challenges in cloud computing," *Computers and Electrical Engineering*, vol. 71, pp. 28–42, 2018.

[33] S. Laniepce, M. Lacoste, M. Kassi-Lahlou, F. Bignon, K. Lazri *et al.,* "Engineering intrusion prevention services for IaaS clouds: The way of the hypervisor, " in *IEEE Seventh Int. Sym. on Service-Oriented System Engineering*, San Francisco, USA, 2013.

[34] J. Vyas and M. Prashant, "Providing confidentiality and integrity on data stored in cloud storage by hash," *International Journal of Advance Research in Engineering, Science & Technology*, vol. 4, no. 5, pp. 38–50, 2017.

[35] C. Stergiou, K. Psannis, B. G. Kim and B. Gupta, "Secure integration of IoT and cloud computing," *Future Generation Computer Systems*, vol. 78, pp. 964–975, 2018.

[36] S. Sahmim and H. Gharsellaouib, "Privacy and security in internet-based computing: Cloud computing, internet of things, cloud of things: A review," *Procedia Computer Science*, vol. 112, pp. 1516–1522, 2017.

[37] H. Elazhary, "Internet of Things (IoT), mobile cloud, cloudlet, mobile IoT, IoT cloud, fog, mobile edge, and edge emerging computing paradigms: Disambiguation and research directions," *Journal of Network and Computer Applications*, vol. 128, pp. 105–140, 2019.

[38] D. Androcec, N. Vrcek and J. Seva, "Cloud computing ontologies: A systematic review," in *3rd Int. Conf. on Models and Ontology-based Design of Protocols, Architectures and Services*. Chamonix, France, 2012.

[39] M. Al-Sayed, H. Hassan and F. Omara, "Towards evaluation of cloud ontologies," *Journal of Parallel and Distributed Computing*, vol. 126, pp. 82–106, 2019.

[40] A. Boukerche and R. De Grande, "Vehicular cloud computing: Architectures, applications, and mobility," *Computer Networks*, vol. 135, pp. 171–189, 2018.

[41] D. Homocianu and M. Homocianu, "GiPlot: An interactive cloud-based tool for visualizing and interpreting large spectral data sets," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 209, pp. 234–240, 2019.

[42] Z. Du, L. He, Y. Chen, Y. Xiao, P. Gao *et al.,* "Robot cloud: Bridging the power of robotics and cloud computing," *Future Generation Computer Systems*, vol. 74, pp. 337–348, 2017.

[43] Y. Chen, "Service-oriented computing in recomposable embedded systems," in *Joint IARP/IEEERAS/EURON/IFIP 10.4 Workshop on Dependability in Robotics*, Tucson, Arizona, 2006.

[44] Y. Chen, Z. Du and M. Garcia-Acosta, "Robot as a service in cloud computing," in *2010 Fifth IEEE Int. Sym. on Service Oriented System Engineering*, Nanjing, China, 2010.

[45] Y. Bounagui, A. Mezrioui and H. Hafiddi, "Toward a unified framework for cloud computing governance: An approach for evaluating and integrating IT management and governance models," *Computer Standards & Interfaces*, vol. 62, pp. 98–118, 2019.

[46] Y. He, "The lifecycle process model for cloud governance," 2020. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.458.1988&rep=rep1&type=pdf.

[47] B. Varghese and R. Buyya, "Next generation cloud computing: New trends and research directions," *Future Generation Computer Systems*, vol. 79, pp. 849–861, 2018.

[48] S. Hakak, N. F. M. Noor, M. N. Ayub, H. Affal, N. Hussin *et al.,* "Cloud-assisted gamification for education and learning—Recent advances and challenges," *Computers and Electrical Engineering*, vol. 74, pp. 22–34, 2019.

[49] M. Alles, "Examining the role of the AIS research literature using the natural experiment of the 2018 JIS conference on cloud computing," *International Journal of Accounting Information Systems*, vol. 31, pp. 58–74, 2018.

[50] J. C. Kim and K. Chung, "Associative feature information extraction using text mining from health big data," *Wireless Personal Communications*, vol. 105, no. 2, pp. 691–707, 2018.

[51] A. Salatino, T. Thanapalasingam, A. Mannocci, F. Osborne and E. Motta, "The computer science ontology: A large-scale taxonomy of research areas," in *Int. Semantic Web Conf. 2018: The Semantic Web*, Monterey, USA, 2018.

[52] G. Csardi and T. Nepusz, "The igraph software package for complex network research," *InterJournal Complex Systems*, pp. 1695, 2006.

[53] P. Lula, S. Wiśniewska and K. Wójcik, "Ontology-based system for automatic analysis of job offers," in *Proc. of the 21st Int. Conf. on Information Technology for Practice*, Ostrava, Czech Republic, 2018.

[54] J. Ooms, "Hunspell: High-performance stemmer, tokenizer, and spell checker," 2018. [Online]. Available: https://cran.r-project.org/web/packages/hunspell/index.html [Accessed 25 May 2020].