AI

MDPI

Article

Latent Dirichlet Allocation and t-Distributed Stochastic Neighbor Embedding Enhance Scientific Reading Comprehension of Articles Related to Enterprise Architecture

Nils Horn [ID], Fabian Gampfer [ID] and Rüdiger Buchkremer *[ID]

Institute of IT Management and Digitization Research (IFID), FOM University of Applied Sciences,
40476 Düsseldorf, Germany; nils.horn@fom-net.de (N.H.); fabian.gampfer@fom-net.de (F.G.)
* Correspondence: ruediger.buchkremer@fom.de

Abstract: As the amount of scientific information increases steadily, it is crucial to improve fast-reading comprehension. To grasp many scientific articles in a short period, artificial intelligence becomes essential. This paper aims to apply artificial intelligence methodologies to examine broad topics such as enterprise architecture in scientific articles. Analyzing abstracts with latent dirichlet allocation or inverse document frequency appears to be more beneficial than exploring full texts. Furthermore, we demonstrate that t-distributed stochastic neighbor embedding is well suited to explore the degree of connectivity to neighboring topics, such as complexity theory. Artificial intelligence produces results that are similar to those obtained by manual reading. Our full-text study confirms enterprise architecture trends such as sustainability and modeling languages.

Keywords: enterprise architecture; natural language processing; latent dirichlet allocation; text mining; t-distributed stochastic neighbor embedding; reading comprehension

1. Introduction

Comprehending a scientific article (reading comprehension) is a sophisticated cognitive process that depends on numerous extrinsic and intrinsic factors [1]. First of all, precise keywords have to be grasped. The correlations in which these keywords occur are crucial. Additional proximal terms increase the dimensionality of the data acquisition and complicate the analysis with artificial intelligence (AI). However, this may be reduced by lowering the dimensionality. One method to cope with this process is "neighbor embedding". It allows topics that are not directly related to the core area to be captured. This gives insight beyond the core area. It can also predict whether a trend is connected to an external field. Therefore, the question arises as to whether it makes more sense to capture the whole text or perhaps only the summaries (abstracts) to explore this interconnectedness arises. In this paper, we want to demonstrate that for this purpose, it makes more sense to focus on the abstracts by examining both.

The number of publications in the enterprise architecture (EA) research area is continuously growing [2]. In his much-quoted article, Zachman, who is regarded as one of the most significant researchers in modern EA, describes the need for a logical architecture to structure companies' systems as early as 1987 [3]. While early approaches were heavily information-technology focused, the focus has shifted towards business-related contexts such as business processes and organizational goals, as described by Winter and Fischer in 2006 [4].

Today, various methods and approaches support existing architectures' management and for further development and transformation into future state-of-the-art EA [5].

After three decades of scientific research, a broad range of topics has emerged within the EA research area, growing every year [6]. A holistic analysis of current trends and topics

within EA is therefore difficult to perform by conventional means. However, a solution-oriented approach can be found in the field of artificial intelligence. Modern methods, such as topic modeling, make it possible to carry out full-text analyses systematically [7]. Thus, a large number of publications in the field of EA can be investigated simultaneously.

This study aims to provide an overarching view of current developments in the research area EA and to complement existing research carried out in the past. The identified topics and trends are discussed and examined for practical relevance.

To conduct the trend analysis based on artificial intelligence practices, design science research is used, defined in more detail by Hevner et al. [8] and Winter [9]. Using topic modeling as an algorithm for the automatic evaluation of scientific publications is not new. Buchkremer et al. [7] also used the methodology in their studies.

This article is structured as follows: In the next section, theoretical basics are explained to create a deeper understanding of the technology in use and serve as a theoretical introduction to the overall topic. In Section 3, the data preparation for the topic modeling analysis is described in detail. The data preparation and the number of topics to be examined will be determined and discussed. Furthermore, the selection of the training algorithm is explained. Next, the application is described with particular attention to the parameterization of the analysis procedure. Subsequently, in Section 4, the determined topics are presented, described individually, and discussed. Finally, in Section 5, the procedure is critically reflected, and limitations of the analysis are pointed out. The elaboration comes to an end with a conclusion, including an outlook on possible further research approaches.

2. Background and Related Works

2.1. Enterprise Architecture

As Saint-Louis et al. [10] confirmed in a recent study, heterogeneous definitions and descriptions of the term 'Enterprise Architecture' can be found in the literature. In terms of this study, we rely on the definition pronounced by the international standard ISO/IEC/IEEE 42010:2011 [11], which describes EA as a methodology for managing and developing enterprise architecture and defines the term as follows: 'Enterprise Architecture is a discipline that manages the fundamental organization of an enterprise, which is embodied in its components, their relationships to one another and the environment, and the principles that govern its design and evolution.'

2.2. State-of-the-Art Reviews on Enterprise Architecture

Since our previous study, conducted in 2016, roughly four years have passed, and the body of knowledge of EA has continued to grow. More than 300 peer-reviewed scientific contributions concerning EA have been published each year [6]. Among these contributions, there are also several recent state-of-the-art reviews: Kitsios and Kamariotou [12] examine how existing EA modeling frameworks cover business strategy optimization and provide insight into a special subarea of EA. Zhang et al. [13] analyze the subject EA with its links to the closely related subject of business-it alignment. Ansyori et al. [14] consider precisely the subject of critical success factors to implement EA. Dumitriu and Popescu [15] cover the design of EA frameworks in their review.

However, none of the reviews conducted since 2016 come from an overarching point of view that considers the subject of EA as a whole. We want to address this gap with our work. As shown previously, natural language processing (NLP) is well suited to achieve this objective, so we have chosen it again for our research. Moreover, this work presents an opportunity to validate the predictions made in 2016 and indicate whether an NLP-supported systematic review can provide an accurate prognosis on developing a scientific research field.

2.3. Topic Modeling as a Part of Natural Language Processing

NLP is empirically referred to as a collective term for the machine processing of natural language. It combines computational linguistics, information technology, artificial intelligence, and cognitive sciences [16]. The combination of computer science and linguistics can create possibilities to process natural language by machine, for example, through stochastic algorithms and program logic [17]. In doing so, mainly recognized language areas, syntactic features, and semantics from linguistics are used [18]. By using intelligent and self-learning algorithms of artificial intelligence, machines can cognitively interpret and process natural language [19].

An essential part of artificial intelligence research in machine learning (ML) was defined by Samuel [20] in Empirical Research. Employing ML, computers can learn and continuously improve independently and without explicit programming [20]. In general, ML can be defined by different algorithms. In essence, however, a distinction is made between supervised and unsupervised learning [20].

A combination of both is used with topic modeling, which identifies various topics for given documents by unsupervised data clustering [21]. As Sun et al. [22] explain, topic modeling consists of finding topics *T* that best describe the text's content. It is assumed that each document is a mixed model of topics, where *T* is given, and a multinomial distribution of words described for each topic. Hong and Davison [23] describe topic modeling as an automatism for extracting dominant topics from a text corpus. Often, the Latent Dirichlet Allocation (LDA) algorithm presented by Blei et al. [24] is considered the basis of topic modeling. Anwar et al. [25] describe the algorithm as a flexible, generative, probabilistic topic model for collecting discrete data, which considers documents to collect a topic selection. Each topic is represented as a list of words with probabilities of belonging. Haidar and Kurimo [26] also understand LDA as a generative, probabilistic topic model and define it as a three-level Bayesian model. In a recent paper, Hussein et al. explain that in addition to classical deep learning methods, transformer technologies such as BERT (Bidirectional Encoder-Representations from Transformers) can be applied to explore trends in texts. Similar to BERT, LDA and IDF also target the frequency of words in a text. t-SNE, similar to transformers, examines the proximity of words to each other [27]. Hao et al. tackle cross-domain sentiment alignment by applying stochastic word embedding [28].

3. Applying Topic Modeling to Enterprise Architecture Research

3.1. Topic Modeling Methodology for Literature Reviews

To carry out the analysis of the given publications without disturbances, text preparation steps must be carried out in advance [7]. Since the research contributions are initially available as a PDF and are therefore encrypted, they must be converted into a processable text format. As Welbers et al. [29] show, the R-framework 'pdftools', developed by Ooms [30], offers simple possibilities for converting the text. The texts are vectorized for more efficient processing and then cleaned up using various text mining methods [31,32]. As the original wording of the publications is to be used for evaluation, punctuations and stop words such as 'and' or 'the' are removed from the texts, as these do not offer any significant added value for the analysis. Yaram [33] uses the R-Framework 'tm' for text cleansing, which was introduced by Feinerer [34] and will also be used in this work. To perform a topic modeling using LDA afterward, it is necessary to transform the already vectorized texts into a document term matrix (DTM) [35]. The DTM includes the number of respective words per document and represents them numerically. In the following figure (see Figure 1), *d* denotes the documents and *w* the weighting or occurrence of the words *t* per document [36]. According to this, for example, W_{11} represents the weighting or occurrence of word 1 to document 1. *m* and *n* are used as counters for words and documents considered.

$$A_{mxn} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ w_{m1} & w_{m2} & \cdots & w_{mn} \end{bmatrix} \begin{matrix} \leftarrow t_1 \\ \leftarrow t_2 \\ \vdots \\ \leftarrow t_m \end{matrix}$$

$$\begin{matrix} d_1 & d_2 & \cdots & d_n \\ \downarrow & \downarrow & & \downarrow \end{matrix}$$

Figure 1. Structure of a document term matrix.

After the transformation into a DTM, a list of words with the most robust frequency is generated and analyzed. This list shows, for example, that words like 'IEEE,' 'vol' or 'city' are located in the corpus at high frequency, which can lead to a distortion of the results of the topic modeling [37,38]. These very words can be classified as stop words and should be cleaned up. For exclusion, a separate function is implemented in R to exclude a range of freely selectable words from the text.

Modeling is mainly influenced by the number of topics to be identified [39]. To determine the ideal number of topics, empirical studies often rely on log-likelihood measurement [40,41].

To validate the results determined by the LDA algorithm, cross-validation is applied [42]. As part of this procedure, the training text corpus is divided into two parts. The first and usually more extensive body of the text is used for training and the second for testing. The test corpus w consists of documents that have not been considered. The training model is described as a topic matrix with Φ as topic-word distribution. The parameter Θ is not considered because it represents the training set's document-topic distribution and is therefore unsuitable for evaluation.

$$L(w) = \log p(w\,|\Theta, \Phi) = \sum\nolimits_d \log p(w_i\,|\Theta, \Phi). \tag{1}$$

Thus, considering the logarithmic probability of a set of unseen documents, w_i considers the topic-word distribution Φ for the document-topic distribution θ [43]. In particular, the perplexity of the training data on the test corpus is used as a measure to determine the transferability of the model [44]. The perplexity is defined as:

$$perplexity(w) = \exp\left\{ -\frac{L(w)}{N} \right\}. \tag{2}$$

where N is the total number of words in the test corpus [43].

A decreasing function of the logarithmic probability *L(w)* can be seen. Therefore, it is generally true that the lower the calculated perplexity, the better the performance of the trained topic modeling model [45,46].

For the analyses of this paper, we used the full text of 231 scientific publications in the subject area of EA from various scientific libraries. The process of data retrieval is explained in more detail in Section 3.4.

In this paper, the evaluation of the optimal number of topics is implemented in the programming language R. A separate function that performs multiple k-fold cross-validations is developed. For this purpose, a model is trained in each case, then transferred to a test corpus to measure the perplexity. A value range between 2 and 300 subjects is analyzed to determine the optimal number of subjects. Since the evaluation is computationally intensive, several computing clusters have to be formed to parallelize the calculations. The following graphic in Figure 2 visualizes the results of the analysis.
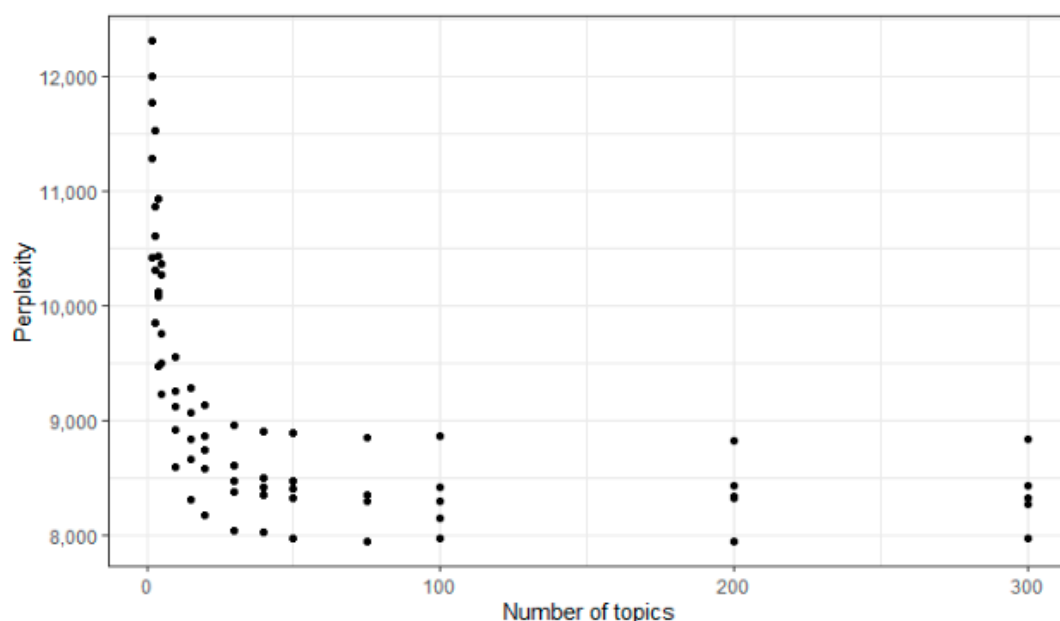
Figure 2. Cross-validation of perplexity versus the number of topics.

The optimum number of topics for the application is between 40 and 100 iterations (see Figure 2). Another possibility for evaluating the best possible number of topics is the R-framework 'ldatuning', developed by Murzintcev [47]. The framework uses several metrics for measurement. We use three of them for our study. Griffiths 2004 [48] represents an approach where the number of topics is optimal when the log-likelihood for the data becomes maximum. CaoJuan 2009 [49] metric measures the stability of topic structure using an average cosine distance between every pair of topics. Arun 2010 [50] finds the optimal number of topics by applying a symmetric Kulback-Leibler divergence on the distributions generated from topic-word and document-topic matrices, as they viewed a topic model as matrix factorization [51].

All three metrics (see Figure 3) suggest that the optimal number of topics is between 40 and 60. A low value is preferred in the upper graph, while in the lower graph, a high value is favored. This confirms the perplexity analysis carried out previously. Thus, for this study, 50 topics are to be determined.
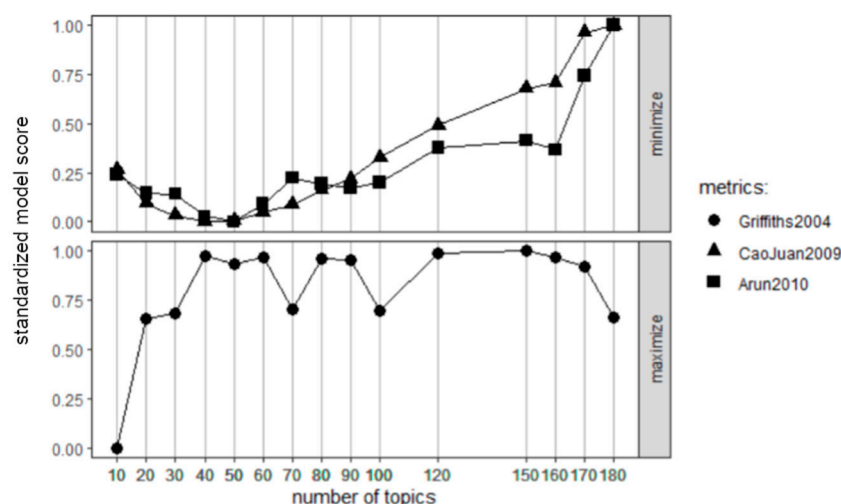

Figure 3. Determining the optimal number of topics according to Gibbs.

The R-framework 'topicmodels' presented by Grün and Hornik [52] will be used for the investigation intended in this paper. This framework includes implementing the topic modeling algorithm LDA and other basic methods required for the analysis, for example, perplexity. The framework is based on the R-Framework 'tm', which is also used in this study [53]. In general, a distinction can be made between the two approaches VEM and Gibbs sampling, evaluate the topics found [52]. In this study, the Gibbs sampling algorithm implemented by Griffiths and Steyvers [48] will be used, as it has proven itself in various studies [37,40,54]. The algorithm performs topic modeling as follows: Given are the vectors of all words →w and →z and their topic assignment by data collection W. The topic assignment depends on the assignment of all other word positions, namely, the topic assignment of a word t is taken from the following multinomial distribution [55].

$$p\left(z_i = k \middle| \vec{z} \neg i, \vec{w}\right) = \frac{n_k^{(t)}, \neg i + \beta_t}{\left[\sum_{v=1}^{V} n_k^{(V)} + \beta_v\right] - 1} \frac{n_m^{(k)}, \neg i + \alpha_k}{\left[\sum_{i=1}^{K} n_m^{(j)} + \alpha_j\right] - 1}. \tag{3}$$

In the formula, $n_k^{(t)}, \neg i$ denotes the number of assignments of the word t to the topic k. $\sum_{v=1}^{V} n_k^{(V)}$ represents the total number of words t assigned to the topic k. Furthermore, $n_m^{(k)}, \neg i$ deals with the number of words in the document m assigned to topic k. All the terms described above exclude the current assignment. The formula element $\sum_{i=1}^{K} n_m^{(j)}$ stands for the total number of words in the document m except for the word t that is currently being dealt with. α and β denote the Dirichlet parameters, which are symmetric [55].

In addition to selecting the training algorithm, an optimal parameterization of the methodology according to the requirements is necessary.

3.2. t-Distributed Stochastic Neighbor Embedding for Topic Model Visualization

The tSNE algorithm, introduced by van der Maaten and Hinton [56], is used to overcome the challenge of representing high-dimensional data, and it has a wide range of applications, e.g., life sciences, analysis of deep learning networks [57–59].

In general, tSNE reduces data's dimensionality and produces 2D or 3D embeddings, preserving local structures in the high-dimensional data. Typical tasks performed by users of tSNE are based on identifying relationships between data points and their origin. The tasks often include identifying visual clusters and their verification, e.g., using parallel coordinate plots [60].

The algorithm tSNE calculates two joint probability distributions P to represent the total distance between data points in the high-dimensional space and Q describing the similarity in the low-dimensional space. The goal of this logic is to achieve a faithful representation of P in low-dimensional space by Q. This can be achieved by minimizing the cost function C given by the Kullback-Leibler divergence between the joint probability distributions P and Q to optimize the positions in the low-dimensional space [60]. The minimization of the Kullback-Leibler divergence and the change in position of the low-dimensional points for each step of the gradient descent is defined as [61]:

$$C(\varepsilon) = KL(P\|Q) = \sum_{i \neq j} p_{ij} \log \frac{p_{ij}}{q_{ij}}. \tag{4}$$

where ε denotes an s-dimensional embedding, p_{ij} defines the joint probabilities that measure the pairwise similarity between two high-dimensional input data points and q_{ij} the embedding similarity between two low-dimensional points (low-dimensional version of data points of p_{ij}) [61]. Maaten describes that the objective function focuses on modeling high values of p_{ij} (similar objects) by high values of q_{ij} (nearby points in the embedding space) [61]. This is because Kullback-Leibler divergence is asymmetric. In case of the em-

bedding ε, the objective function is non-convex and is therefore typically minimized by using the Gradient descent function:

$$\frac{\delta C}{\delta y_i} = 4 \sum_{j \neq i} (p_{ij} - q_{ij}) q_{ij} Z(y_i - y_j). \tag{5}$$

where y_i and y_j denote the low-dimensional input data points, and Z is defined as normalization term [61]:

$$Z = \sum_{k \neq l} \left(1 + \|y_k - y_l\|^2 \right)^{-1}. \tag{6}$$

We refer to van der Maaten and Hinton [56] and Maaten [61] for more details on the algorithm.

In this paper, we use the R-framework 'Rtsne' introduced by Krijthe and van der Maaten [62], which implements the Barnes-Hut tSNE algorithm to reduce the computational complexity [61].

While the original tSNE uses a brute-force approach that increases computational complexity and memory complexity (which is $O(n^2)$), the Barnes-Hut variant uses a quadtree, resulting in a reduction in computational complexity of $O(nlog(n))$ and the memory complexity to $O(N)$ per iteration [60].

For more details, we refer to van der Maaten [61].

3.3. Comparison to the Methodology of Previous Studies

Gampfer et al. [6] ran an EA trend analysis supported by NLP in 2016, published in 2018. Table 1 shows a comparison of the methods used in both papers. While the overall approach and the subject are similar, there are also crucial differences that will be addressed subsequently.

Table 1. Comparison of the methods used.

Characteristic	Methodology of Gampfer et al. [5]	Methodology in This Work
Input	3799 documents	231 documents
Input Type	Title and abstract of papers	Full text or papers
Method	K-Means Clustering with Davis-Bouldin	Latent Dirichlet Allocation (LDA)
Tools	RapidMiner & SAS	R
Results	8 clusters/trends	12 clusters/trends
Result Type	n-to-n relationship between input documents and trends	n-to-1 relationship between input documents and trends

To better judge the comparability of the methods used, we applied the methodology presented in this study to Buchkremer and coworkers [6].

The application shows that the topics cloud, agile/adapt, smart, big data, sustainable, entrepreneurial, complexity theory, and IoT could also be determined by the new methodology. Table 2 shows the characteristic terms of the EA publications and the trend and topics determined based on these terms:

Table 2. Characteristic terms of EA publications and identified trends of the previous study.

Identified Terms	Related Trend/Topic
saas, cloud, computing	cloud
agile, methodology, adapt	agile/adapt
smart, machines	smart
framework, big, data, analysis	big data
green, bio, sustainable	sustainable
entrepreneurial, enterprise, enterpriselevel	entrepreneurial
complexity, theory	complexity theory
iot, things	internet of things

The direct comparison shows that the same trends and topics could be determined with both algorithms. Hence, from a high-level point of view, both methods produce similar results. However, a significant difference becomes evident when looking at how individual documents can be mapped to trends. Gampfer et al. [6] use an n-to-n mapping between documents and trends—meaning a document can belong to multiple trends. This work uses n-to-1 mapping—meaning a document can belong to one trend only.

3.4. Information Retrieval: Publication Search and Selection Process

For a careful consideration of current research topics within EA, publications from the years 2019 and 2020 were extracted from multiple relevant databases. In the selection of scientific publications, the focus is on peer-reviewed journals and conference proceedings. Furthermore, potential duplicates were cleaned up based on a manual comparison of the publication titles. The following combined search string is used to identify as many publications with topic-specific focus as possible: '"Enterprise Architecture" OR "Enterprise Architecture Management."' The individual search result was limited to the publication period between 2019 and 2020 to ensure that the publications are up to date. The search results of the following databases are looked at:

- IEEE Xplore® Digital Library (https://ieeexplore.ieee.org/Xplore/home.jsp)
- ACM Digital Library (https://dl.acm.org/)
- Science Direct–Elsevier (https://www.sciencedirect.com/)
- Springer Link (https://link-springer.com/)

Within the scope of data acquisition, 271 documents on the subject area EA were attained from 30 December 2019 to 10 January 2020. After the manual review and the cleaning of potential duplicates, 231 documents were identified, examined within the study.

3.5. Application of Algorithm and Parametrization

The parameter seed is often used in the programming language to reproduce the results [63,64]. The hyperparameter marks a starting point for the generation of a sequence of random numbers. If the random number generator is identical, the results' reproducibility can be achieved with the same configuration. For this study, the hyperparameter should be set to *seed* = 2020. This selection was made arbitrarily. In their studies, Green and Hornik specify a repetition rate between 1000 and 2000 [52]. Séaghdha uses a repetition rate of 1000 in his studies [65]. Our tests have shown that the best results can be achieved with 2000 repetitions. Accordingly, the algorithm is parameterized with *iteration* = 2000. In principle, the hyperparameter *best* can be used to decide whether all training runs should report a result—for this purpose, the parameter *best = FALSE* is set—or only those which, in retrospect, show the best logarithmic probability [52]. In this study, only results with the best possible probabilities should be used. Thus, the parameter *best = TRUE* is defined. The control value k indicates the number of topics to be determined and is set to $k = 50$ for this investigation, as described in the previous chapter. The estimation parameter α is set to $50/k$, based on the recommendation of Griffiths and Steyvers [48]. Furthermore, the DTM, as the basis of the analysis, is transferred to the function.

4. Current Enterprise Architecture Research Trends

To identify the current EA trend topics, we analyzed full-text publications and applied an approach that combines unsupervised and supervised techniques. First, we ran an utterly unsupervised algorithm to obtain clusters of terms that occur together in the documents. We identified current trends based on the clusters and validated them by checking assigned documents in a second step. In the next step, we used both steps to analyze the topics in more detail. In the analysis, we compared the trend we had identified with other studies and the Gartner Hype Cycle for Enterprise Architecture [66] to examine the relevance of the practice topics.

4.1. Identifying and Measuring Current EA Trends

For the topic identification, we used the algorithm LDA as an unsupervised method for clustering terms. The algorithm defines the clusters based on the occurrence of terms that belong to a topic or the probability of words belonging to a topic; for details on the method, see Section 3. The method's application resulted in a list of manually reviewed terms to identify relevant subjects in terms of content. The terms that cannot be mapped to a topic were excluded from the analysis.

Table 3 shows the result of the review and the topic assignment.

Table 3. Characteristic terms of recent EA publications and identified trends of this work.

Identified Terms	Related Trend/Topic
pattern, fuzzy	pattern recognition
security, attack, protection	security
saas, cloud, computing	cloud computing
sustainable, ecosystem	sustainability
complexity, theory	complexity theory
archimate, bpmn, modeling	modeling languages
digital, transformation, innovation	digital transformation
internet, things, sensor	internet of things
data, big, veracity	big data
release, cycle, development	continuous planning

We tagged the topic with the highest fit as per probability calculated by the algorithm for each document. According to the results shown in Figure 4, the most significant relevance is Sustainability, while the Internet of Things showed the smallest number of assigned documents.
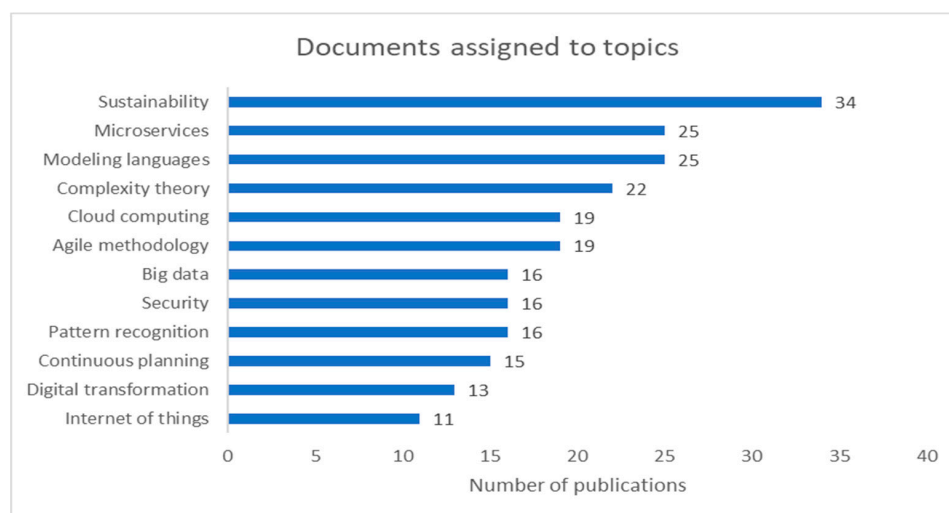

Figure 4. Distribution and assignment of the examined documents to the identified topics.

The distribution of EA trends identified by this work mostly confirms the predictions of Buchkremer and coworkers [6] made in 2016. Sustainability is clearly on the rise, while Agile methodology has lost researchers' attention in the field. One striking deviation of the prediction concerns the field of complexity theory. While the forecast of 2016 indicated that this topic would play a niche role, current results show an increasing interest in the subject. In Section 4 of this paper, we have a more detailed look at the individual trends.

4.2. Significance of Full-Text Mapping and the Deployment of t-SNE in Analyzing EA Trends

Compared to the 2018 study by Buchkremer and coworkers, it can be seen that complete text analysis yields similar results to that of abstract analysis. To identify trends for a topic with many publications, we recommend examining abstracts instead of full texts. The results are comparable, and it is more cost-effective and less computationally intensive. t-SNE is helpful in identifying the degree of interconnectedness. Thus, trends are indexed that the average expert might not have determined to be directly related to the topic area, such as identifying complexity theory. t-SNE also shows that the discipline of EA is becoming more interconnected and multidisciplinary overall, and an increasing number of trends can be identified that are not part of the core EA topic (see Figure 5).

The following figure shows that many documents have more than one topic assigned. This is an indicator that topics are related to each other.
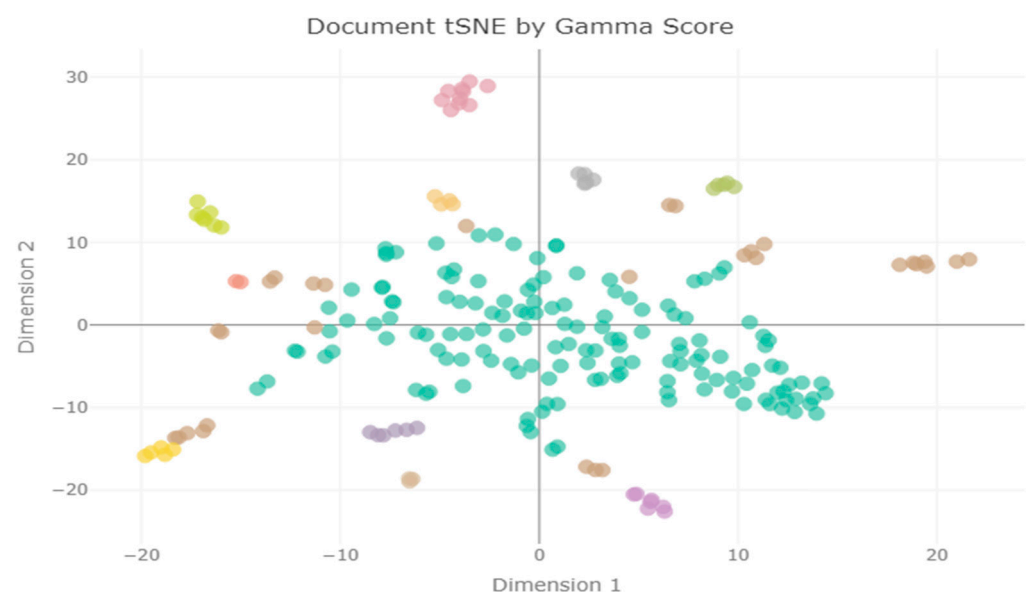

Figure 5. Document grouping based on their similarity using tSNE.

4.2.1. Cloud Computing and EA

The investigation shows that cloud computing [67,68] as an overarching theme currently forms a trend within EA. Gampfer et al. [6] forecast a continuing trend of the topic, which can also be seen in this analysis, as the focus of modern IT architectures, in particular, has moved away from stationary system landscapes to cloud deployment in recent years [69]. In practice, the relevance of cloud computing seems to be declining. While the topic has already been listed as 'sliding into the trough' in Gartner's 2017 'Hype Cycle' [70], it is no longer identified as a trend in the 2019' Hype Cycle' [66].

4.2.2. Sustainability and EA

As the study by Gampfer et al. [6] shows, there is a growing interest in the topic of sustainability in the EA sector. In general, sustainability is considered one of the significant challenges of contemporary society [71]. Likewise, more and more companies adopt a sustainability strategy, which underlines the relevance of the economy's topic [72].

Thus, sustainability is also in demand in the EA context, especially regarding how developments can be made long-term and sustainable [73]. In the 'Hype Cycle' by Gartner for EA, this topic does not appear [66].

4.2.3. Digital Transformation and EA

Digital transformation is the umbrella term for the digitization of today's society [74]. According to Zimmermann et al. [75], the term covers technological megatrends such as big data, artificial intelligence, and cloud computing. From an economic perspective, digital transformation enables new technologies to achieve competitive advantages [76]. EA's task is to support the digital transformation process by continuously evaluating and reconfiguring a company's value creation mechanisms. The resulting change interacts with all information systems and affects the existing system architectures [77]. The other trends identified can be assigned to this topic. Gartner's current 'Hype Cycle' covers a sub-discipline of digital transformation and focuses on the change within companies through the hype 'Digital Business Transformation' [66].

4.2.4. Pattern Recognition and EA

A current trend in pattern recognition or machine learning in EA can be identified based on the topic analysis. The 'Hype Cycle' for Gartner's EA area also shows a current trend in this topic area [66]. Machine Learning is currently being taken up in practice and research and influences the digital transformation [78]. In particular, the opportunity to automate processes and interactions as far as possible based on continuous automatic learning is prompting companies to implement machine learning [79].

4.2.5. Complexity Theory and EA

Although Gampfer et al. [6] only forecast a continuation of the trend complexity theory until 2017, this work's investigation shows an ongoing trend. In principle, complexity can be used in EA to understand architectures and measure their complexity [80]. The Gartner hype cycle does not attach any relevance to the topic complexity theory [66].

4.2.6. Modeling Languages and EA

EA modeling languages heavily center around Archimate. Fritscher and Pigneur [81] describe Archimate as a support language for modeling structures within EA. Landthaler et al. [82] describe Archimate as standard within the modeling of modern EA. Archimate does not appear in the Gartner hype cycle for EA [66], although Perez-Castillo et al. [83] show its high relevance in various modeling applications in the field of EA. In empirical studies, Archimate can be identified as a trend within EA.

4.2.7. Big Data and EA

As Lu and Liu [84] point out, there is a steady increase in Big Data technology publications from 2011 onwards. In practice, Big Data can be beneficial in EA decision making and strategy development [85]. The topic modeling conducted in this study confirms a continuation of the trend in EA. The current Gartner hype cycle for EA does not identify Big Data as a trend [66].

4.2.8. Microservices and EA

According to the analysis results of this study, microservices can be classified as a trend within EA. Zimmermann et al. [75] describe microservices as the core area of digitization. In principle, microservices can be defined as individual applications with independent functionalities and the opposite of large monolithic systems [86]. Although not included in the Gartner hype cycle [66], microservice architectures are gaining more importance in practice due to their flexibility and independence [87].

4.2.9. Security and EA

Another EA trend can be found in the area of security or cybersecurity. This trend is understandable since the increasing number of IT-supported processes and the ongoing digitalization of companies mean an equally large degree of cyber-threats [88]. As Halawi et al. [5] describe, empirically, this trend is not new but firmly connected with IT and EA. In the current Gartner hype cycle, the topic of security architecture is currently at a peak and can therefore be considered relevant in practice [66].

4.2.10. Internet of Things and EA

Based on the topic modeling results, the topic area Internet of Things (IoT) [89,90] can be classified as relevant. It is necessary to integrate IoT into the EA as an essential part of 'Industry 4.0' [91]. Zimmermann et al. [75] describe IoT as a core aspect of digitization and a megatrend in digital architecture. Gampfer et al. [6] forecast a continuous growth of technology in the following years and note a high impact on EA. Although IoT was identified as a trend by Gartner in the hype cycle of 2017 [70], the topic is no longer considered as such in the current analysis [66].

4.2.11. Agile Methodology and EA

According to Gartner's hype cycle [66], Agile methodology or Agile Architecture is a trend that has great significance. Although it is seen as a challenge to combine agile methods and EA [92], companies trust efficient software management and higher software maintenance flexibility through agile approaches [93]. The present analysis shows that the topic of agile methodology is also a current trend in EA science.

4.2.12. Continuous Planning and EA

According to the identified word-groups of the topic modeling, Continuous Planning is another current trend within EA. Continuous Planning allows us to meet the challenges of a constant need for new requirements [94]. The planning methodology, which is significantly based on release management, uses agile methods and attempts to create a close integration between planning and implementation of changes [95] and a more frequent provision of new functionalities in the software area. The Gartner hype cycle shows that this topic is currently becoming a trend in practice [66].

5. General Discussion and Conclusions

In this paper, an analysis of issues and trends of the years 2019–2020 in EA was conducted using machine learning. Therefore, a method was developed, making it possible to perform arbitrary topic analyses in empirical studies and practice. The implementation was based on the automatic evaluation of scientific papers published by Buchkremer et al. [7]. To enable a high-performance implementation in the programming language R, various existing functionalities were extended by parallel processing, which proved helpful. Furthermore, relevant findings of using the R programming language in machine learning were collected, which will be helpful in future analyses.

As a result, of the investigation, the trends in the area of EA predicted by Gampfer et al. [6] could be verified and confirmed. Some new topics and research trends were also identified. Finally, the topics of cloud computing, sustainability, capability management, digital transformation, pattern recognition, complexity theory, modeling languages, big data, microservices, security, the internet of things, agile methodology, and continuous planning were identified as relevant. To establish a link to practice, all identified topics were compared with the current Gartner hype cycle for EA. In this way, discrepancies between empiricism and practice were evaluated.

The automated analysis of full texts has proven beneficial to uncover additional insights, especially when there is a limited number of documents to be analyzed. The analysis of 231 full texts yielded similar results compared to 3799 abstracts. However, in terms of the methodology, selecting abstracts only reveals the most relevant content. Therefore, as

a general rule of thumb for any future study applying a methodology similar to ours, we recommend analyzing not full texts but only abstracts—given that the corpus has sufficient size. Still, we also want to emphasize that the context of the state-of-the-art analysis has a significant impact, which is why we suggest iteratively model the approach in a way that is both goal-oriented and fit for purpose.

In summary, this paper shows that the use of state-of-the-art methods such as machine learning are beneficial for the topic and trend analysis, as it allows for an overarching view of the literature in contrast to classical systematic literature analysis.

What is new is that different methods lead to nearly the same result, and the exciting thing is that the investigation of full texts does not provide any significant added value in the trend analysis. It should be noted, however, that we have studied this phenomenon only for EA. By LDA/t-SNE, a known method to capture the complexity of texts, it can be quickly determined how far a trend is already differentiated from other trends. No one has ever done a study on EA using text analysis tools. Moreover, predictions about EA, which have been studied by text analysis, have been confirmed for the first time.

References

1. Just, M.A.; Carpenter, P.A.; Woolley, J.D. Paradigms and processes in reading comprehension. *J. Exp. Psychol. Gen.* **1982**, *111*, 228–238. [CrossRef]
2. Abdallah, A.; Abran, A. Enterprise Architecture Measurement: An Extended Systematic Mapping Study. *Int. J. Inf. Technol. Comput. Sci.* **2019**, *11*, 9–19. [CrossRef]
3. Zachman, J. A Framework for Information Systems Architecture. *IBM Syst. J.* **1987**, *38*, 276–292. [CrossRef]
4. Winter, R.; Fischer, R. Essential Layers, Artifacts, and Dependencies of Enterprise Architecture. In Proceedings of the 2006 10th IEEE International Enterprise Distributed Object Computing (EDOCW'06), Hong Kong, China, 16–20 October 2006; p. 30.
5. Halawi, L.; McCarthy, R.; Farah, J. Where We are with Enterprise Architecture. *J. Inf. Syst. Appl. Res.* **2019**, *12*, 4–13.
6. Gampfer, F.; Jürgens, A.; Müller, M.; Buchkremer, R. Past, current and future trends in enterprise architecture—A view beyond the horizon. *Comput. Ind.* **2018**, *100*, 70–84. [CrossRef]
7. Buchkremer, R.; Demund, A.; Ebener, S.; Gampfer, F.; Jagering, D.; Jurgens, A.; Klenke, S.; Krimpmann, D.; Schmank, J.; Spiekermann, M.; et al. The Application of Artificial Intelligence Technologies as a Substitute for Reading and to Support and Enhance the Authoring of Scientific Review Articles. *IEEE Access* **2019**, *7*, 65263–65276. [CrossRef]
8. Hevner, A.; March, S.; Park, J.; Ram, S. Design science research in information systems. *MIS Q.* **2004**, *28*, 75–105. [CrossRef]
9. Winter, R. Design science research in Europe. *Eur. J. Inf. Syst.* **2008**, *17*, 470–475. [CrossRef]
10. Saint-Louis, P.; Morency, M.C.; Lapalme, J. Defining Enterprise Architecture: A Systematic Literature Review. In Proceedings of the 2017 IEEE 21st International Enterprise Distributed Object Computing Workshop (EDOCW), Quebec City, QC, Canada, 10–13 October 2017; pp. 41–49.
11. ISO/IEC. *Systems and Software Engineering–Architecture Description*; IEEE: Piscataway Township, NJ, USA, 2011.
12. Kitsios, F.; Kamariotou, M. Business strategy modelling based on enterprise architecture: A state of the art review. *Bus. Process Manag. J.* **2019**, *25*, 606–624. [CrossRef]
13. Zhang, M.; Chen, H.; Luo, A. A Systematic Review of Business-IT Alignment Research with Enterprise Architecture. *IEEE Access* **2018**, *6*, 18933–18944. [CrossRef]
14. Ansyori, R.; Qodarsih, N.; Soewito, B. A systematic literature review: Critical success factors to implement enterprise architecture. *Procedia Comput. Sci.* **2018**, *135*, 43–51. [CrossRef]
15. Dumitriu, D.; Popescu, M.A.-M. Enterprise Architecture Framework Design in IT Management. *Procedia Manuf.* **2020**, *46*, 932–940. [CrossRef]

16. Li, L.-S.; Gan, S.-J.; Yin, X.-D. Feedback recurrent neural network-based embedded vector and its application in topic model. *J. Embed. Syst.* **2017**, *2017*, 5. [CrossRef]

17. Horn, N.; Erhardt, M.S.; Di Stefano, M.; Bosten, F.; Buchkremer, R. Vergleichende Analyse der Word-Embedding-Verfahren Word2Vec und GloVe am Beispiel von Kundenbewertungen eines Online-Versandhändlers. In *Künstliche Intelligenz in Wirtschaft & Gesellschaft*; Springer Fachmedien Wiesbaden: Wiesbaden, Germany, 2020; pp. 559–581. ISBN 9783658295509.

18. Wang, Y.; Berwick, R.C. On Formal Models for Cognitive Linguistics. In Proceedings of the 11th IEEE International Conference on Cognitive Informatics and Cognitive Computing, Kyoto, Japan, 22–24 August 2012; pp. 7–17.

19. Fahad, S.K.A.S.A.; Yahya, A.E. Inflectional Review of Deep Learning on Natural Language Processing. In Proceedings of the 2018 International Conference on Smart Computing and Electronic Enterprise, Shah Alam, Malaysia, 11–12 July 2018; pp. 1–4.

20. Samuel, A.L. Some Studies in Machine Learning Using the Game of Checkers. *IBM J. Res. Dev.* **1959**, *3*, 210–229. [CrossRef]

21. Shubhankar, K.; Singh, A.P.; Pudi, V. A Frequent Keyword-Set Based Algorithm for Topic Modeling and Clustering of Research Papers. In Proceedings of the 2011 3rd Conference on Data Mining and Optimization (DMO), Putrajaya, Malaysia, 28–29 June 2011; pp. 96–102.

22. Sun, Y.; Han, J.; Gao, J.; Yu, Y. iTopicmodel: Information Network-Integrated Topic Modeling. In Proceedings of the 2009 Ninth IEEE International Conference on Data Mining, Miami Beach, FL, USA, 6–9 December 2009; pp. 493–502.

23. Hong, L.; Davison, B. Empirical Study of Topic Modeling in Twitter. In Proceedings of the First Workshop on Social Media Analytics, Washington, DC, USA, 25–28 July 2010; pp. 80–88.

24. Blei, D.; Ng, A.; Jordan, M. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.

25. Anwar, W.; Bajwa, I.S.; Choudhary, M.A.; Ramzan, S. An Empirical Study on Forensic Analysis of Urdu Text Using LDA-Based Authorship Attribution. *IEEE Access* **2018**, *7*, 3224–3234. [CrossRef]

26. Haidar, M.A.; Kurimo, M. Lda-Based Context Dependent Recurrent Neural Network Language Model Using Document-Based Topic Distribution of Words. In Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017; pp. 5730–5734.

27. Hussain, A.; Tahir, A.; Hussain, Z.; Sheikh, Z.; Gogate, M.; Dashtipour, K.; Ali, A.; Sheikh, A. Artificial intelligence-enabled analysis of UK and US public attitudes on Facebook and Twitter towards COVID-19 vaccinations (Preprint). *J. Med. Internet Res.* **2020**. [CrossRef]

28. Hao, Y.; Mu, T.; Hong, R.; Wang, M.; Liu, X.; Goulermas, J.Y. Cross-Domain Sentiment Encoding through Stochastic Word Embedding. *IEEE Trans. Knowl. Data Eng.* **2020**, *32*, 1909–1922. [CrossRef]

29. Welbers, K.; Van Atteveldt, W.; Benoit, K. Text analysis in R. *Commun. Methods Meas.* **2017**, *11*, 245–265. [CrossRef]

30. Ooms, J. Pdftools: Text Extraction, Rendering and Converting of PDF Documents. Available online: https://cran.r-project.org/web/packages/pdftools/index.html/ (accessed on 21 April 2021).

31. Khanna, P.; Kumar, S.; Mishra, S.; Sinha, A. Sentiment analysis: An approach to opinion mining from twitter data using r. *Int. J. Adv. Res. Comput. Sci.* **2017**, *8*, 252–256. [CrossRef]

32. Suri, P.; Roy, N.R. Comparison between LDA & NMF for Event-Detection from Large Text Stream Data. In Proceedings of the 3rd IEEE International Conference on "Computational Intelligence and Communication Technology" (IEEE-CICT 2017), Ghaziabad, India, 9–10 February 2017; pp. 1–5.

33. Yaram, S. Machine Learning Algorithms for Document Clustering and Fraud Detection. In Proceedings of the 2016 IEEE International Conference on Data Science and Engineering (ICDSE), Cochin, India, 23–25 August 2016; pp. 1–6.

34. Feinerer, I. An introduction to text mining in R. *Newsl. R Proj.* **2008**, *8*, 19.

35. Wang, X.; Lee, M.; Pinchbeck, A.; Fard, F.H. Where Does LDA Sit for GitHub? In Proceedings of the 2019 34th IEEE/ACM International Conference on Automated Software Engineering Workshop (ASEW), San Diego, CA, USA, 11–15 November 2019; pp. 94–97.

36. Hidayat, E.Y.; Firdausillah, F.; Hastuti, K.; Dewi, I.N. Azhari Automatic Text Summarization Using Latent Dirichlet Allocation (LDA) for Document Clustering. *Int. J. Adv. Intell. Inform.* **2015**, *1*, 132–139. [CrossRef]

37. O'Callaghan, D.; Greene, D.; Carthy, J.; Cunningham, P. An Analysis of the Coherence of Descriptions in Topic Modeling. *Expert. Syst. Appl.* **2015**, *42*, 5645–5657. [CrossRef]

38. Xu, A.; Qi, T.; Dong, X. Analysis of the Douban online review of the MCU: Based on LDA topic model. *J. Phys. Conf. Ser.* **2020**, *1437*, 012102. [CrossRef]

39. Huang, L.; Ma, J.; Chen, C. Topic Detection from Microblogs Using T-LDA and Perplexity. In Proceedings of the 2017 24th Asia-Pacific Software Engineering Conference Workshop, Nanjing, China, 4–8 December 2017; pp. 71–77.

40. Chen, Q.; Yao, L.; Yang, J. Short Text Classification Based on LDA Topic Model. In Proceedings of the 2016 International Conference on Audio, Language and Image Processing (ICALIP), Shanghai, China, 11–12 July 2016; pp. 749–753.

41. Shiryaev, A.; Dorofeev, A.; Fedorov, A.; Gagarina, L.; Zaycev, V. LDA Models for Finding Trends in Technical Knowledge Domain. In Proceedings of the 2017 IEEE Conference on Russian Young Researchers in Electrical and Electronic Engineering (ElConRus), St. Petersburg and Moscow, Russia, 1–3 February 2017; pp. 551–554.

42. Shao, J. Linear Model Selection by Cross-Validation. *J. Am. Stat. Assoc.* **1993**, *88*, 486–494. [CrossRef]

43. Pleplé, Q. Perplexity To Evaluate Topic Models. Available online: http://qpleple.com/perplexity-to-evaluate-topic-models/ (accessed on 20 April 2021).

44. Slutsky, A.; Hu, X.; An, Y. Tree Labeled LDA: A Hierarchical Model for Web Summaries. In Proceedings of the 2013 IEEE International Conference on Big Data, Silicon Valley, CA, USA, 6–9 October 2013; pp. 134–140.

45. Jiang, J. Modeling Syntactic Structures of Topics with a Nested HMM-LDA. In Proceedings of the 2009 Ninth IEEE International Conference on Data Mining, Miami Beach, FL, USA, 6–9 December 2009; pp. 824–829.

46. Jingrui, Z.; Qinglin, W.; Yu, L.; Yuan, L. A Method of Optimizing LDA Result Purity Based on Semantic Similarity. In Proceedings of the 2017 32nd Youth Academic Annual Conference of Chinese Association of Automation (YAC), Hefei, China, 19–21 May 2017; pp. 361–365.

47. Murzintcev, N. Select Number of Topics for LDA Model. 2019. Available online: https://cran.r-project.org/web/packages/ldatuning/vignettes/topics.html (accessed on 20 April 2021).

48. Griffiths, T.; Steyvers, M. Finding Scientific Topics. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 5228–5235. [CrossRef] [PubMed]

49. Cao, J.; Xia, T.; Li, J.; Zhang, Y.; Tang, S. A density-based method for adaptive LDA model selection. *Neurocomputing* **2009**, *72*, 1775–1781. [CrossRef]

50. Arun, R.; Suresh, V.; Madhavan, C.V.; Murthy, M.N. On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations. In Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining, Singapore, 11–14 May 2020; pp. 391–402.

51. Gao, S.; Janowicz, K.; Couclelis, H. Extracting urban functional regions from points of interest and human activities on location-based social networks. *Trans. GIS* **2017**, *21*, 446–467. [CrossRef]

52. Grün, B.; Hornik, K. topicmodels: An R Package for Fitting Topic Models. *J. Stat. Softw.* **2011**, *40*, 1–30. [CrossRef]

53. Meyer, D.; Hornik, K.; Feinerer, I.; Hornik, K.; Meyer, D. Text Mining Infrastructure in R. *J. Stat. Softw.* **2008**, *25*, 1–54.

54. Li, C.; Feng, S.; Zeng, Q.; Ni, W.; Zhao, H.; Duan, H. Mining Dynamics of Research Topics Based on the Combined LDA and WordNet. *IEEE Access* **2018**, *7*, 6386–6399. [CrossRef]

55. Phan, X.-H.; Nguyen, L.-M.; Horiguchi, S. Learning to Classify Short and Sparse Text & Web with Hidden Topics from Large-scale Data Collections. In Proceedings of the 17th international conference on World Wide Web, Beijing, China, 21–25 April 2008; pp. 91–100.

56. van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.

57. Yasaswi, J.; Kailash, S.; Chilupuri, A.; Purini, S.; Jawahar, C.V. Unsupervised learning based approach for plagiarism detection in programming assignments. *ACM Int. Conf. Proceeding Ser.* **2017**, 117–121.

58. Pezzotti, N.; Thijssen, J.; Mordvintsev, A.; Höllt, T.; Van Lew, B.; Lelieveldt, B.P.F.; Eisemann, E.; Vilanova, A. GPGPU Linear Complexity t-SNE Optimization. *IEEE Trans. Vis. Comput. Graph.* **2020**, *26*, 1172–1181. [CrossRef] [PubMed]

59. Chang, C.-Y.Y.; Lee, S.-J.J.; Lai, C.-C.C. Weighted word2vec Based on the Distance of Words. In Proceedings of the 2017 International Conference on Machine Learning and Cybernetics, ICMLC 2017, Ningbo, China, 9–12 July 2017; Volume 2, pp. 563–568.

60. Pezotti, N.; Lelieveldt, B.P.F.; van der Maaten, L.; Hölt, T.; Eisemann, E.; Vilanova, A. Approximated and User Steerable tSNE for Progressive Visual Analytics. *IEEE Trans. Vis. Comput. Graph.* **2017**, *23*, 1739–1752. [CrossRef]

61. Van Der Maaten, L. Accelerating t-SNE using Tree-based Algorithms. *J. Mach. Learn. Res.* **2015**, *15*, 3221–3245.

62. Krijthe, J.; Van Der Maaten, L. Package "Rtsne". 2018. Available online: https://cran.r-project.org/web/packages/Rtsne/index.html (accessed on 20 April 2021).

63. Toomet, O.; Henningsen, A. Sample Selection Models in R: Package sampleSelection. *J. Stat. Softw.* **2008**, *27*, 1–23. [CrossRef]

64. Liaw, A.; Wiener, M. Classification and Regression by randomForest. *R News* **2002**, *2*, 18–22.

65. Séaghdha, D. Latent Variable Models of Selectional Preference. In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Uppsala, Sweden, 11–16 July 2010; pp. 435–444.

66. Allega, P.; Santos, J. *Hype Cycle for Enterprise Architecture 2019*; Gartner: Stamford, CT, USA, 2019.

67. Armbrust, M.; Fox, A.; Griffith, R.; Joseph, A.D.; Katz, R. A view of cloud computing. *Commun. ACM* **2010**, *53*, 50–58. [CrossRef]

68. Dillion, T.; Wu, C.; Chang, E. Cloud Computing: Issues and Challenges. In Proceedings of the 2010 24th IEEE International Conference on Advanced Information Networking and Applications, Perth, Australia, 20–23 April 2010; pp. 27–33.

69. Jadeja, Y.; Modi, K. Cloud Computing—Concepts, Architecture and Challenges. In Proceedings of the 2012 International Conference on Computing, Electronics and Electrical Technologies [ICCEET], Nagercoil, India, 21–22 March 2012; pp. 877–880.

70. Blosch, M.; Burton, B. *Hype Cycle for Enterprise Architecture*; Gartner: Stamford, CT, USA, 2017.

71. Manzhynski, S.; Figge, F. Coopetition for sustainability: Between organizational benefit and societal good. *Bus. Strateg. Environ.* **2020**, *29*, 827–837. [CrossRef]

72. Espahbodi, L.; Espahbodi, R.; Juma, N.; Westbrook, A. Sustainability priorities, corporate strategy, and investor behavior. *Rev. Financ. Econ.* **2019**, *37*, 149–167. [CrossRef]

73. Lapalme, J.; Gerber, A.; Van Der Merwe, A.; Zachman, J.; De Vries, M.; Hinkelmann, K. Exploring the future of enterprise architecture: A Zachman perspective. *Comput. Ind.* **2016**, *79*, 103–113. [CrossRef]

74. Bauer, W.; Hämmerle, M.; Schlund, S.; Vocke, C. Transforming to a hyper-connected society and economy—Towards an "Industry 4.0". *Procedia Manuf.* **2015**, *3*, 417–424. [CrossRef]

75. Zimmermann, A.; Schmidt, R.; Sandkuhl, K. Multiple Perspectives of Digital Enterprise Architecture. In Proceedings of the 14th International Conference on Evaluation of Novel Approaches to Software Engineering (ENASE 2019), Crete, Greece, 4–5 May 2019; pp. 547–554.

76. Korhonen, J.J.; Halen, M. Enterprise Architecture for Digital Transformation. In Proceedings of the 2017 IEEE 19th Conference on Business Informatics, Thessaloniki, Greece, 24–26 July 2017; pp. 349–358.

77. Zimmermann, A.; Schmidt, R.; Sandkuhl, K.; Jugel, D.; Bogner, J.; Möhring, M. Evolution of Enterprise Architecture for Digital Transformation. In Proceedings of the 2018 IEEE 22nd International Enterprise Distributed Object Computing Workshop, Stockholm, Sweden, 16–19 October 2018; pp. 87–96.

78. Kaidalova, J.; Sandkuhl, K.; Seigerroth, K. How Digital Transformation affects Enterprise Architecture Management—A case study. *Int. J. Inf. Syst. Proj. Manag.* **2018**, *6*, 5–18.

79. Sapna, R.; Monikarani, H.G.; Mishra, S. Linked Data through the Lens of Machine Learning: An Enterprise View. In Proceedings of the 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, India, 20–22 February 2019; pp. 1–6.

80. Schuetz, A.; Widjaja, T.; Kaiser, J. Complexity in Enterprise Architecture: Conceptualization and Introduction of A Measure from a System Theoretic Perspective. In Proceedings of the 21st European Conference on Information Systems, Utrecht, The Netherlands, 5–8 June 2013; pp. 1–12.

81. Fritscher, B.; Pigneur, Y. Business IT Alignment from Business Model to Enterprise Architecture. In Proceedings of the International Conference on Advanced Information Systems Engineering, London, UK, 20–24 June 2011; pp. 4–15.

82. Landthaler, J.; Uludag, Ö.; Bondel, G.; Elnaggar, A.; Nair, S.; Matthes, F. A Machine Learning Based Approach to Application Landscape Documentation. In Proceedings of the IFIP Working Conference on The Practice of Enterprise Modeling, Vienna, Austria, 31 October–2 November 2018; pp. 71–85.

83. Perez-Castillo, R.; Ruiz, F.; Piattini, M.; Ebert, C. Enterprise Architecture. *IEEE Softw.* **2019**, *36*, 12–19. [CrossRef]

84. Lu, L.; Liu, J. The Major Research Themes of Big Data Literature. In Proceedings of the 2016 IEEE International Conference on Computer and Information Technology, Nadi, Fiji, 8–10 December 2016; pp. 586–590.

85. Veneberg, R.K.; Iacob, M.E.; van Sinderen, M.J.; Bodenstaff, L. Enterprise Architecture Intelligence Combining Enterprise Architecture and Operational Data. In Proceedings of the 2014 IEEE International Enterprise Distributed Object Computing Conference, Ulm, Germany, 1–5 September 2014; pp. 22–31.

86. Bogner, J.; Zimmermann, A. Towards Integrating Microservices with Adaptable Enterprise Architecture. In Proceedings of the 2016 IEEE 20th International Enterprise Distributed Object Computing Workshop (EDOCW), Vienna, Austria, 5–9 September 2016; pp. 1–6.

87. Taibi, D.; Lenarduzzi, V.; Pahl, C. Processes, motivations, and issues for migrating to microservices architectures: An empirical investigation. *IEEE Cloud Comput.* **2017**, *4*, 22–32. [CrossRef]

88. Larno, S.; Seppänen, V.; Nurmi, J. Method Framework for Developing Enterprise Architecture Security. *Complex Syst. Inform. Model. Q.* **2019**, *117*, 57–71. [CrossRef]

89. Atzori, L.; Iera, A.; Morabito, G. The internet of things: A survey. *Comput. Netw.* **2010**, *54*, 2787–2805. [CrossRef]

90. Gubbi, J.; Buyya, R.; Marusic, S.; Palaniswami, M. Internet of Things (IoT): A vision, architectural elements, and future directions. *Futur. Gener. Comput. Syst.* **2013**, *29*, 1645–1660. [CrossRef]

91. Schmidt, R.; Möhring, M.; Härting, R.-C.; Reichstein, C.; Neumaier, P.; Jozinovic, P. Industry 4.0—Potentials for Creating Smart Products: Empirical Research Results. In Proceedings of the International Conference on Business Information Systems, Poznań, Poland, 24–26 June 2015; pp. 16–27.

92. Canat, M.; Català, N.; Jourkovski, A.; Petrov, S.; Wellme, M.; Lagerström, R. Enterprise Architecture and Agile Development Friends or Foes? In Proceedings of the 2018 IEEE 22nd International Enterprise Distributed Object Computing Workshop, Stockholm, Sweden, 16–19 October 2018; pp. 176–183.

93. Xiong, W.; Carlsson, P.; Lagerström, R. Re-Using Enterprise Architecture Repositories for Agile Threat Modeling. In Proceedings of the 2019 IEEE 23rd International Enterprise Distributed Object Computing Workshop (EDOCW), Paris, France, 28–31 October 2019; pp. 118–127.

94. Fitzgerald, B.; Stol, K.-J. Continuous software engineering and beyond: Trends and challenges. In Proceedings of the 1st International Workshop on Rapid Continuous Software Engineering, Hyderabad, India, 31 May–7 June 2014; pp. 1–9.

95. Knight, R.; Rabideau, G.; Chien, S.; Engelhardt, B.; Sherwoord, R. Casper: Space exploration through continuous planning. *IEEE Intell. Syst.* **2001**, *16*, 70–75.