Past Achievements and Future Opportunities in Combining Conceptual Modeling with VR/AR: A Systematic Derivation

Fabian Muff ● and Hans-Georg Fill ●

University of Fribourg, Research Group Digitalization and Information Systems
`fabian.muff|hans-georg.fill@unifr.ch`

Abstract. Despite the increased interest in virtual and augmented reality in recent years, they are not yet mainstream technologies for everyday use in industry. We argue that a promising approach to facilitate the application of virtual and augmented reality is to combine it with conceptual modeling. In this paper, we thus conducted a systematic literature review on the combination of conceptual modeling with virtual and augmented reality within the last two decades. For this purpose, we reverted to a manual literature search, computational topic modeling, and an expert-driven classification process. This analysis highlights the areas in which such a combination of virtual and augmented reality and conceptual modeling already exists, as well as the aspects that are not yet covered or that would offer opportunities for further research.

Keywords: Bibliometric Analysis · Augmented Reality · Virtual Reality · Conceptual Modeling · Latent Dirichlet Allocation.

1 Introduction

Throughout the last years, the application of virtual reality (VR) and augmented reality (AR) technologies to business scenarios has been increasingly studied by the research community [37]. In VR, the user's perception is based entirely on *virtual* information in a *virtual* world. In AR, computer-generated information is provided to the user in addition to data collected from real life, enhancing the user's perception of *reality*. Due to the recent technological progress [42], affordable and mobile VR and AR devices became widely available and enabled the broad application of the technology in industrial scenarios such as for maintenance tasks or training [16]. A study from PwC estimates that VR and AR will deliver an enormous boost to the global economy until 2030 [13]. Further, a recent study from 2022 indicates that a majority of U.S. executives are highly interested in exploring AR and VR as a foundation for the Metaverse [31].

However, the development of such applications still requires considerable technical know-how. Thus, the provision of *systematic* and at the same time *flexible* approaches for *designing* VR and AR applications is regarded as a prerequisite for a more widespread adoption, cf. [41]. Conceptual modeling, e.g., as

used in enterprise modeling, may serve as a solution for both aspects [35]. On the one hand it aims to reduce complexity by structuring a particular domain for improving human understanding [28,8]. This may involve the use of novel technologies, e.g., in three-dimensional space [2]. On the other hand, the knowledge made explicit in such models may be processed algorithmically, e.g., as found in model-driven engineering for easing the creation of software applications [7] or for fueling knowledge into existing applications [14].

This leads us to propose two main directions for virtual and augmented reality in relation to conceptual modeling. First, the use of functionalities of VR and AR for modeling itself. We will denote this as *VR/AR-assisted modeling*. Second, the incorporation of information from the model space into VR or AR applications, which we will denote as *knowledge-based VR/AR*. This second direction includes both design-time and run-time aspects, i.e., the modeling and model-driven generation of VR/AR applications as well as the fueling of model contents into existing VR/AR applications.

The paper at hand aims to explore the multitude of approaches proposed in academic research for combining conceptual modeling with virtual and augmented reality. Despite numerous contributions, no structured analysis of them has been undertaken so far to the best of our knowledge. Therefore, we conducted a systematic literature review on the combination of conceptual modeling with VR and AR within the last two decades. Further, we employed a computational content analysis to identify distinct research streams that have been explored in this field. Finally, we analyzed and refined the results of our analysis with the help of expert classification. The contribution of this study is to provide a comprehensive understanding of the main contributions for combining conceptual modeling with virtual and augmented reality, identify the main topics that have been studied in the past, and highlight the areas that require further research.

The remainder of the paper is structured as follows. In Section 2, we will describe the research methodology used for the review. Section 3 will describe the literature search results, which were used as input for Latent Dirichlet Allocation to computationally derive a first set of topics. Further, it will be shown how these topics have been refined using expert classification and the allocation of papers to the final set of topics. Finally, we will discuss the results of the analysis and derive points for future research in Section 4, as well as related work in Section 5.

2   Research Methodology

The methodology that we followed in this study is mainly based on the recommendations by Kitchenham [19] for conducting systematic literature reviews. This includes the three phases *Planning*, *Conducting*, and *Reporting*. The planning phase includes the identification of the need of the review as described above, as well as the definition of a research protocol, as shown in Fig. 1. The research protocol describes each step of the review process according to Booth et al. [5]. For the conduction phase, we further reverted to the guidelines by Webster and Watson [39], who recommend in particular the screening of dedicated

outlets and the application of forward- and backward searches. In addition, we performed a computational literature analysis followed by an expert classification for deriving the topics of the different research streams.
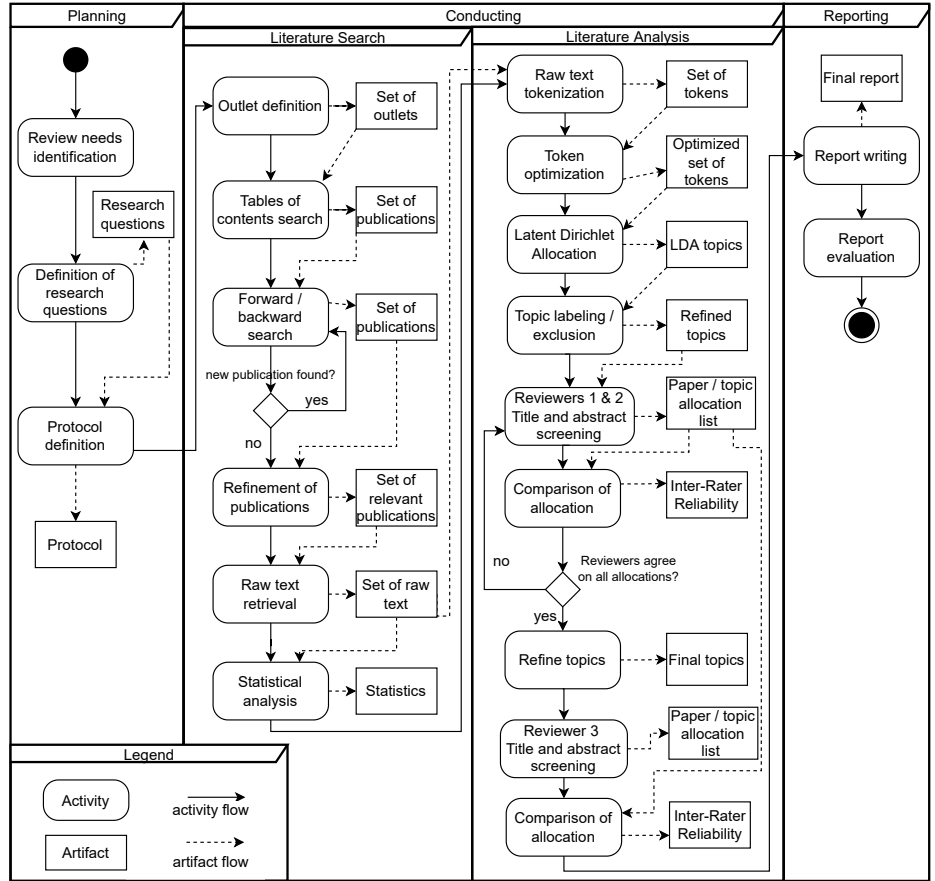

Fig. 1: Description of the research protocol. The protocol is divided into the three main areas as proposed by Kitchenham [19]. The process shows the undertaken steps together with the resulting artifacts.

2.1 Aims and Scope of the Study

The aim of this work is to identify the main research topics that combine conceptual modeling with virtual and augmented reality. Further, the study shall give detailed insights on the proposed concepts of *VR/AR-assisted modeling* and *knowledge-based VR/AR*. The investigated time frame includes academic papers

that have been published between the years 2000 and the first half of 2022, with the goal to show the most recent research developments in these areas.

2.2 Literature Collection

For identifying the main research contributions on combining conceptual modeling with VR/AR, we reverted primarily to the method proposed by Webster and Watson [39] to determine an initial set of relevant sources. We describe in the following the steps as shown in the *Literature Search* section of the research protocol in Figure 1. We first identified the nine most important outlets in the field of conceptual modeling, based on a recent review by Härer and Fill [17]. According to this source, many topics in conceptual modeling are strongly related to enterprise modeling. For example, business / business process models, or data models and schemas. In addition, we added five outlets in the area of *Business Informatics* and *Information Systems* with potentially relevant contributions (*Outlet definition*) - see Figure 2.
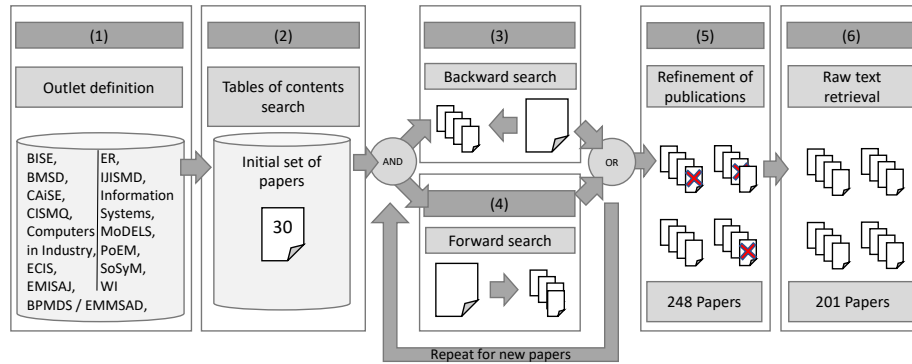

Fig. 2: Data collection process following Webster & Watson [39]: (1) Identification of the relevant outlets. (2) Screening of tables of contents. (3 and 4) Iterative forward- and backward search, based on the newly added relevant papers. (5) Selection refinement by a more profound inspection of the selected papers, resulting in 201 relevant papers of which the raw texts were retrieved (6).

We analyzed the tables of contents of the outlets to identify relevant contributions (*Tables of contents search*). For each of the found contributions, we applied a forward- and backward search, i.e., finding for each paper relevant cited and citing articles using *semanticscholar.org* and *google.scholar.com* (*Forward / backward search*). We repeated this step until no new papers were found. We then reviewed the set of papers for excluding wrongly selected papers (*Refinement of publications*). Finally, we retrieved the raw texts of the papers for further analysis (*Raw text retrieval*), and calculated quantitative indicators of the set of relevant papers (*Statistical analysis*). A detailed description of the results of this literature search will follow in Section 3.1.

2.3   Content-based Data Analysis

To derive the research contribution in terms of previously studied topics, we conducted a computational analysis and complemented it with an expert-driven classification of relevant papers into distinct topical domains. The following steps refer to the *Literature Analysis* section in the research protocol as defined in Figure 1.

Computational Data Analysis: For the compilation of an initial set of topics describing the main directions in the papers of the literature analysis, we resorted to the technique of *topic modeling*. This required the tokenization of the raw text of each document and preliminary tasks such as minimal stemming, stopword filtering, case transformation, synonym replacement, and single character filtering (*Raw text tokenization* and *Token optimization*). On this basis, we performed an LDA (*Latent Dirichlet Allocation*), which is an established method in computational topic modeling that has been successfully applied in previous literature reviews [17,27]. For the LDA, we used an iterative approach, which tries to optimize the hyperparameters for the topic generation, i.e., the number of topics, alpha and beta heuristics, as well as some evaluation measures like the topic coherence and the topic perplexity [23]. At the end of this iterative process, we decided on pursuing an analysis with ten topics. Details and results of this process will be described in Section 3.2.

Expert Analysis and Refinement: The topics proposed by the LDA were then labeled and refined manually by the authors and one external expert in an iterative procedure. By looking at the different words allocated by the LDA to the topics and by considering the list of the most probable allocated topic for each paper, we allocated labels to each topic (*Topic labeling/exclusion*). After this first topic labeling, the papers were manually allocated to one of the topics. As proposed by Vessey et al. [38], two experts allocated the papers independently from each other to exactly derive one topic by screening the titles of the papers. Then, each disagreement was discussed iteratively to find a consensus based on the abstracts of the contributions (*Title and abstract screening*).

For checking the reviewers' agreement, we calculated the inter-rater reliability (IRR) by using *Cohen's Kappa* (κ) [12] (*Comparison of allocation*). These steps were repeated until reviewers one and two reached an agreement on their allocation. Thereby, the topics could also be refined by renaming them or by merging similar topics, if found necessary, during the manual evaluation (*Refine topics*). This resulted in the final list of topics.

As an extension of the labeling process for two reviewers proposed by [38], a third reviewer manually assigned the papers to the final topics derived by reviewers one and two through a title and abstract screening (*Reviewer 3 Title and abstract screening*). The goal was to validate the reliability of the final assignment of reviewers one and two. Again, the IRR between the decision of the third reviewer and the joint assignment of reviewers one and two was calculated (*Comparison of allocation*).

3   Results

In this section, we describe the results obtained from the literature search process defined in Section 2.2, as well as of the content-based data analysis process described in Section 2.3.

3.1   Literature Search for Combining Virtual and Augmented Reality With Conceptual Modeling

As described in the methodology section above, we initially examined 15 outlets. We went manually through the outlets' tables of contents and searched for the terms *AR*, *VR*, *augmented reality*, *virtual reality*, and *3D*. The abstracts of the resulting papers were used to decide whether they are relevant for the analysis. A paper was considered relevant if it addressed at least one of the above areas, as well as conceptual modeling. In the context of this paper, we regard conceptual modeling in a broad sense, i.e., relating to the formal description of some aspect of the world around us based on a schema for the purpose of human understanding and communication [28,17]. The initial screening of these outlets led to a list of 30 relevant papers. The forward- and backward searches resulted in a list of 248 papers. Subsequently, a more detailed analysis of whether each paper indeed involved conceptual modeling was performed. Through a manual review of abstracts and/or full texts, we identified and excluded papers that are not based on a schema. This process resulted in a final list of 201 relevant papers. Due to space restrictions, the documentation of the whole process is available in the online Appendix A.

Regarding the number of publications over time, there is a clearly increasing trend in the number of published papers with a slope of $m = 0.4675$ when excluding the values from 2022 – see the right side of Figure 3. In addition, the publications are distributed over many outlets. Only 30 out of the 201 relevant papers were published in one of the initially defined 15 outlets. In total, the 201 papers were published in 143 different outlets and only 12 of these outlets had three or more publications in the observed time span – see left side of Figure 3. From the initial 15 outlets only *BMSD*, *CAiSE*, *ECIS* and *Computers in Industry* have three or more relevant publications.

3.2   Computational Topic Modeling

For the content-based analysis, we used computational topic modeling. Two common methods are LDA (Latent Dirichlet Allocation) [4] and NMF (Non-Negative Matrix Factorization) [36], which have been used for a long time. NMF is increasingly used for document collections with large noise, e.g., prepositions, abbreviations, or slang words. LDA can struggle with noise, but can be used in an iterative, semi-supervised way to produce a good ground truth of topics [11]. When the ground assumption of non-correlating topics does not hold, alternatives such as CTM (Correlated Topic Models) and STM (Structural Topic Models) may be used. CTM relaxes the assumption of independent topics [20].
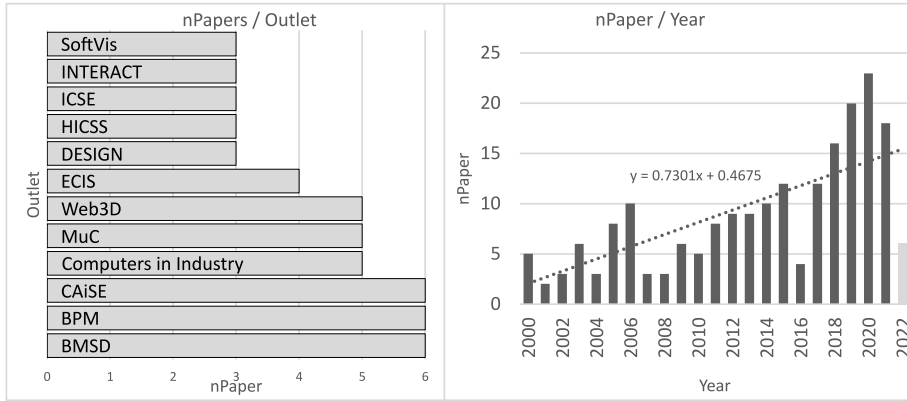
Fig. 3: Outlets with three or more relevant papers obtained from the literature search (left) and number of articles published per year with a linear trend line (right). The year 2022 was not considered since not all publications were yet available at the cut-off date of the analysis.

STM is a mixture model, in which each document can belong to a mixture of the specified k topics [34] and is often used for documents containing questionnaire data with open-ended questions. For datasets consisting mainly of short texts such as social media posts, specific methods have been developed, among others, SATM [33], or ETM [32]. Since our dataset consists exclusively of scientific papers, we decided to exclude the recent methods for short texts. We assumed that the topics in our analysis should be unique and independent. Further, we aimed to achieve the clearest possible assignment of a paper to a topic. Thus, we chose to exclude CTM and STM. Finally, we selected the traditional LDA as our basic methodology, which has been validated by several empirical studies as being capable of extracting semantically meaningful topics from texts and categorizing texts according to these topics [6,9,22,24].

We used MALLET (**MA**chine **L**earning for **L**anguag**E T**oolkit), as well as the LDA implementation that is part of RapidMiner Studio 9.5. As topic modeling is an unsupervised process, the evaluation of the results of an LDA presents some challenges. First, the quality of topics can be measured and compared by the *coherence value* of the topics [9,24]. It gives an overview of the semantic interpretability of the topics [24]. Second, *perplexity* measures of how well a probability model predicts a given sample. However, Chang et al. [9] showed, that human judgement and perplexity often do not correlate. Since the goal of our analysis was to get distinguishable topics that are human-interpretable, we focused on coherence rather than perplexity. Regarding the number of tokens assigned to each topic (topic size), there is no optimal topic value according to Mimno et al. [24]. However, smaller topics seem to be of better quality.

Based on this information, we performed different iterations of LDA for seven to thirteen topics and compared the corresponding average coherence values C_{UMass}. The values varied between -3.369 and -4.257, where lower values

are considered as better [24]. Since C_{UMass} decreases rapidly at the beginning and remains relatively stable between the LDA with ten and 13 topics, we decided to analyze the model with ten topics having an average coherence value of $C_{UMass} = $ -4.203. Further, we chose five tokens per topic as topic size. The left side of Fig. 4 shows the ten initial topics delivered by the LDA with the five most weighted words for each topic. For example, *Topic 0* has the most weighted terms *system*, *maintenance*, *context*, *user*, and *information*. The order of the topics has no specific meaning. Further, the LDA delivered a list of all papers with the according allocation probability to the different topics. Over a set of documents, each document d is represented by a statistical distribution θ_d over its different topics. That means, that each topic has a certain probability or weight for d, and for each topic k a distribution of words $\theta_{d,k}$ [3]. The hidden variables of the distributions are computed with the Gibbs sampling scheme by using parallel processing, where the weights per word are determined to maximize their probability of occurring in a given topic [29].

Only 27 (13%) papers had a most probable allocation to one of the ten topics of < 0.5. The remaining 174 papers had a most probable allocation of ≥ 0.5 and 101 papers (50%) had a most probable allocation of ≥ 0.7.

For our study, the LDA was intended as an objective ground truth for further analysis. For this reason, we do not elaborate further on the original topics of the LDA, but rather focus on the additional findings through the manual topic refinement and the paper assignment process in the next section.

3.3   Topics and their Contribution

Since there is almost no human interference, LDA is a relatively objective process. The results of the LDA require however some interpretation and contextualization to increase their value. In this section, we therefore show the results of the labeling and revision of the ten initial topics through expert assessment, as well as the allocation of the different papers to these topics.

Refined Topics: For the labeling of the ten topics, the two authors considered the words allocated to the topics by the LDA together with the list of the most probable topic for each paper as defined by the *Literature Analysis* section of the research protocol visible in Figure 1. Thereby, the most probable topic for each paper is the one to which the LDA assigns the paper with the highest probability.

We then decided commonly on a label for each LDA topic. Some topics required specific treatment: *Topic 8* consists of the terms *system*, *service*, *glass*, *smart*, and *information*. This indicates a focus on smart glasses, which have been explicitly researched in several of the selected papers. Since this is a hardware-specific category, it was decided to exclude this topic from the subsequent steps. Further, *Topic 7* and *Topic 9* were considered as similar in terms of their research area. The terms *sysml*, *uml*, *diagram*, and *visualization* were interpreted as related to software or system visualization. Thus, they were merged in one topic with the label *Software and System Visualization*. As shown in Figure 4
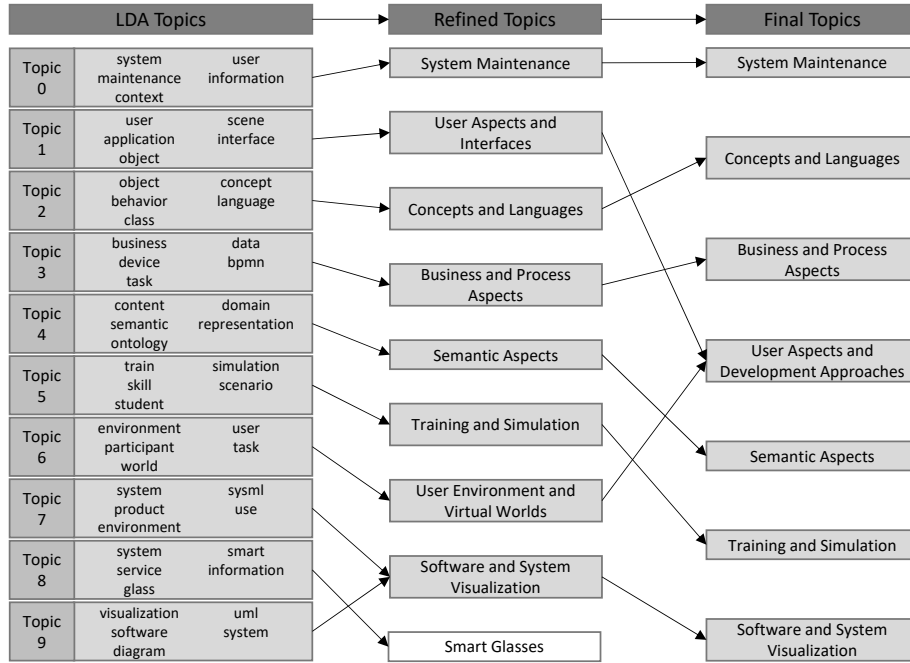
Fig. 4: Visualization of the topic evolution over the different refinement steps. **LDA Topics:** Initial topics delivered by the LDA analysis with the five most weighted words each. The order of the topics has no systematic ranking. **Refined Topics:** Topics according to the expert topic labeling. **Final Topics:** Final seven topics after the last refinement step.

(Refined Topics), out of the 10 LDA topics, eight topics were kept for the further analysis.

Paper Allocation and Final Topics: After the initial topic labeling, the papers were manually allocated to one of the topics by the two authors to express the core focus of each paper through a single assignment. The resulting inter-rater reliability (IRR) in the form of *Cohen's Kappa* [12] was $\kappa = 0.617$ after the first allocation. According to Landis and Koch [21] values between 0.6 and 0.8 indicate a substantial agreement. After agreeing on the allocation of papers to the various topics, reviewers one and two discussed and refined the topics again. Thereby, the topics *User Aspects and Interfaces*, and *User Environment and Virtual Worlds* were merged into one topic entitled *User Aspects and Development Approaches*, which was regarded as a more suitable, common label when inspecting the underlying papers. This resulted in the final set of the seven topics visible in Figure 4 (Final Topics).

As shown in Table 1, 63 papers (31.3%) were allocated to the topic *Business and Process Aspects*, followed by 37 papers (18.4%) allocated to *Software and System Visualization*, 31 papers to *User Aspects and Development Approaches* (15.4%), 26 papers to *Semantic Aspects* (12.9%), 23 papers to *Training and*

Simulation (11.4%), 14 papers to *Concepts and Languages* (7%), and 7 papers to *System Maintenance* (3.5%). We will discuss the final topics and its main contributions in more detail in Section 4.

Topic Name	nPapers	Dist. 2000-2022
Business and Process Aspects	63	
Concepts and Languages	14	
Semantic Aspects	26	
Software and System Visualization	37	
System Maintenance	7	
Training and Simulation	23	
User Aspects and Development Approaches	31	
Total	201	

Table 1: Distribution of the 201 papers (nPapers) over the final seven topics in alphabetical order after the final allocation by reviewers one and two, and a visual distribution of the papers over time.

Quality Audit: For additional quality assurance it was reverted to a third reviewer who assigned the 201 papers to the final seven topics by considering only the titles of the papers. The resulting IRR in comparison to the final allocation of reviewers one and two was $\kappa = 0.520$, indicating moderate agreement [21]. Following the title screening, the third reviewer was then additionally presented with the abstracts of those papers to which he had not assigned the same topic as reviewers one and two, without revealing the assignments of the other reviewers to him. He could then decide whether to assign a different topic or maintain his selection. After this step, the resulting IRR in comparison to the final allocation of reviewers one and two increased to $\kappa = 0.655$, which indicates a substantial agreement [21].

4   Discussion

With the insights gained above, we can now advance to the discussion of our findings in regard to the initially proposed directions of *VR/AR-assisted modeling* and *knowledge-based VR/AR*. Further, we will highlight areas that have not yet been covered by research.

The main research contributions are the following: First, the research question on the main contributions of combining conceptual modeling with virtual and augmented reality can be answered directly in terms of the literature search (Section 3.1). The 201 relevant papers are distributed across many different outlets, with no outlet dominating. The research area reviewed in this paper shows a clearly increasing trend in publications, which is a promising sign for future research. Second, by discussing the results from Section 3.3 and by reflecting on

possible application areas that would push research and industry forward, we can identify the main topics that have been studied in the past, and highlight the areas that require further research. Regarding the identification of the main topics, we need to consider the final topics, their interpretation, the allocation of the papers to these topics by the reviewers, as well as some exemplary contributions to these topics. It shall be noted that the labeling of the different topics is a subjective task and that other reviewers may allocate different labels. However, we tried to mitigate this subjective factor by conducting the objective LDA analysis as a ground truth for further investigation. Further, the labeling of the different topics was conducted by two reviewers in an iterative process and dissenting opinions were discussed. In the following, we discuss the final topics and their interpretation as well as some sample papers that the reviewers assigned to these topics.

Papers assigned to the topic *Business and Process Aspects* deal mainly with business process management. With regard to the traditional business process life cycle [40], Design/Analysis [R90, R169, R244][1], Configuration [R72, R193], and Enactment [R152, R173, R219] have been subject of research related to VR/AR. However, we could not yet discover research on the *Evaluation* of business processes related to VR/AR. This is surprising, since VR/AR devices provide a variety of sensor data that would be predestined for process evaluation. The areas *VR/AR-assisted modeling*, e.g., [R96] and *knowledge-based VR/AR*, e.g., [R1, R158] are both present in research.

Concepts and Languages contain contributions like languages for modeling VR/AR systems, or for authoring VR/AR content. Thereby, we could identify the three main streams: *content creation* [R88, R186], *metamodeling* [R147, R26], and *concepts for model-driven code generation* [R184, R119]. All these research streams can be related to *knowledge-based VR/AR*, either for design-time, or for run-time, i.e., real-time content creation. What seems to have not been covered so far is the combination of *knowledge-based VR/AR* and *VR/AR-assisted modeling* in a generic way, e.g., for allowing VR-based model-driven engineering of VR/AR applications, which could be useful for simulating the interaction with 3D environments in VR prior to their realization using AR.

For structuring the papers allocated to the topic of *Semantic Aspects*, we found that they can be related to the seven components of the semantic web framework derived in [15]. Considering these components, we found approaches for *Querying and Reasoning* [R4, R148, R206], *Ontology Engineering* [R41, R205], *Ontology Instance Generation* [R160, R208], and *Semantic Web Services* [R188]. The assignment to *VR/AR-assisted modeling* or *knowledge-based VR/AR* is not always clear. It depends on whether the semantic aspects are used for modeling ontology-driven VR/AR applications [R41], for semantic aspects such as reasoning for AR during run-time [R148], or for generating models by analyzing the

---

[1] For space reasons, references to articles from our literature analysis are linked in a separate document. Clicking on the [R] reference opens a web page showing the selected reference at the top.

sensor data of VR/AR devices. This last point seems to be missing so far in the found papers.

In *Software and System Visualization* the focus lies on *requirement gathering and analysis*, *designing*, *coding*, *testing*, and *maintenance and support*, i.e., on the software development life cycle [18]. Most of the discovered papers deal with analyzing [R58, R142, R155, R156 R157] (*knowledge-based VR/AR*) and designing [R105, R177] (*VR/AR-assisted modeling*) software and systems. Only few addressed testing and maintenance of software and systems [R9, R85] and none addressed so far the coding phase.

System Maintenance is an area where VR/AR is used in relation to maintenance activities, e.g., modeling languages and VR/AR systems guiding maintenance processes on the basis of conceptual models [R78, R99]. This refers mainly to the area of *knowledge-based VR/AR* as described at the beginning of the paper. Looking at the different types of maintenance, e.g., *improving*, *preventing*, and *correcting* [25], all types are covered by the found approaches, since most of them are not bound to a particular maintenance type.

In the *Training and Simulation* topic, contributions focus mainly on training and simulation aspects, e.g., in business process training. Mostly we can refer here again to *knowledge-based VR/AR for design- or run-time*. Most research is conducted in training applications involving virtual worlds for desktop applications [R8, R34, R121] followed by VR training environments [R182, R234]. Very little research has been done in the area of AR training applications combined with conceptual modeling [R75, R228]. This is an area that should be explored further, as training in AR offers many potential application scenarios.

The topic *User Aspects and Development Approaches* is twofold. First, contributions focusing on the user, i.e., *user interaction* [R57], *user interfaces* [R29], and *collaboration* [R215]. Second, research focusing on development approaches, i.e., approaches investigating *content authoring* [R42, R102], *model-driven development* [R30, R46], and the development of *virtual worlds* [R25]. Both of these main streams primarily cover design-time aspects, and thus, belong to *knowledge-based VR/AR*. Only very few contributions dealt with pedagogic or learning aspects [R132]. This is surprising as there is a lot of ongoing research on general VR/AR learning approaches, as recently shown by Chen et al. [10].

From the above descriptions and the mentioned papers, it becomes clear that most of the contributions found in our analysis are positioned in the area of *knowledge-based VR/AR* where models are used as input for VR/AR applications. Currently, there exist very few approaches where modeling in VR/AR, or the automated elicitation of models is considered. Further, only some contributions focus on pedagogic and learning aspects in AR modeling. Regarding missing areas, some aspects are not covered at all by research yet. For example, approaches combining knowledge-based VR/AR and VR/AR-assisted modeling, allowing the interplay of these two areas. Further, we could not yet identify approaches on the evaluation of business processes using VR/AR and no approaches for the semantic elicitation of conceptual models during run-time, e.g., for generating conceptual models on the basis of the user context.

Further, in comparison to the most promising industry use cases as proposed by the *Augmented Reality for Enterprise Alliance (AREA)* [1], which acts under the umbrella of the Object Management Group (OMG), 11 out of the 13 use case areas are covered also by our analysis. Only the areas *remote assistance* and *marketing and sales* did not become apparent in our study. This large overlap illustrates the relevance of the topics researched in academia for industry.

5   Related Work

Based on the wide research and analysis that we conducted, we can confidently state that to date, there has been no literature review that systematically investigates the combination of conceptual modeling with VR and AR. While there has been a previous review in the field conducted by Poehler and Teutenberg [30], it is important to note that their focus was specifically on the application of VR for business processes, rather than conceptual modeling as a whole. Thus, our findings highlight the novelty and importance of our review in filling this gap in the existing body of literature.

6   Conclusion

In this paper we conducted a systematic literature review, a computational bibliometric study, as well as an expert driven classification of papers combining conceptual modeling with VR/AR. The analysis suggests that there is a clear upward trend in the number of publications in this research area. There are no specific venues for this area so far, but the contributions are rather spread across many different outlets. The elaborated research areas include research in both VR/AR-assisted modeling, as well as knowledge-based VR/AR. However, the focus so far lies strongly on knowledge-based VR/AR. Only few publications deal with VR/AR-assisted modeling.

Despite the large number of publications that we reviewed, this study is however not without limitations. First, the initial selection of outlets for the literature search could have been extended to include further venues. However, since we performed a comprehensive forward- and a backward search for each paper, we are confident that we found most relevant papers. Second, we performed a computer-assisted content analysis using only unigrams. We did not consider bi-grams or n-grams, as this would have increased the complexity. This could be considered for an extension of the study in the future. Third, we only allowed papers to be allocated to one single topic. This follows the proposal of Vessey et al. [38]. However, this could be extended to multiple allocations, thereby permitting greater insight into the overlap of topics.

The results of our study offer valuable insights into the combination of conceptual modeling with virtual and augmented reality, which we believe will be of great interest to both the research community and industry practitioners. We hope that our findings will stimulate discussions and lead to further research in this evolving field. Moreover, we plan to standardize our process and share our insights with other members of the AR modeling community, which will help

to advance the field and drive future innovation, e.g., in AR-related enterprise modeling, like the use cases derived in [26].

Appendix A   Dataset of the Review Process

The bibliographies of the document corpora [R1-R248], as well as the various lists [T1-T8] documenting the whole process shown in Fig. 1 are available as HTML files online². In particular, we provide lists with the initial papers [T2], all papers [T3], papers per journal [T4], the most probable topics per paper [T5], as well as the assignments of the reviewers during the review process [T6, T7, T8, T9].

References

1. AREA: AREA AR use case descriptions. Tech. rep., Augmented Reality for Enterprises Alliance (AREA) (2022), https://thearea.org/wp-content/uploads/2022/05/AR-Use-Case-Descriptions-and-Examples.pdf, (Date accessed: 2022-11-07)
2. Betz, S., Eichhorn, D., Hickl, S., Klink, S., Koschmider, A., Li, Y., Oberweis, A., Trunko, R.: 3d representation of business process models. In: Modellierung betrieblicher Informationssysteme. LNI, vol. P-141, pp. 73–87. GI (2008), https://dl.gi.de/20.500.12116/23619
3. Blei, D.M.: Probabilistic topic models. Commun. ACM **55**(4), 77–84 (2012). https://doi.org/10.1145/2133806.2133826
4. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003)
5. Booth, A., Sutton, A., Papaioannou, D.: Systematic approaches to a successful literature review. Sage, second edition edn. (2016)
6. Boyd-Graber, J., Mimno, D., Newman, D.: Care and Feeding of Topic Models: Problems, Diagnostics, and Improvements. CRC Handbooks of Modern Statistical Methods, CRC Press (2014)
7. Brambilla, M., Cabot, J., Wimmer, M.: Model-Driven Software Engineering in Practice. Morgan & Claypool (2017). https://doi.org/10.2200/S00751ED2V01Y201701SWE004
8. Cabot, J., Vallecillo, A.: Modeling should be an independent scientific discipline. Softw. Syst. Model. **21**(6), 2101–2107 (2022). https://doi.org/10.1007/s10270-022-01035-8, https://doi.org/10.1007/s10270-022-01035-8
9. Chang, J., Boyd-Graber, J., Gerrish, S., Wang, C., Blei, D.M.: Reading tea leaves: How humans interpret topic models. In: NIPS'09 COnference. p. 288–296 (2009)
10. Chen, P., Liu, X., Cheng, W., Huang, R.: A review of using augmented reality in education from 2011 to 2016. In: Innovations in Smart Learning. pp. 13–18 (2017). https://doi.org/10.1007/978-981-10-2419-1_2
11. Churchill, R., Singh, L.: The evolution of topic modeling. ACM Comput. Surv. (2022). https://doi.org/10.1145/3507900
12. Cohen, J.: A coefficient of agreement for nominal scales. Educational and Psychological Measurement **20**(1), 37–46 (1960). https://doi.org/10.1177/001316446002000104

_____

² https://doi.org/10.5281/zenodo.7794278

13. Dalton, J., Gillham, J.: Seeing is believing (2019), https://www.pwc.com/gx/en/industries/technology/publications/economic-impact-of-vr-ar.html, (Date accessed: 2022-11-07)

14. Fill, H., Härer, F., Muff, F., Curty, S.: Towards augmented enterprise models as low-code interfaces to digital systems. In: Business Modeling and Software Design BMSD. vol. 422, pp. 343–352. Springer (2021). https://doi.org/10.1007/978-3-030-79976-2_22

15. Garcia-Castro, R., Gómez-Pérez, A., Muñoz-García, Ó., Nixon, L.J.B.: Towards a component-based framework for developing semantic web applications. In: ASWC Conference. pp. 197–211. Springer (2008). https://doi.org/10.1007/978-3-540-89704-0_14

16. Grambow, G., Hieber, D., Oberhauser, R., Pogolski, C.: A Context and Augmented Reality BPMN and BPMS Extension for Industrial Internet of Things Processes, pp. 379–390. Springer (2022). https://doi.org/10.1007/978-3-030-94343-1_29

17. Härer, F., Fill, H.: Past trends and future prospects in conceptual modeling - A bibliometric analysis. In: ER'2020. pp. 34–47. Springer (2020). https://doi.org/10.1007/978-3-030-62522-1_3

18. Khan, N.A.: Research on various software development lifecycle models. In: Proceedings of the Future Technologies Conference (FTC) 2020, Volume 3. pp. 357–364. Springer (2021). https://doi.org/10.1007/978-3-030-63092-8_24

19. Kitchenham, B.: Procedures for performing systematic reviews. Tech. rep. (2004)

20. Lafferty, J., Blei, D.: Correlated topic models. In: Advances in Neural Information Processing Systems. vol. 18, pp. 147–154. MIT Press (2005)

21. Landis, J.R., Koch, G.G.: The measurement of observer agreement for categorical data. Biometrics **33**(1), 159 (1977). https://doi.org/10.2307/2529310

22. Lau, J.H., Newman, D., Baldwin, T.: Machine reading tea leaves: Automatically evaluating topic coherence and topic model quality. In: Conference of the European Chapter of the Assoc. for Computational Linguistics. pp. 530–539. ACL (2014). https://doi.org/10.3115/v1/E14-1056

23. McCallum, A.K.: Mallet: A machine learning for language toolkit (2002), http://mallet.cs.umass.edu, (Date accessed: 2022-11-07)

24. Mimno, D., Wallach, H.M., Talley, E., Leenders, M., McCallum, A.: Optimizing semantic coherence in topic models. In: EMNLP '11. p. 262–272. ACL (2011)

25. Mobley, R.: Maintenance Fundamentals. Plant Engineering, Elsevier Science (2011). https://doi.org/10.1016/b978-0-7506-7798-1.x5021-3

26. Muff, F., Fill, H.: Use cases for augmented reality applications in enterprise modeling: A morphological analysis. In: BMSD Symposium. pp. 230–239. Springer (2022). https://doi.org/10.1007/978-3-031-11510-3_14

27. Muff, F., Härer, F., Fill, H.: Trends in academic and industrial research on business process management - A computational literature analysis. In: HICSS. pp. 1–10. ScholarSpace (2022), http://hdl.handle.net/10125/80215

28. Mylopoulos, J.: Conceptual modelling and Telos. In: Conceptual Modeling, Databases, and Case: An Integrated View of Information Systems Development. pp. 49–68. John Wiley & Sons, Inc. (1992)

29. Newman, D., Asuncion, A.U., Smyth, P., Welling, M.: Distributed algorithms for topic models. J. Mach. Learn. Res. **10**, 1801–1828 (2009). https://doi.org/10.5555/1577069.1755845

30. Pöhler, L., Teuteberg, F.: Closing spatial und motivational gaps: Virtual reality in business process improvement. In: European Conference on Information Systems, AIS (2021), https://aisel.aisnet.org/ecis2021_rp/151

31. PricewaterhouseCoopers: Us metaverse survey: Build a metaverse strategy to deliver sustainable business outcomes (2022), https://www.pwc.com/us/en/tech-effect/emerging-tech/metaverse-survey.html, (Date accessed: 2022-11-07)

32. Qiang, J., Chen, P., Wang, T., Wu, X.: Topic modeling over short texts by incorporating word embeddings. In: PAKDD. vol. 10235, pp. 363–374 (2017). https://doi.org/10.1007/978-3-319-57529-2_29

33. Quan, X., Kit, C., Ge, Y., Pan, S.J.: Short and sparse text topic modeling via self-aggregation. In: Joint Conference on Artificial Intelligence. pp. 2270–2276. AAAI (2015), http://ijcai.org/Abstract/15/321

34. Roberts, M.E., Stewart, B.M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S.K., Albertson, B., Rand, D.G.: Structural topic models for open-ended survey responses. American Journal of Political Science **58**(4), 1064–1082 (2014). https://doi.org/https://doi.org/10.1111/ajps.12103

35. Sandkuhl, K., Fill, H., Hoppenbrouwers, S., Krogstie, J., Matthes, F., Opdahl, A.L., Schwabe, G., Uludag, Ö., Winter, R.: From expert discipline to common practice: A vision and research agenda for extending the reach of enterprise modeling. Bus. Inf. Syst. Eng. **60**(1), 69–80 (2018). https://doi.org/10.1007/s12599-017-0516-y

36. Shahnaz, F., Berry, M.W., Pauca, V.P., Plemmons, R.J.: Document clustering using nonnegative matrix factorization. Inf. Process. Manag. **42**(2), 373–386 (2006). https://doi.org/10.1016/j.ipm.2004.11.005

37. de Souza Cardoso, L.F., Mariano, F.C.M.Q., Zorzal, E.R.: A survey of industrial augmented reality. Comput. Ind. Eng. **139** (2020). https://doi.org/10.1016/j.cie.2019.106159

38. Vessey, I., Ramesh, V., Glass, R.L.: Research in information systems: An empirical study of diversity in the discipline and its journals. J. Manag. Inf. Syst. **19**(2), 129–174 (2002). https://doi.org/10.1080/07421222.2002.11045721

39. Webster, J., Watson, R.T.: Analyzing the past to prepare for the future: Writing a literature review. MIS Q. **26**(2) (2002)

40. Weske, M.: Business Process Management - Concepts, Languages, Architectures, Third Edition. Springer (2019). https://doi.org/10.1007/978-3-662-59432-2

41. Wild, F., Perey, C., Hensen, B., Klamma, R.: IEEE standard for augmented reality learning experience models. In: IEEE International Conference on Teaching, Assessment, and Learning for Engineering, TALE 2020. pp. 1–3. IEEE (2020). https://doi.org/10.1109/TALE48869.2020.9368405

42. Yin, K., He, Z., Xiong, J., Zou, J., Li, K., Wu, S.T.: Virtual reality and augmented reality displays: advances and future perspectives. Journal of Physics: Photonics **3**(2) (2021). https://doi.org/10.1088/2515-7647/abf02e