Past Trends and Future Prospects in Conceptual Modeling - A Bibliometric Analysis

Felix Härer[1[0000−0002−2768−2342]] and Hans-Georg Fill[1[0000−0001−5076−5341]]

Digitalization and Information Systems Group,
University of Fribourg, 1700 Fribourg, Switzerland
felix.haerer|hans-georg.fill@unifr.ch
https://www.unifr.ch/inf/digits/en/

Abstract. Research in conceptual modeling is today undertaken in a large number of fields. We describe the results from a bibliometric analysis of major outlets for conceptual modeling research for investigating the evolution of research topics. As a basis for the study we used the openly accessible DBLP dataset which we enriched with data from related publisher's websites and databases. Besides a descriptive analysis of the publication data, we conducted a content-based analysis of more than 3.200 papers using Latent Dirichlet Allocation. This permits to gain insights into the past trends in conceptual modeling research and derive future prospects for the community.

Keywords: Conceptual Modeling · Research Communities · Bibliometric Analysis.

1 Introduction

The benefits of conceptual modeling for digital transformation as implied by this year's conference theme seem to be largely undisputed within the conceptual modeling community itself [2,23,12,14,3]. However, from the viewpoint of other disciplines, practitioners and newcomers it is often not so obvious what conceptual modeling stands for, which topics are investigated and where according research takes place, cf. [8]. And even for a scientific community itself it is from times to times beneficial to reflect on its past achievements, on the evolution of topics and the identification of visions and future directions [21].

The initial impulse for our investigation was a statement by an anonymous reviewer in an evaluation report, who stated that some geographical regions would dominate certain areas in conceptual modeling. Whereas any member of the community might have some gut feelings of whether this may be true or not, we could not find previous analyses that would have permitted us to verify or counter this statement. Some indication exists that there is a dependency between organizational and educational cultures and the use of certain diagrams [13]. To the best of our knowledge, the only bibliometric analysis available so far is the paper by Chen et al. who had looked at papers from the ER conference from 1979-2005 with a focus on authors and citation counts [7]. As conceptual modeling is however today spread across many different outlets, we decided to conduct a more comprehensive analysis. The research questions we defined for this undertaking were as follows:

- RQ 1: In which geographical regions is conceptual modeling research conducted?
- RQ 2: How has the quantity of conceptual modeling papers evolved over time, taking into account specific regions and the communities of specific outlets?
- RQ 3: How many authors are active in conceptual modeling and what is their typical number of papers?
- RQ 4: What are the major conceptual modeling topics in terms of published research, how do they differ per outlet and how did they evolve over time?
- RQ 5: With the topical evolution, where is conceptual modeling positioned now and are there indications of research gaps or opportunities?
- RQ 6: Is there an indication of prospective topics, application areas, or domains where conceptual modeling might be used in the future?

For answering these questions we will report in the following results from a bibliometric analysis that we have conducted based on the openly accessible DBLP dataset as the main source. Due to several manual steps that were required for the analysis, we restricted the outlets to nine core outlets in conceptual modeling: the International Symposium on Business Modeling and Software Design (BMSD), the joint publication of Business Process Modeling, Development and Support in conjunction with Evaluation and Modelling Methods for Systems Analysis and Development (BPMDS/EMMSAD), the International Conference on Conceptual Modeling (ER), the International Conference on Model Driven Engineering Languages and Systems (MoDELS), and the Practice of Enterprise Modeling (PoEM) conference as well as the Complex Systems Informatics and Modeling Quarterly (CISMQ), the Enterprise Modelling and Information Systems Architectures - International Journal of Conceptual Modeling (EMISAJ), the International Journal of Information System Modeling and Design (IJISMD), and the Software and Systems Modeling (SoSyM) journal. We further enriched the data of the time span from 2005-2019 with information from publishers' websites for adding further author attributes, e.g. for determining their geographical origin and institution. In addition, we retrieved the full-texts of approx. 3.200 papers to conduct content-based analyses.

The remainder of the paper is structured as follows: in Section 2 we describe the data collection and analysis process in detail, followed by a descriptive analysis in Section 3 and a content-based analysis in Section 4. The results are discussed in Section 5 and conclusions are drawn in Section 6.

2 Data Collection and Analysis Process

For the data collection and analysis process we reverted to the well-known KDD process commonly used in data mining [10]. It contains the steps *data selec-*

tion, *preprocessing*, *transformation*, and *data mining* and applies the Extract-Transform-Load (ETL) procedure [1]. The data collection and analysis process is outlined in Figure 1.

2.1 Data Collection Process

By using the DBLP database dump from 2019-10-11[1] as a starting point, nine major English-language conference and journal outlets for conceptual modeling were selected. For each of the 4.131 entries matching the relevant identifiers of the outlets, metadata in the form of *Title*, *Authors*, *Year*, *Outlet*, *URL*, and *DOI* in most cases, was collected in a JSON file. For further enrichment, data was retrieved from the publishers' websites, e.g. IEEE, and included metadata on DOI, affiliation, and country, with additional manual extractions of countries from the affiliation in case of known and exact matches.

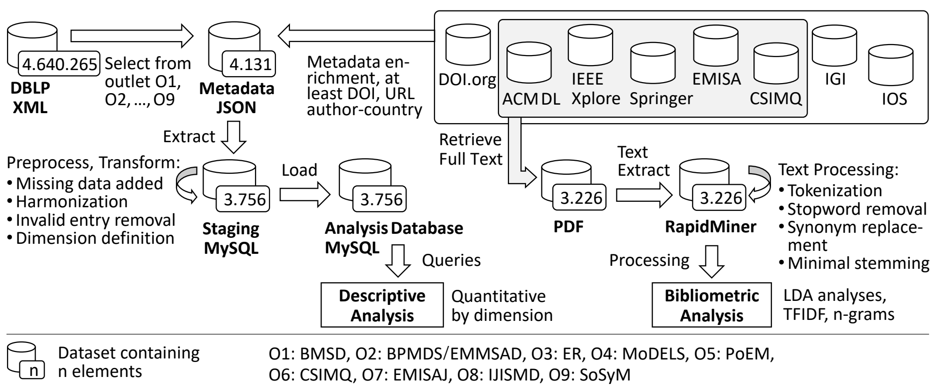

Fig. 1. Data collection and analysis process based on the DBLP XML dataset

Metadata was extracted into a staging database while cleaning operations were applied regarding the addition of missing data for DOI and authors, the manual harmonization of all names, countries, and outlets, as well as the removal of invalid entries including non-paper articles such as editorials and placeholders with missing authors. For an analysis along multiple dimensions, the 3.756 remaining entries were transformed into the star schema shown in Figure 2 with partially normalized dimensions or "snowflaking". With this standard practice in data warehousing, redundancy is accepted for never-changing historized data to minimize query latency [15, p. 55]. Finally, we loaded 3.756 entries for analysis[2].

Full text documents were available to us from all publishers except IGI and IOS. The 3.226 available full texts of the publishers indicated in Figure 1 were retrieved using Node.js scraping scripts. Subsequently, the documents were loaded

[1] https://dblp.uni-trier.de/xml/
[2] For all publications see https://zenodo.org/record/3982628.

into RapidMiner Studio 9.5. There, NLP operations were applied such as tokenization, stopword removal, synonym replacement and some minimal stemming for the normalization of plural and common inflected forms[3].

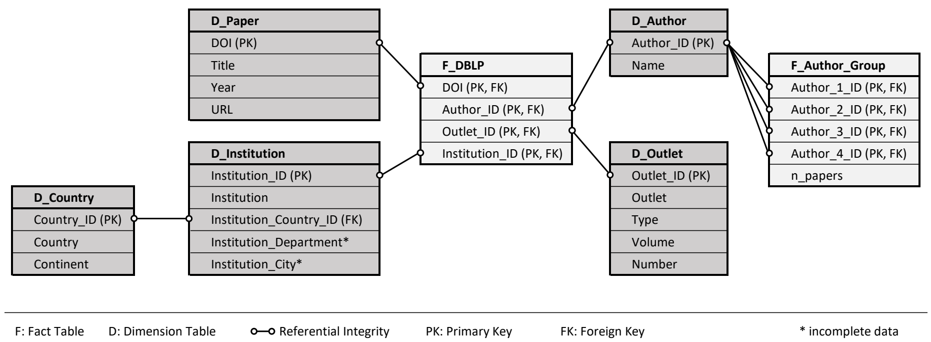

Fig. 2. Schema of the analysis database. Fact tables (prefix F) for the storage of DBLP publications and author groups reference according dimension tables (prefix D).

2.2 Data Analysis Process

The analysis database served as a source for descriptive statistics. By using queries over the dimensions, the quantitative analysis yielded exact answers such as the absolute frequencies of publications. The database architecture proved to be useful for multi-dimensional queries, e.g. authors affiliated with institutions from countries at varying levels of granularity such as specific countries, continents or all countries. A second analysis of bibliographic data was based on the full texts of the documents. Latent Dirichlet Allocation (LDA) was applied, which is a statistical method for the identification of topics in documents [5,4]. Further operations for TF-IDF, n-gram occurrences, and clustering were deferred to the future analysis of the dataset and will not be regarded here.

3 Descriptive Analysis of the Dataset

For a first descriptive analysis, Table 1 outlines the properties of the dataset, showing the contained outlets with publications, author figures and availability. For the following analysis, the time period 2005 to 2019 is considered and grouped into time frames where appropriate. The interpretation of these results will be discussed in Section 5.

[3]The NLP and analysis processes for RapidMiner as well as stopwords and synonyms are available at https://zenodo.org/record/3982628.

For investigating the worldwide distribution of conceptual modeling research as described in RQ1, we conducted a quantitative analysis by querying the affiliations of each paper with their known geographical locations. As shown in Table 2, 2.466 or 66% of publications originated from Europe, 603 or 16% from North America, and 10% or 359 from Asia.

Outlet	Type	n papers	n authors	y_1	y_2
BMSD	C	119	214	2011	2019
BPMDS/EMMSAD	C	320	651	2009	2019
ER	C	1155	2115	1992	2019
MoDELS	C	712	1388	2005	2018
PoEM	C	230	442	2008	2018
CSIMQ	J	112	262	2014	2019
EMISAJ	J	131	303	2005	2018
IJISMD	J	161	411	2010	2019
SoSyM	J	816	1703	2002	2019
Σ		3756			

Table 1. Dataset overview indicating for each outlet (C: Conference, J: Journal) the number of papers and authors with the first year y_1 and last year y_2 of available data.

Concerning the evolution of the number of publications over time (RQ2), it can be stated that from the total of 3.199 publications, most publications fall into the period 2015-2019, i.e. 1.347 (42%), compared to 1.101 (34%) in the timeframe 2010-2014 and 751 (23%) in 2005-2009 as shown in Table 2 per continent and in Table 3[4] per outlet. Considering specific outlets, most publications in the dataset are from SoSyM, i.e. 762 papers, followed by 712 papers in MoDELS and 652 papers in ER.

In RQ3 we considered the number of authors who are active in conceptual modeling and their typical number of papers. In our dataset we identified $n = 5.141$ authors, indicated per outlet in Table 1. For the typical number of publications, a skew has been observed in that 69% of authors have exactly one publication while 31% have between 2 and 61 publications. Considering percentile measures, 80% of the observed authors have less than $P_{80} = 2$ publications. Furthermore the progression $P_{85} = 3$, $P_{90} = 4$, $P_{95} = 6$ indicates authors with several publications. The indication is a power law distribution, sometimes observed in social networks [6], as the steep increase to $P_{96} = 6$,

[4]Note that countries located on multiple continents cause the total sum of publications in Table 2 to exceed the one of Table 3.

Continent	Year			Σ
	2005 - 2009	2010 - 2014	2015 - 2019	
Africa	3	17	26	46
Asia	58	104	197	359
Europe	532	882	1052	2466
North America	202	199	202	603
Oceania	37	44	46	127
South America	26	45	75	146

Table 2. Publications originating from authors affiliated with institutions in the given continents over time frames of five years.

$P_{97} = 7$, $P_{98} = 10$, $P_{99} = 14$ shows. It can be concluded that most of the interactions or publications are from a relatively small community of authors. One first interpretation is that many of the 5.141 authors might not be from the core conceptual modeling community but rather from collaborating fields.

4 Content-based Analysis of Publications

In a second step, we analyzed the contents of the papers contained in our dataset. For this we used the LDA implementation of MALLET (MAchine Learning for LanguagE Toolkit[5]) that is part of RapidMiner 9.5[6]. In comparison to simpler methods such as term frequency and n-gram analysis, LDA is a topic modeling approach operating on the level of documents with the goal of identifying their topics. Thereby, LDA assumes that several topics are present in each document. Given a document collection, any document d is described by a statistical distribution θ_d over its topics. That is, each topic has a particular probability or weight for d, and a distribution of words $\theta_{d,k}$ for any topic k [4]. Each topic k therefore is represented by the top n words according to their probability or weight. The particular weights can be considered hidden variables, determined with the Gibbs sampling scheme, which is carried out iteratively for each word such that its likelihood of appearing in a particular topic is maximized [20].

For all LDA analyses that we performed, we show the top eight topics inferred from the top five words according to their weight (cf. [4,22]). In addition, the topics are ordered by the cumulative weight of the top five words, where a weight of topic k and word w is an absolute measure of the occurrences for w assigned

[5] http://mallet.cs.umass.edu/topics.php

[6] Specifically, a concurrent implementation for the detection of topics [24] was used in combination with a specialized sampling scheme based on Gibbs sampling [20].

Outlet	Year			Σ
	2005 - 2009	2010 - 2014	2015 - 2019	
BMSD	0	30	89	119
BPMDS/EMMSAD	33	161	126	320
ER	214	217	221	652
MoDELS	286	248	178	712
PoEM	36	92	102	230
CSIMQ	0	4	108	112
EMISAJ	30	45	56	131
IJISMD	0	82	79	161
SoSyM	152	222	388	762
Σ	751	1101	1347	3199

Table 3. Publications of individual outlets over time frames of five years.

to k in all documents[7]. The LDA analysis of full-text documents between 2005 and 2019 was performed for the top eight topics in the whole data set, in each outlet, over each continent and over consecutive time-frames of five years. In the following we will only discuss a subset of the results due to limitations of space.

The results for the LDA analyses are shown in the following tables, with Table 4 demonstrating overall results and Table 5 an analysis of an outlet dimension by using the ER conference as example. We can observe for the topics

Topic 1		Topic 2		Topic 3		Topic 4	
Word	Weight	Word	Weight	Word	Weight	Word	Weight
process	83634	model	84526	model	59466	model	33915
model	56049	language	29097	transformation	41565	system	28213
business	47149	tool	18822	rule	27674	test	17407
service	30829	metamodel	16466	graph	18114	software	17176
system	23810	uml	16429	element	13825	feature	15532

Topic 5		Topic 6		Topic 7		Topic 8	
Word	Weight	Word	Weight	Word	Weight	Word	Weight
model	58638	state	32081	class	21079	data	28064
use	14060	event	24238	constraint	21019	schema	10742
case	12277	model	16437	model	19568	database	10260
software	12206	transition	13856	type	14748	query	8659
tool	10573	system	13477	object	13784	set	8190

Table 4. LDA analysis of papers from 2005 - 2019 ordered by cumulative topic weight

[7]Note that the ordering is solely for presentation as it can only consider known weights. Weights beyond the top words possibly involve all words of all documents.

Topic 1		Topic 2		Topic 3		Topic 4	
Word	**Weight**	**Word**	**Weight**	**Word**	**Weight**	**Word**	**Weight**
model	16398	schema	5967	process	7853	goal	6068
conceptual	4162	data	4344	model	3571	model	4222
system	3748	node	3423	service	3017	requirement	3058
design	3207	database	3215	data	2993	value	2085
information	2864	query	3200	event	2256	system	2082

Topic 5		Topic 6		Topic 7		Topic 8	
Word	**Weight**	**Word**	**Weight**	**Word**	**Weight**	**Word**	**Weight**
model	3820	ontology	4668	data	5482	set	2987
type	3530	model	2782	user	2713	constraint	2115
class	3253	type	2356	concept	1718	relation	1890
relationship	2962	concept	2018	web	1714	data	1601
object	2406	relation	1840	result	1702	tuple	1498

Table 5. LDA analysis of papers published at the ER conference between 2005 and 2019 ordered by cumulative topic weight

of the ER conference that the word *model* occurs in all topics except for Topics 2, 7, and 8. In Topic 2 *schema* is obviously prominent. Most weight is accumulated by Topic 1 with the highest-weighted words *model* and *conceptual*, *system*, *design*, and *information*. This is followed by Topic 2 in the context of schemas and databases, Topic 3 in the context of processes and services, and Topic 4 in the context of goals and requirements[8]. The remaining topics cover additional themes typically considered for the ER conference, probably with the exception of Topic 7 that leaves room for interpretation.

Topic 1		Topic 2		Topic 3		Topic 4	
Word	**Weight**	**Word**	**Weight**	**Word**	**Weight**	**Word**	**Weight**
model	70196	model	48435	model	47688	process	60317
language	27875	business	24000	transformation	35420	model	28576
class	16830	process	21852	rule	22523	business	15904
uml	14970	system	19973	graph	15231	event	9714
metamodel	14570	information	18395	element	11086	activity	9536

Topic 5		Topic 6		Topic 7		Topic 8	
Word	**Weight**	**Word**	**Weight**	**Word**	**Weight**	**Word**	**Weight**
model	41349	system	22905	state	24032	data	18289
test	15518	model	22082	set	11431	schema	9311
software	14148	service	14566	system	10664	set	7180
feature	13442	component	13142	event	10508	database	7134
case	11564	time	12182	model	9555	ontology	6331

Table 6. LDA analysis of papers by authors affiliated with institutions in Europe between 2005 and 2019 ordered by cumulative topic weight

[8]Note the absence of the term *ER* that is part of the stop word list described in section 2.1 due to frequent mentions of the conference name.

Excerpts from the LDA analysis results per continent are shown in Table 6 for Europe and in Table 7 for North America. The topic with the most cumulative weight for publications with authors from European institutions contains the terms *model* and *language*, which are also found in the first topic for North America. There, *tool* and *software* seem to be more prominent than *language*, whereas *UML* and *metamodel* are prominently observed in Europe. A further observation is that the terms contained in Topic 8 for Europe around databases do not occur in a similar way in any of the top topics for North America. Conversely, topics related to *goal* and *requirement* are absent in the top European topics. Topics around business processes are strongly represented in Europe through Topic 2 and Topic 4 and only considered in Topic 5 in North America.

Topic 1		Topic 2		Topic 3		Topic 4	
Word	**Weight**	**Word**	**Weight**	**Word**	**Weight**	**Word**	**Weight**
model	19129	transformation	9864	state	5663	model	10141
tool	5872	model	9814	model	5598	feature	4989
software	4951	rule	6553	system	4567	product	2472
language	3641	metamodel	3604	time	3721	element	2138
system	3171	graph	3444	event	2798	change	1959

Topic 5		Topic 6		Topic 7		Topic 8	
Word	**Weight**	**Word**	**Weight**	**Word**	**Weight**	**Word**	**Weight**
process	7551	model	4869	goal	5776	data	3806
business	4085	system	4017	model	5648	type	2195
service	3318	test	3448	requirement	3012	attribute	2142
model	2395	uml	3110	system	1797	conceptual	2102
system	2359	case	2799	analysis	1656	relationship	2011

Table 7. LDA analysis of papers by authors affiliated with institutions in North America between 2005 and 2019 ordered by cumulative topic weight

In Table 8 and Table 9 the results from the LDA analysis of two time frames are shown. It can be observed that the predominant topic in the earlier time frame (2010-2014) includes the terms *model*, *transformation*, *rule*, *element*, and *graph*. In the more recent time frame (2015-2019), the predominant topic contains the terms *model*, *transformation*, *language*, *element*, and *metamodel*.

5 Discussion

The discussion of the results of our analysis will be divided into *a*. the discussion of the descriptive analysis results of the retrieved data and *b*. the discussion of the content-based analysis results through the application of LDA to the contents of the retrieved papers.

The development of the overall numbers of publications as shown in Table 3 shows the positive development of the conceptual modeling community in general. With regard to the three time periods 2005-2009, 2010-2014, and 2015-2019, most outlets show increasing numbers of publications. The only exceptions being

Topic 1		Topic 2		Topic 3		Topic 4	
Word	**Weight**	**Word**	**Weight**	**Word**	**Weight**	**Word**	**Weight**
model	23689	model	27849	process	26140	model	24810
transformation	14645	software	8282	model	10824	process	5977
rule	9910	tool	7229	business	7740	system	5492
element	5783	language	7227	activity	5153	design	5129
graph	5688	feature	6932	event	3965	information	4766

Topic 5		Topic 6		Topic 7		Topic 8	
Word	**Weight**	**Word**	**Weight**	**Word**	**Weight**	**Word**	**Weight**
service	12192	model	9312	constraint	7639	data	8192
business	9311	system	7684	model	5512	schema	4331
model	7703	state	6073	semantic	5150	query	3618
goal	6756	test	5423	class	5065	database	3028
system	5717	time	5055	set	4477	user	2886

Table 8. LDA of papers in 2010 - 2014 ordered by cumulative topic weight

here the MoDELS conference that shows a decline of approx. -13% and -28%, the BPMDS/EMMSAD conference with a decline of -22% in the last period and the IJISMD journal with a slight decline of -4%.

A closer inspection for the MoDELS conference revealed that this decline is primarily not due to a decrease in the acceptance rate but rather due to lower submission numbers, i.e. for example 172 submissions and 35 accepted full papers in the foundations track in 2015 compared to 89 submissions and 18 full papers in 2019 [19,25].

Topic 1		Topic 2		Topic 3		Topic 4	
Word	**Weight**	**Word**	**Weight**	**Word**	**Weight**	**Word**	**Weight**
model	43627	process	39247	model	20638	model	25545
transformation	20563	model	20585	system	12527	software	9678
language	11544	business	10396	business	11773	feature	9389
element	10093	event	5792	information	10395	case	8671
metamodel	8930	task	5171	service	9866	use	8267

Topic 5		Topic 6		Topic 7		Topic 8	
Word	**Weight**	**Word**	**Weight**	**Word**	**Weight**	**Word**	**Weight**
model	19817	model	19027	state	9975	data	13071
system	15845	type	7502	set	8150	database	4640
component	6957	class	7203	model	6258	time	4554
time	6507	level	6502	constraint	5998	set	4125
state	6377	instance	5349	rule	5231	value	4093

Table 9. LDA of papers in 2015 - 2019 ordered by cumulative topic weight

At the same time, the MoDELS conference hosted 18 workshops in 2019, which may outweigh the decline in submissions to the main conference and which are not part of our analysis. The reasons for such a shift would need to be investigated more closely, e.g. through expert interviews with authors and organizers.

One reason could be the generally higher acceptance rate of workshops and the increasing pressure to publish results that has been reported for many fields [18], which could make submissions to workshops more attractive.

Overall, the top three outlets in terms of publication numbers in 2005-2019 are the SoSyM journal, the MoDELS conference, and the ER conference. Several smaller outlets such as BMSD, PoEM, CSIMQ, and EMISAJ started to make good progress. It has to be noted that some outlets were not yet present in all time periods, i.e. BMSD, CSIMQ, and IJISMD began their activities only after 2009.

When analyzing the origins of the authors of the respective papers across geographical regions, the strong predominance of researchers with a European affiliation stands out. Whereas the number of contributions from authors in North America stayed relatively constant, conceptual modeling research is clearly on the rise in the last two time periods (2010-2014 / 2015-2019) in Europe (+19% / +134%), Asia (+89% / +82%), and South America (+67% / +95%). Though the overall increase in publications can be witnessed across disciplines[9], the relative increases per region should be considered by editors and conference organizers.

In terms of the content-based analysis using LDA, the results for the whole data set over all periods show the broad range of topics that have been investigated by the conceptual modeling community in the past – see Table 4. We can make two major observations here: first, several topics are related to technical, fundamental aspects of modeling and schemas, e.g. in terms of metamodels, languages, transformation, constraints, and schemas (Topics 2, 3, 7, 8), which one would expect in such a research community. Second however, the application domains seem to be centered so far mostly around business / business process models, software and data models, with process and business modeling topics on the forefront. Other domains such as the humanities, e.g. [17], the legal domain, e.g. [11] or natural and exact sciences, e.g. [9], have not yet gained high visibility - at least in the investigated outlets. It could be an opportunity for the future orientation of the conceptual modeling community to broaden its scope to further domains and thus increase its relevance and impact.

The results of the LDA analysis per outlet permit to assess the scope of topics that have been primarily published in a particular publication source and also give an estimation of the importance of topics within this source. The topics in Table 5 nicely show the primary topics for the ER conference in the past. Although we were not able to integrate the results from the analysis of other outlets due to space restrictions, we can conclude that such insights can serve well in guiding potential authors in choosing an outlet for publishing their research and they support our understanding based on our own experience with these outlets. The results may be further used to assess the strategies of the different outlets, e.g. of journals publishing conceptual modeling research to decide which audiences they want to address and for conference organizers to decide about the tracks and workshops.

[9]See the large increases in the number of publications recorded in DBLP https://dblp.uni-trier.de/statistics/newrecordsperyear.

The LDA analysis per continents as shown for Europe and North America revealed that process modeling and business information systems are discussed to a higher extent in Europe than in North America as shown by Topics 2 and 4. As we had already noted in the descriptive analysis, goal and requirements modeling seems to have a stronger standing in North America than in Europe, whereas a topic around databases and data schemas is only present among the first eight topics in Europe but not in North America.

Finally, the LDA analysis per time periods as shown here for 2010-2014 and 2015-2019 can aid in tracking the importance of topics over time. Based on the cumulative weights of the topics and the resulting ordering, it can be derived that model transformation has been of constant interest and less papers have recently focused on software models while more papers have been published on process and business information systems modeling. Interpretations for these shifts have to be made carefully as we did not investigate the reasons for these shifts. Together with the results shown for the distribution of papers across outlets and the topics discussed for each outlet, one interpretation would be that this is linked to the decrease of papers in our dataset for the MoDELS conference, which is strongly positioned in software modeling and UML.

There are several limitations of this study that need to be noted. This concerns foremost the selection of only a subset of outlets for publishing research on conceptual modeling. Due to the highly inter-disciplinary nature of the field, results of conceptual modeling research are often published in other, domain-oriented outlets apart from the traditional modeling outlets. Furthermore, the investigated conferences typically host a large number of workshops that also publish high volumes of papers, which have not been considered. This also applies to conferences and journals that are not published in English - e.g. the German Modellierung conference[10] - that have a long tradition of conceptual modeling research and have neither been included. This may change some results of the analysis. On the other hand, we believe that we have based our analysis on a good sample of conferences and journals that are relevant for the community.

6 Conclusion and Future Prospects

In summary we can draw the following conclusions: conceptual modeling research is well-established and shows a positive development in terms of the number of publications as well as the number of outlets available for presenting results. Despite these good news, it should be considered to widen the scope of traditional conceptual modeling outlets or to create new outlets for investigating novel applications of conceptual modeling to further domains. Whereas topics related to conceptual modeling and databases, business and information systems have been well covered, other domains seem to be underrepresented so far. Although workshops, which are traditionally held in conjunction with the major conferences, may serve this purpose, many of them only take up specialized topics in

[10]See https://dblp.uni-trier.de/db/conf/modellierung/

the traditional domains. With these results, our next steps encompass the design of tests on available and ongoing data to complete the cycle of hypothesis testing and hypothesis generation [16] for informing future analyses.

Based on the insights we have gained during our analysis, we see future prospects for conceptual modeling in a multitude of new domains, which would increase the overall relevance and impact. Potential candidates for such domains are the humanities, the legal domain or natural sciences. First indications for such directions are the 1st International Workshop on Conceptual Modeling for Life Sciences initiated by Bernasconi, Canakoglu, Palacio, and Román, which is hosted at this year's ER conference[11] as well as the ongoing workshops on Characterizing the Field of Conceptual Modeling initiated by Delcambre, Pastor, Liddle, and Storey[12].

References

1. Ali, S.M.F., Wrembel, R.: From conceptual design to performance optimization of ETL workflows: Current state of research and open problems. The VLDB Journal **26**(6), 777–801 (Dec 2017)
2. Babar, Z., Yu, E.S.K.: Digital transformation - implications for enterprise modeling and analysis. In: 23rd IEEE International Enterprise Distributed Object Computing Workshop, EDOC Workshops 2019, Paris, France, October 28-31, 2019. pp. 1–8. IEEE (2019). https://doi.org/10.1109/EDOCW.2019.00015
3. Berzisa, S., Bravos, G., González, T.C., Czubayko, U., España, S., Grabis, J., Henkel, M., Jokste, L., Kampars, J., Koç, H., Kuhr, J., Llorca, C., Loucopoulos, P., Pascual, R.J., Pastor, O., Sandkuhl, K., Simic, H., Stirna, J., Valverde, F.G., Zdravkovic, J.: Capability driven development: An approach to designing digital enterprises. Bus. Inf. Syst. Eng. **57**(1), 15–25 (2015). https://doi.org/10.1007/s12599-014-0362-0
4. Blei, D.M.: Probabilistic topic models. Communications of the ACM **55**(4), 77–84 (2012). https://doi.org/10.1145/2133806.2133826
5. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. Journal of Machine Learning Research **3**, 993–1022 (2003)
6. Broido, A.D., Clauset, A.: Scale-free networks are rare. Nature Communications **10**(1), 1–10 (Mar 2019). https://doi.org/10.1038/s41467-019-08746-5
7. Chen, C., Song, I.Y., Zhu, W.: Trends in conceptual modeling: Citation analysis of the er conference papers (1979-2005). In: Proceedings of the 11th International Conference on the International Society for Scientometrics and Informatrics. pp. 189–200. CSIC (2007)
8. Delcambre, L.M.L., Liddle, S.W., Pastor, O., Storey, V.C.: A reference framework for conceptual modeling. In: Trujillo, J.C., Davis, K.C., Du, X., Li, Z., Ling, T.W., Li, G., Lee, M.L. (eds.) Conceptual Modeling. pp. 27–42. Springer (2018)
9. Döller, V.: Provis - probability visualized: A modeling tool for teaching stochastics. In: Companion Proceedings of Modellierung 2020 Short, Workshop and Tools & Demo Papers. CEUR Workshop Proceedings, vol. 2542, pp. 222–226. CEUR-WS.org (2020)

[11] http://www.bioinformatics.deib.polimi.it/cmls/
[12] http://www.nwpu-bioinformatics.com/ER2018/file/ccm.pdf

10. Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P.: From data mining to knowledge discovery: An overview. In: Advances in Knowledge Discovery and Data Mining, pp. 1–34. American Association for Artificial Intelligence, USA (Feb 1996)
11. Fill, H.G.: Towards requirements for a meta modeling formalism to support visual law representations. In: Schweighofer, E., Kummer, F., Hoetzendorfer, W. (eds.) Internationales Rechtsinformatik Symposium 2012. Salzburg (2012)
12. Gray, J., Rumpe, B.: Models for the digital transformation. Software and Systems Modeling **16**(2), 307–308 (2017). https://doi.org/10.1007/s10270-017-0596-7
13. Jaakkola, H., Thalheim, B.: Culture-Adaptable Web Information Systems. In: Information Modelling and Knowledge Bases XXVII, Frontiers in Artificial Intelligence and Applications, vol. 280 (2016)
14. Karagiannis, D., Mayr, H.C., Mylopoulos, J.: Domain-specific Conceptual Modeling. Springer (2016)
15. Kimball, R., Ross, M.: The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. Wiley Publishing, 3rd edn. (2013)
16. Knobloch, B.: A Framework for Organizational Data Analysis and Organizational Data Mining. In: Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications: Concepts, Methodologies, Tools, and Applications, pp. 449–462. IGI Global (2008)
17. Kropp, Y.O., Thalheim, B.: Conceptual modelling and humanities. In: Michael, J., Bork, D., Fill, H., Fettke, P., Karagiannis, D., Köpke, J., Koschmider, A., Mayr, H.C., Rehse, J., Reimer, U., Striewe, M., Tropmann-Frick, M., Ullrich, M. (eds.) Companion Proceedings of Modellierung 2020 Short, Workshop and Tools & Demo Papers co-located with Modellierung 2020, Vienna, Austria, February 19-21, 2020. CEUR Workshop Proceedings, vol. 2542, pp. 13–21. CEUR-WS.org (2020)
18. Kun, Á.: Publish and who should perish: You or science? Publications **6**(2), 18 (2018). https://doi.org/10.3390/publications6020018
19. Lethbridge, T., Cabot, J., Egyed, A.: Message from the chairs. In: 2015 ACM/IEEE 18th International Conference on Model Driven Engineering Languages and Systems (MODELS). pp. iii–iii (2015)
20. Newman, D., Asuncion, A., Smyth, P., Welling, M.: Distributed algorithms for topic models. Journal of Machine Learning Research **10**, 1801–1828 (2009)
21. Pastor, O.: Conceptual modeling of life: Beyond the homo sapiens. In: Comyn-Wattiau, I., Tanaka, K., Song, I., Yamamoto, S., Saeki, M. (eds.) Conceptual Modeling - 35th International Conference, ER 2016, Gifu, Japan, November 14-17, 2016, Proceedings. Lecture Notes in Computer Science, vol. 9974, pp. 18–31 (2016)
22. Rosen-Zvi, M., Griffiths, T.L., Steyvers, M., Smyth, P.: The author-topic model for authors and documents. In: UAI '04, Proceedings of the 20th Conference in Uncertainty in Artificial Intelligence, Banff, Canada, July 7-11, 2004. pp. 487–494 (2004)
23. 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)
24. Yao, L., Mimno, D., McCallum, A.: Efficient methods for topic model inference on streaming document collections. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 937–946. KDD '09, Association for Computing Machinery, New York, NY, USA (2009)
25. Yue, T., Kessentini, M., Pretschner, A., Voss, S.: Preface. In: 2019 ACM/IEEE 22nd International Conference on Model Driven Engineering Languages and Systems (MODELS). pp. 10–11 (2019)