# Evaluating a Framework of Conceptual Modelling Research

Jose Ignacio Panach[1[0000-0002-7043-6227]], Óscar Pastor [2[0000-0002-1320-8471]], Stephen W. Liddle [3[0000-0001-7671-4729]], Veda C. Storey[4[0000-0002-8735-1553]], Heinrich C. Mayr [5[0000-0001-5770-8091]] and Bernhard Thalheim [6[0000-0002-7909-7786]]

[1]Escola Tècnica Superior d'Enginyeria, Universitat de València, Valencia, Spain;
[2]VRAIN, Universitat Politècnica de València, Valencia, Spain;
[3]Brigham Young University, Provo, Utah, USA
[4]Georgia State University, Atlanta, Georgia, USA
[5]Alpen-Adria-Universität Klagenfurt, Klagenfurt, Austria
[6]Christian-Albrechts-Universität Kiel, Kiel, Germany

joigpana@uv.es, opastor@dsic.upv.es, liddle@byu.edu, vstorey@gsu.edu, Heinrich.Mayr@aau.at, bernhard.thalheim@email.uni-kiel.de

**Abstract.** Conceptual modelling aims to abstractly represent aspects of the perceivable world for a wide range of application areas so that the relevant part of an application can be represented and appropriate data included in a useful information system. Such abstraction efforts can be challenging and the resulting research papers have a wide variety of reported results. This makes it difficult to correctly identify and categorize the contribution of a specific research paper. Prior research has proposed a framework for characterizing the contributions of conceptual modelling research. In this paper, we assess the effectiveness of the framework's use to understand whether and how it could be applied to the large field of conceptual modeling. An experiment was conducted with 24 junior researchers as subjects. The evaluation of the experiment consisted of analysing the level of agreement among the subjects in the classification of a research paper's characteristics and analysing the correctness when classifying a paper. The data analysis showed a medium level of agreement and a good level of correctness. The outcome implies that framework helps identify the type of contribution made by a research paper, even when subjects are non-experts in conceptual modelling.

**Keywords:** conceptual modelling; empirical software engineering; evaluating research contributions, Characterizing Conceptual Modeling Research.

## 1 Introduction

Conceptual Modelling (CM) focuses on understanding, conceptualizing and representing 'reality' in different domains of interest, with a special emphasis on information systems design and development [1]. Consequently, the field of conceptual modelling research has become broad, diverse, and not always easy to understand. Prior research has studied the process of learning conceptual models (e.g., Bera et al. [2]). Delcambre et al. [3, 4] introduced a framework for understanding CM research and for characterizing CM research contributions, called Characterizing Conceptual Modeling Research (CMMR), which is the basis of experiments reported in this paper.

The CCMR framework is intended to be capable of capturing the main characteristics of CM research reported in a research publication. The diversity of conceptual modelling is accommodated while maintaining the core notions that are frequently associated with CM; for example, abstraction of the real world for which an application needs to be developed and communication mechanisms. One significant aspect is how the framework materializes the goal of providing authors with a standardized way to describe their work while highlighting their main contribution. This would allow researchers to position their work and enable others to assess the value of their CM contribution. The framework aims to be suitable for any user, even for non-experts in CM.

The main contribution of this paper is the analysis of the effectiveness of using a specific framework to characterize the contribution of a conceptual modelling research publication. The evaluation is conducted in terms of agreement in answering questions of the framework, and correctness when answering questions. The sample is composed of 24 master's students who have limited knowledge of CM but have some experience with research activities. The results show a medium value for the agreement (half of the questions yield a high agreement level), and a high level of correctness.

Section 2 describes related research efforts to classify research on conceptual modelling. Section 3 describes the CCMR framework. Section 4 describes our experiment, and Section 5 presents the results. Section 6 concludes the paper.

## 2 Related Works

Several frameworks have been developed to help classify and understand research articles. Some papers focus on the evaluation of the article review process, such as the work of Bianchi et al. [5] which analyses Royal Society journals from 2006 to 2017, measuring manuscript changes during the peer review from their initial submissions. Their evaluation is based on metrics for the number of reviewers, the evaluation of reviewers, and their citations after publication. A similar work was conducted by Bolek et al. [6], which aims to analyse reviewers' attitudes towards different sections of a manuscript during its review. They examined the consistency of a reviewer's evaluation throughout the review of the different parts of the paper, whether it corresponds to the recommendations of the editors, and whether the paper needs a new revision. Leung et al. [7] studies top-quality reviewers' perceptions of peer reviews and synthesizes their approaches to build a guide that helps reviewers. The results do not identify a single, structured approach to reading and reviewing manuscripts. Brown et al. [8] proposed a framework for the peer review process. This framework aims to be useful for both academics and anyone with enough knowledge to understand the proposal.

There are other works in literature that have proposed frameworks and guidelines for reading papers but not in a systematic manner. Snyder [9], defines a guideline to conduct literature reviews. The guidelines consider different types of reviews, as well as some recommendations to know how to both conduct and evaluate a literature review paper. Storey et al. [10] proposed a framework to capture the main beneficiary of a research study (who), the main type of research contribution produced ( what), and the research strategies used in the study (how). Franzago et al. [11] defined a classification framework considering as input the result of a mapping study to analyse collaborative model-driven engineering. The framework is composed of a set of concepts, references among concepts and attributes representing the set of data items extracted from each primary study. Vost et al. [12] was proposed to simplify and design case studies for comparing testing tools and make the results more reliable. That framework aims to facilitate the replication of case studies, helping in the aggregation of secondary studies.

These works provide an interesting background to classify and evaluate papers, but they are generic and do not focus specifically on CM. They are not suitable for a context beyond research. Moreover, none of the review guidelines and frameworks described here have empirical validation. Rather, they have been used to classify or aggregate papers, but lack an exhaustive analysis of their positive and negative aspects. As a step towards solving the gaps identified in the related works, this paper proposes the evaluation of a framework to classify papers that deal with conceptual modelling. The framework is composed of questions the reader must answer while (or after) reading the paper. The questions help to classify the paper and identify the contribution. They are defined so that even non-experts in conceptual modelling can apply the framework.

## 3 A framework to classify conceptual modelling research papers

This section summarizes the Characterizing Conceptual Modeling Research (CCMR) framework presented in Delcambre et al. [4]. The framework is composed of multiple questions that aim to characterize the conceptual modelling context of the work and the type of contribution associated with it. Table 1 shows the questions that compose the framework. The first part of the CCMR framework is intended to characterize the context (questions from Q1 to Q4). Concerning the possible types of contribution, three significant and distinct types of research contributions are proposed: 1) those centered on a model, language, metamodel, representation, or notation for CM (thus exploring "what" is being represented); 2) those centered on a method, process, tool, or algorithm for a CM-related purpose (thus examining "how" representations can be used); and 3) those that contribute CM vision, philosophy, principles, or a review (thereby encompassing a variety of CM aspects in a high-level way). For convenience, we refer to these types as "model", "method", or "vision" as a shorthand.

Once the subject chooses the contribution type (model, method, or vision), the questions are specific to that type. Questions specific to the model are from Q5 to Q6, questions specific to the method are from Q7 to Q9, and the question specific to vision is Q10. Q11 appears in all three contribution types, letting users introduce an additional free-text description. We are aware that some of the questions in Table 1 could be seen as ambiguous and open to multiple interpretations. Our experiment aimed to analyze how this ambiguity manifested in the specific practical setting we selected.

**Table 1.** Questions of the CCMR framework

| | |
|---|---|
| **1. Characterize the conceptual modelling context of this work** | |
| Q1 | List the representation used/treated as a conceptual model in this work |
| Q2 | Which groups of users are likely to create, modify, read or otherwise use this representation |
| Q3 | What is captured in this representation |
| Q4 | What is the level of abstraction for this representation |
| **2. Contribution type: a model, language, metamodel, notation** | |
| Q5 | List the model, conceptual modelling language, notation, or metamodel |
| Q6 | What is the nature of the significant contribution? |
| Q11 | Briefly describe the contribution |
| **3. Contribution type: method/process, tool, algorithm** | |
| Q7 | List the method, process, tool or algorithm |
| Q8 | What is its purpose? |
| Q9 | What is the nature of the significant contribution? |
| Q11 | Briefly describe the contribution |
| **4. Contribution type: CM vision, philosophy, principles, or review of the CM** | |
| Q10 | What type of high-level contribution is made? |
| Q11 | Briefly describe the contribution |

## 4 Experimental design

The analysis is done for each question, identified from Q1 to Q11 as shown in Table 1. The experiment was conducted at the Universitat Politècnica de València (Spain) in the Master of Engineering and Software System Technology program with 24 subjects. Participation in the experiment was mandatory since it was an evaluable task as part of the course, among other activities. The pedagogical value for students was that they would learn how to interpret and classify CM papers. Even though the CCMR framework can be helpful for CM experts, it is not expected that all CCMR users will have significant experience. We recruited inexperienced researchers to study whether CM characteristics included in the framework are understandable by subjects with minimal CM background. Even though subjects have a minimum level of experience reading research papers, most of them have not dealt with conceptual models previously.

### 4.1 Research questions, response variables, and metrics

We focus on studying the framework from the point of view of helping in the analysis of a selected paper. **RQ1:** *Does the framework help to achieve an agreement among* *several reviewers of the paper evaluation?* The response variable to answer this research question is “agreement.” IEEE [13] defines agreement as “mutual acknowledgement of terms and conditions under which a working relationship is conducted.” By “agreement” we refer to whether the criteria and classification established by each subject are the same as the criteria and classification of the other subjects for the same article. It is important to understand what criteria were used to classify the questions. We measure the agreement level using Blau’s Index [14], which quantifies the probability that two subjects randomly recruited from a population are in different categories of a variable. Its minimum value, zero, would indicate that all members agree on the same category, so there would be no variety. The maximum number, 1, means that subjects selected all the possibilities in the same proportion. Based on the Blau Index, we consider values less than or equal to 0.4 as a high agreement. To classify the possible results into quantifiable metrics, before conducting the experiment we designed an experimenter’s solution, where we divided the possible answers to each question of the framework into several categories. The proposed solution was obtained as a major agreement reached by the authors, senior researchers in the CM domain.

**RQ2:** *Does the framework help in the correctness of the paper evaluation?* The response variable to answer this research question is “correctness.” IEEE [13] defines correctness as “the degree to which software, documentation, or other items meet user needs and expectations, whether specified or not.” By “correctness” we mean whether the options marked in the framework by subjects agree with what experimenters have considered as possible correct answers for the evaluated paper. Gold standard answers were created by major agreement of the experimenter's team, who are experts in CM. Possible values for correctness are 1 or 0. Value 1 means that the subject’s answer agrees with the experimenters’ solution. Value 0 means the contrary. Note that the experimenters’ solution may involve several possible correct answers since there are some subjective aspects that, depending on how a paper is interpreted, may lead to multiple correct answers.

### 4.2 Procedure

Our design consists of reading three papers and applying the framework. We selected three papers to represent the three proposed contribution types (model, method, vision). Fonseca et al. [15], coded as Paper A, represents the “model” contribution type; Bour et al. [16], coded as Paper B, represents the “method” contribution type; and Harer et al. [17], coded as Paper C, represents the “vision” contribution type. All three papers were published recently in the International Conference on Conceptual Modeling (ER). The choice of this conference is justified by the fact that its topics match well the central topic of the framework: conceptual modelling.

Our experiment focuses on studying which aspects of the framework can be enhanced, rather than investigating differences between using our framework and not using it. To avoid bias from the order in which the papers are read, we divided subjects into three groups (G1, G2 and G3) and had them read papers in three balanced rounds of different papers. These are the steps of the experimental procedure:

- **Step 1: Demographic questionnaire**. Subjects fill out a demographic questionnaire.
- **Step 2: Framework introduction**. In a class session, one experimenter describes the framework to be used, and how to work with it.
- **Step 3: Training with the framework.** Subjects use the framework to analyse a paper. All subjects use the same paper for training, guided by the experimenter. The protocol and the instruments used in this training exercise are the same as those used in the experimental tasks. Subjects read the paper and fill out the questionnaire. These data are not analysed, as they are outside of the experiment.
- **Step 4: Applying the framework to the first paper**. The first paper to be analysed depends on the group where each subject has been classified. Subjects first read the paper and then applied the framework, reporting their analysis using a questionnaire.
- **Step 5: Applying the framework to the second paper**. This is a repetition of Step 4 but with another paper.
- **Step 6: Applying the framework to the third paper**. This is a repetition of Step 4 but with another paper (the third one).
- **Step 7: Filling satisfaction questionnaire**. At the end of using the framework, subjects fill out the satisfaction questionnaire.

## 5 Results

The results and replication package are available in Zenodo [18]. First, we analyse the agreement for the specific questions for each contribution type, only considering the subjects that made the correct identification of the type.

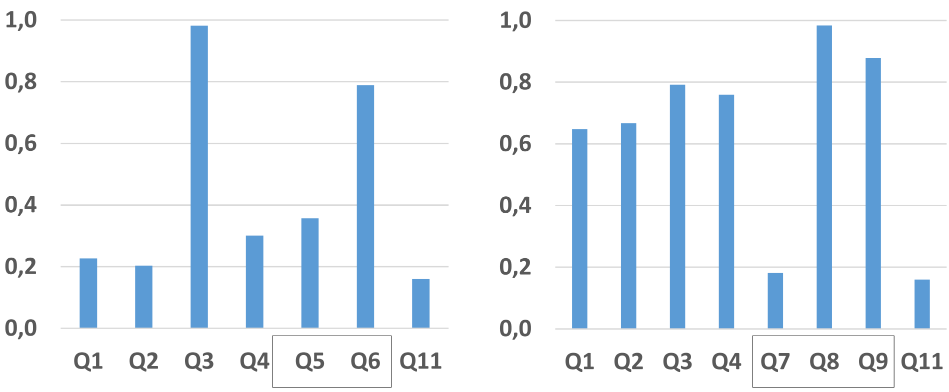


**Figure 1.**a) Results for agreement of Paper A. b) Results of Paper B

Figure 1.a quantifies agreement for Paper A [15], a paper of type "model." Results show that the level of agreement is quite good for most common questions except for Q3. Note that specific questions for a paper of type "model" are highlighted in a rectangle (Q5 and Q6). We can conclude that the identification of the nature of the contribution (Q6) seems to affect agreement analysis. Figure 1.b quantifies agreement for Paper B [16], a method paper. Q1 yields poor agreement. This means that the

determination of the conceptual models used in the paper is fuzzy. Q2 yields a low level of agreement. This leads to thinking that the target audience needs to be clarified. Q11 (brief contribution description) yields good agreement. Q8 and Q9, which are questions specific to this type of paper, yield a low agreement. This means that the purpose and the nature of the contribution are not clearly identified. Q7, which is also specific to the "method" contribution type, yields a high agreement which means that subjects agree on the method identification.

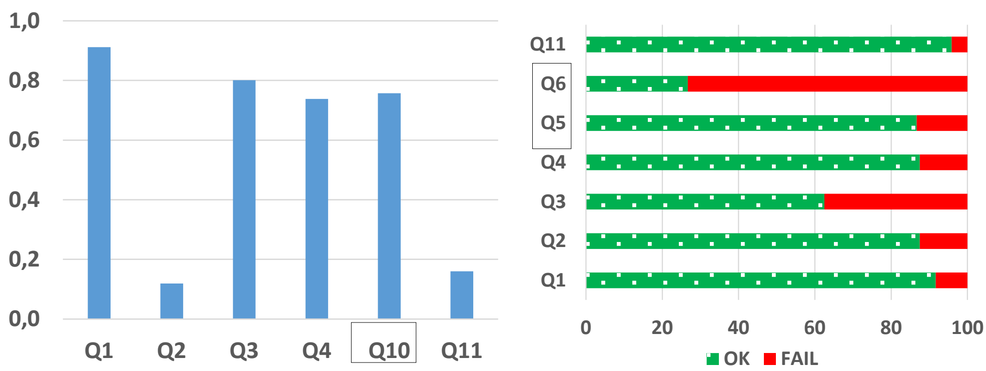


**Figure 2.** a) Results for agreement of Paper C. b) Results for correctness of Paper A

Figure 2.a shows the results for agreement of Paper C [17]. The only strong agreement appears in Q2, representing the groups interested in the conceptual model. This result for Q2 is shared in Paper A and Paper C. Q1 fluctuates between both replications, and Q3 shows a low agreement, similar to the other papers. Q11, which represents the contribution of the article, has a very strong agreement. Q10, which is the only specific question for this type of paper, does not yield a large agreement, which means that there is confusion in delimiting the main contribution of the paper.

Even if it is not easy to generalize the results because there are some discrepancies among papers, we can answer to RQ1 that the agreement level of the questions depends on the analysed question. Some questions have good agreement (Q1, Q2, Q5, Q7, Q11), at least in some cases, while other questions almost always have poor agreement (Q3, Q4, Q6, Q8, Q9, Q10). It seems to be clear that conclusions cannot be generalized for all questions that compose the framework. We want to emphasize that our purpose is to get concrete insights about the framework's use to identify points of improvement, and this goal is achieved with the information collected and reported in the paper.

To analyse the **correctness** of each type of paper independently, questions specific to each type of paper are analysed only if the subject correctly identifies the type of paper. Figure 2.b shows the results of correctness for Paper A [15]. Q1 and Q2 yield the best values, which means that the models used in the paper and the target users are identified with almost no error. The lowest correctness arises in Q6, which means that the nature of the contribution is fuzzy. Q3 also shows a low value, which means that subjects of such replication did not properly identify what is captured in the representation, while Q4 had a good value, indicating that what is the level of abstraction was well-identified. Q5 and Q11 show high correctness, which means that subjects correctly identified the type of contribution and its summary.

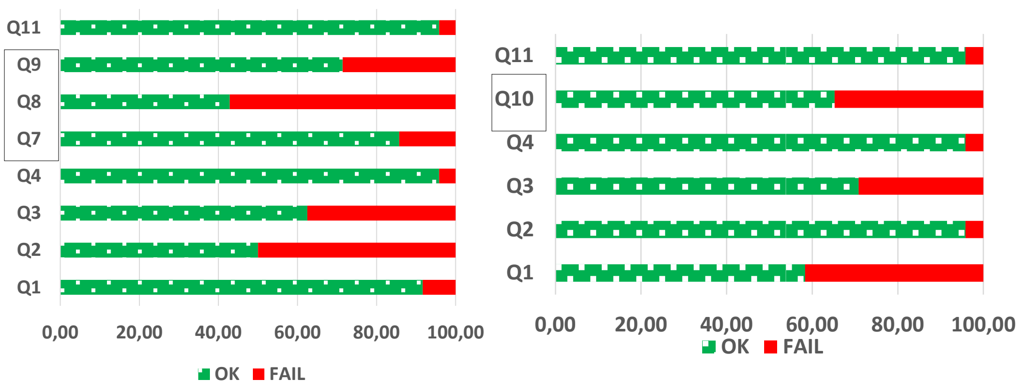


**Figure 3.** a) Results for the correctness of Paper B. b) Results for the correctness of Paper C

Figure 3.a shows the correctness results for Paper B [16]. Q1, Q4, Q7 and Q11 yield the best value. This means that the subjects in this paper identified the conceptual representations, the abstraction level, the method, and the contribution properly. The lowest correctness values arise in Q2, Q8, and Q9. This means that subjects did not identify the paper's target group of users, the purpose, and the nature of the significant contribution. Q3, although showing a low value of correctness, yields that many of the subjects appropriately identified what is captured in the representation (values are close to 60% of correctness). Figure 3.b shows the graphics for the correctness of Paper C [17]. It shows a good correctness level (more than 80% of correctness) for Q2 and Q4. This means that the target groups of the conceptual models and the level of abstraction are identified with a low error. The lowest correctness appears in Q1. There is a wide variety of answers when subjects have to specify the conceptual models used in the paper. Other questions (Q3, Q10, Q11) yield values between 60% and 80% of correctness, which means that there are a few errors when subjects identify what is captured in the conceptual model representation, what the contribution type is, and when the model or notation is listed. Thus, the answer to RQ2 is that, even though most of the questions that compose the framework are answered correctly (Q1, Q4, Q5, Q7, Q10, Q11) in most of the papers, there are some questions with many errors among the subjects (Q2, Q3, Q6, Q8, Q9).

To conclude this section, we discuss potential threats to the validity of our study. Following the conventional approach, we categorize these threats into four types: construct, internal, external, and conclusion validity threats. First, by selecting three specific papers—one for each type of possible contribution—it is possible that this choice introduced some bias. However, these papers were carefully chosen by the authors as adequate representatives of each contribution type, with a high level of agreement among them. Regarding internal validity, it could be affected by the criteria used to classify the selected papers. The team of authors, all of whom are experts in CM, reached a strong consensus on what should be considered the correct classification, and this consensus was used as the baseline, thereby mitigating the threat to internal validity. For conclusion validity, the evaluation of agreement could be influenced by the dispersion and subjectivity associated with some of the responses. To address this threat, we defined and employed a specific measurement criterion (based on the Blau Index, where values less than or equal to 0.4 indicate high agreement). In terms of correctness, another threat to conclusion validity could stem from subjectivity, as the experimenters' solutions might involve several possible correct answers. We mitigated this by collectively discussing the various answers to reach a unanimous decision. Finally, conclusion validity can also be impacted by the specific characteristics of the study population. While the feedback obtained helped us answer the two research questions (RQs), the reliability and generalizability of the results are limited due to the composition and size of the sample, which may introduce bias.

## 6 Conclusion

This paper presents the validation of a framework for characterizing conceptual modeling research that is intended to help understand and classify research papers related to conceptual modeling. The three main types of research contributions proopsed by the CCMR framework guide the work. An experiment was conducted with subjects with a low level of experience in research and conceptual modelling and, thus, were non-experts. The results reveal that, even though subjects are not experts in the conceptual modelling field, they obtained good agreement, and a very high value for correctness. The results have specific implications. First, the CCMR framework works in a similar way for the three possible types of contribution of the paper. Second, the correct application of the framework requires a modest amount of training. A brief introduction of 30 minutes to the framework is enough to apply it correctly. Third, some questions of the framework may lead to different correct answers. Finally, the results helped to identify the most subjective questions. Future research will involve replicating the experiment with different papers and additional subjects to further assess the generalizability of the results independently. We will also refine the CCMR framework by facing subjectivity in the answers that have generated low agreement and/or low correctness.

**Acknowledgments.** This work has been developed with the financial support of the Generalitat Valenciana under the project TENTACLE (GVRTE/2023/4592166) and the Spanish Ministry of Science and Innovation SREC (PID2021-123824OB-I00).

## References

1. Embley, D.W., Liddle, S., Pastor, Ó.: Conceptual-Model Programming: A Manifesto. Handbook of Conceptual Modeling. Springer (2011) 3-16

2. Bera, P., Poels, G.: How quickly do we learn conceptual models? European Journal of Information Systems **28** (2019) 663-680
3. Delcambre, L.M.L., Liddle, S.W., Pastor, O., Storey, V.C.: Characterizing Conceptual Modeling Research. (2019) 40-57
4. Delcambre, L.M.L., Liddle, S.W., Pastor, O., Storey, V.C.: Articulating Conceptual Modeling Research Contributions. In: Reinhartz-Berger, I., Sadiq, S. (eds.): Advances in Conceptual Modeling. Springer International Publishing, Cham (2021) 45-60
5. Bianchi, F., García-Costa, D., Grimaldo, F., Squazzoni, F.: Measuring the effect of reviewers on manuscript change: A study on a sample of submissions to Royal Society journals (2006–2017). Journal of Informetrics **16** (2022) 101316
6. Bolek, M., Bolek, C., Shopovski, J., Marolov, D.: The consistency of peer-reviewers: Assessment of separate parts of the manuscripts vs final recommendations. Accountability in Research (2022) 1-23
7. Leung, D., Law, R., Kucukusta, D., Guillet, B.D.: How to review journal manuscripts: A lesson learnt from the world's excellent reviewers. Tourism Management Perspectives **10** (2014) 46-56
8. Brown, L.M., David, E.A., Karamlou, T., Nason, K.S.: Reviewing Scientific Manuscripts: A Comprehensive Guide for Peer Reviewers. The Annals of thoracic surgery **103** (2017) 1365-1370
9. Snyder, H.: Literature review as a research methodology: An overview and guidelines. Journal of Business Research **104** (2019) 333-339
10. Storey, M.-A., Ernst, N.A., Williams, C., Kalliamvakou, E.: The who, what, how of software engineering research: a socio-technical framework. Empirical Software Engineering **25** (2020) 4097-4129
11. Franzago, M., Ruscio, D.D., Malavolta, I., Muccini, H.: Collaborative Model-Driven Software Engineering: A Classification Framework and a Research Map. IEEE Transactions on Software Engineering **44** (2018) 1146-1175
12. Vos, T.E.J., Marín, B., Escalona, M.J., Marchetto, A.: A Methodological Framework for Evaluating Software Testing Techniques and Tools. 2012 12th International Conference on Quality Software (2012) 230-239
13. IEEE: IEEE standard computer dictionary. A compilation of IEEE standard computer glossaries., Institute of Electrical and Electronics Engineers. New York, EE.UU. (1991)
14. Blau, P.M.: Inequality and heterogeneity: A primitive theory of social structure, Vol. 7. Free Press New York (1977)
15. Fonseca, L., Carbonera, J.L., Rodrigues, F.H., Antunes, C.R., Abel, M.: What Rocks Are Made of: Towards an Ontological Pattern for Material Constitution in the Geological Domain. Conceptual Modeling. Springer-Verlag, Salvador Brazil (2022) 275–286
16. Bour, R., Soule-Dupuy, C., Vallès-Parlangeau, N.: DEMOS: A Participatory Design Approach for Democratic Empowerment of IS Users. In: Laender, A.H.F., Pernici, B., Lim, E.-P., de Oliveira, J.P.M. (eds.): Conceptual Modeling. Springer International Publishing, Cham (2019) 387-394
17. Härer, F., Fill, H.-G.: Past Trends and Future Prospects in Conceptual Modeling - A Bibliometric Analysis. In: Dobbie, G., Frank, U., Kappel, G., Liddle, S.W., Mayr, H.C. (eds.): Conceptual Modeling. Springer International Publishing, Cham (2020) 34-47
18. Panach, J.I., Pastor, O., Liddle, S.W., Storey, V.C., Mayr, H.C., Thalheim, B.: Dataset of the Article Evaluating a Framework of Conceptual Modelling Research. (2024)