What are the used UML diagrams? A Preliminary Survey

Gianna Reggio, Maurizio Leotta, Filippo Ricca, Diego Clerissi

DIBRIS – Università di Genova, Italy
gianna.reggio@unige.it, maurizio.leotta@unige.it,
filippo.ricca@unige.it, diego.clerissi@gmail.com

Abstract. UML is a large notation offering many diagrams and a large set of constructs for each of them covering any possible modelling need. As a result its specification is a huge book, its metamodel is large, and defining/understanding its static and dynamic semantics is difficult. These features have a negative impact on the perception of the UML and lead in some cases to replace it by ad-hoc lean and simple DSLs. Thus, the following question arises: which are the most/less used UML diagrams/constructs? We would like to answer to this question by means of a survey, trying to detect which parts of the UML are the most used. To see how much a diagram is used we preliminarily investigate books, courses/tutorials, and tools covering UML. The less used diagrams will be then analysed to understand the reasons of their low usage.

1 Introduction

UML is a truly large notation offering many different diagrams (14 in the last approved version [15]), and for each diagram it provides a large set of constructs covering any possible need of any modeller for any possible task. As a result, the UML specification is a huge book (732 pages), the UML metamodel is large and quite complex, and the definition and the understanding of its static and dynamic semantics is a truly difficult task, with also the consequence to make difficult to teach it both at the school/university level or in the industry [4]. Moreover, the large number of constructs and the consequent very large metamodel make complex and time consuming developing transformation of UML models and building tools for the UML. Clearly, these features of the UML have a negative impact on how the UML is perceived by the modellers hindering its adoption, and leading in some cases to replace the UML by ad-hoc lean and simple DSLs (Domain Specific Languages).

On the other hand, users naturally tend to consider and use only a portion of its diagrams/constructs, and forgetting about some other ones. On his blog I. Jacobson states "For 80% of all software only 20% of UML is needed" [6]. Furthermore, few people try to learn the UML by reading its specification [15], instead the large majority of the users relies on books and courses/tutorials or just start to use some tools for drawing UML diagrams that in general do not cover the whole UML. For this reason, in many cases, the UML users will never

become aware of the existence of many specific constructs (e.g., how many of you do know the existence of the "Parameter Set" for activities or has even used this construct?).

We would like to asses by means of a survey which parts of the UML (diagrams and constructs) are the most used in practice and which the less ones, using again the words of Jacobson trying to see if an "essential UML" [6] emerges. To discover how much a UML diagram/construct is used, we chose to preliminarily investigate: (1) the books about the UML, (2) the IT University courses covering also the UML, (3) the tutorials presenting the UML to the practitioners, and (4) the tools for producing UML models. Later, similarly to [2, 4], we will conduct a personal opinion survey [5] asking to UML users of different kinds (e.g., industrial practitioners and academics) which parts of the UML they know and which they have never used.

For a given UML diagram, we have proceeded as follows. We have investigated the books to discover if they were citing such diagram, and if they were giving an example of it. Similarly for the courses/tutorials, whereas for the tools we have tried to produce a model containing such diagram. Finally, we have computed descriptive statistics to present the results.

The results of this survey, of which this work constitutes the first step, and of the future personal opinion survey should be of help to many different categories of people:

- **teachers and instructors**: allowing to offer courses and/or tutorials concentrating only a smaller language made out of the most used UML diagrams/constructs;
- **tool builders/users**: obvious advantages since the tools covering the most used diagrams/constructs will be simpler to implement/use.
- **notation designers**: interested in discovering scarcely used constructs, and understanding for which reasons they have been added to the language. Moreover, other interesting questions arise: are the scarcely used constructs derived[1] or primitive? can the scarcely used constructs be applied only in specific cases? It will be interesting to investigate whether the metamodel (and subsequently the UML specification book) may be easily simplified to cover only the most used constructs.

In our opinion having to handle a notation with a large set of constructs where a portion of them are scarcely used, if not almost unknown, is problematic because it can cause a waste of effort and of resources by who want/must use it (e.g., in Italy, some contracts with the public administration must be accompanied by a UML model). From the trivial fact that to print the reference requires 700 sheets, to the fact that understanding the metamodel/preparing for the certification/deciding what to teach to the students/reading a UML book require a large number of hours, and we do not have to forget that also maintaining the official specification and any related item requires a large amount of effort due to its size. Also the OMG has recently recognised the need to simplify

[1] A derived construct may be replaced by a combination of other constructs.

the UML with the initiative "UML Simplification" [12] which will result in the next UML version (2.5), but in this case the simplification concerns only the way UML is defined without any impact about its constructs.

In this paper we present the results of the first step of our survey, i.e., focusing on the usage level of the 14 types of UML diagrams and covering books, courses/ tutorials, and tools. The remainder of the paper is organized as follows. In Sect. 2, we present related work literature regarding empirical study about the UML. In Sect. 3, we sketch the relevant aspects of the conducted empirical work such as: goals, research questions, followed process and analysis methodology. The results of the survey are presented in Sect. 4, while threats to validity are discussed in Sect. 5. Finally, Sect. 6 concludes the paper.

2 Related Work

The systematic literature review (SLR) by Mohagheghi et al. [8] about model-based software development states that "the UML is currently the most widely used modelling language". A similar result has also been obtained in [14] where a personal opinion survey with 155 Italian professionals has been conducted. Another personal opinion survey about UML (171 professionals in total), by Dobing and Parsons [2], points out another strong statement: "regular usage of UML components were lower than expected". The authors of [2] suggest that the difficulty of understanding many of the notations "support the argument that the UML may be too complex". The same claim, in more or less different forms, is present in several blogs, where several proposals of UML simplification are arising[2]. Maybe, the most authoritative is the one of Ivar Jacobson entitled "Taking the temperature of UML" [6], where he wrote: "Still, UML has become complex and clumsy. For 80% of all software only 20% of UML is needed. However, it is not easy to find the subset of UML which we would call the 'Essential' UML. We must make UML smarter to use". The need to simplify the UML is also shown by the recently released OMG draft proposal about this topic [12].

In the tentative to find the "essential UML", Erickson and Siau [3] have conducted a Delphi study[3] with the goal of identifying a UML kernel for three well-known UML application areas: Real-Time, Web-based, and Enterprise systems. The participants to the study (44 experts in total) were asked to rate the relative importance of the various UML diagrams in building systems. UML overall results (i.e., non-domain specific) were: 100% for Class and Statechart diagrams, 95.5% for Sequence diagrams, 90.9% for Use Case diagrams. All the others diagrams received a percentage lesser than 50% (e.g., 27.3% for Activity diagrams). Another personal opinion survey (sample size = 131) about UML [4] confirms the results of Erickson and Siau. Results indicate that the three most important diagrams are Use Case diagrams, Class diagrams and Sequence diagrams.

[2] e.g., `www.devx.com/architect/Article/45694` and `blogs.msdn.com/b/sonuarora/archive/2009/11/02/simplify-uml.aspx`

[3] It attempts to form a reliable consensus of a group of experts in specialized areas.

The main conclusions from another SLR by Budgen et al. [1] about empirical evidence of the UML are two:

– while there are many studies that use the UML in some way, including to assess other topics, there are relatively few for which the UML is itself the object of study, and hence that assess the UML in some way (e.g., UML studies of adoption and use in the field).
– there is a need to study the UML and its forms much more rigorously and to identify which features are valuable, and which could be discarded.

3 Study Definition

The instrument we selected to take a snapshot of the state of the practice concerning UML usage is that of a survey. In the survey's design and execution phases we followed as much as possible the guidelines provided in [5] and used the same presentation format of [13, 11].

The survey has been conducted through the following steps: (1) goals selection, (2) goals transformation into research questions, (3) identification of the population, sample and process, (4) data extraction and, (5) analysis of results and packaging.

We conceived and designed the survey with the goal of understanding *which are the less/most used parts of the UML in practice*. Within the scope of this work, in this paper we aim at addressing the main research question related to the above described goal:

Which of the 14 types of UML diagrams are the most/less used in practice?

The first step to conduct a survey is defining a target population. The target population of our study consists of *sources* concerning UML. In particular, in this study we considered the following four kinds of sources: books, tools, courses and tutorials.

To sample the population and select the sources to consider in our study we: (1) conducted a systematic search performed using Internet resources, Web search engines and electronic databases and (2) used non-probabilistic (convenience sampling) methods [7]. Moreover, in making decisions about whether (or not) to include a source in the study, we adopted some inclusion/exclusion criteria (see subsection below).

After having collected the sources, we extracted the data of interest for our research questions and finally we performed the analysis. Given the nature of this survey, that is mainly descriptive (it describes some condition or factor found in a population in terms of its frequency and impact) and exploratory, we mainly applied descriptive statistics and showed our findings by means of charts.

3.1 The Inclusion and Exclusion Criteria

The inclusion/exclusion criteria can be common for all the kind of sources or specific. For all the sources we adopted the following inclusion criterion: only sources concerning UML versions ≥ 2.0. Concerning books, in case of different editions of the same book we opted (when possible) for the last one. Moreover, we excluded

elements of "grey" literature, i.e., books without ISBN. Concerning tools, we included only UML modelling tools (both commercial and non-commercial) and excluded: (1) general graphics editor (e.g., Inkscape), (2) tools providing only a specific type of diagram (e.g., class diagrams), (3) really unstable, not complete or preliminary tools (e.g., tools in beta version). About courses, we considered only university courses concerning IT studies. We considered courses offered also in languages different from English (e.g., Italian and Spanish). Concerning tutorials, we considered only tutorials provided on Internet as written documents (either on-line or downloadable) and video (where a person gives instructions on how to do something) but we have excluded tutorials taking the form of a screen recording (screencast) and interactive tutorial. For selecting a document of this kind we used the common meaning/perception of tutorials: a tutorial is more interactive and specific than a book or a lecture; a tutorial seeks to teach by example.

3.2 The Process

The process followed to conduct a survey should be as much as possible well defined in order to ensure that such a study be objective and repeatable. For each category of sources, we followed a different process to collect the sources.

Books. We started by the Amazon website[4] and used the search form to find UML related books. We selected the "Computers & Technology" category in the books section. Then, we experimented with several different search criteria using different combinations of strings. Finally, the one that retrieved the highest number of useful items was the simple string "UML 2". Starting from this long list of books ordered by relevance[5] we filtered out books not satisfying the inclusion criteria explained above. Then, we tried to recover them using the facilities provided by our library. Finally, we collected and analysed 30 books. Note that, 18 of them are in the top 24 books ranked ordering the list by relevance by Amazon. The complete list of books is not reported here for space reason and can be found on the online technical report [10]. It includes, for instance, *"The Unified Modeling Language Reference Manual"* and *"The Unified Modeling Language User Guide"* by J. Rumbaugh, I. Jacobson, G. Booch and *"UML Distilled"* by M. Fowler.

Tools. We started by the "List of Unified Modeling Language tools" Wikipedia page[6] containing 49 UML tools. Then, we considered also the UML-tools website[7]. A full Internet search was also carried out using Google. Also in this case, we experimented with several different search criteria using different combinations of strings to provide to Google ("UML tools", "UML tools list" and "UML free tools"). For each tool of our list, we found the official website and selected only the tools satisfying the inclusion criteria explained above. Then, we downloaded and installed the most recent version of all the selected tools. In case

[4] http://www.amazon.com

[5] 2.726 books on July 20, 2013.

[6] http://en.wikipedia.org/wiki/List_of_Unified_Modeling_Language_tools

[7] http://www.uml-tools.com

of commercial tools, we selected a "free for not commercial use" version or a version with university licence or a trial version. At the end, we collected and analysed 20 different tools. The complete list of tools is not reported here for space reason and can be found on the online technical report [10]. It includes, for instance, *"Visual Paradigm"*, *"MagicDraw"* and *"IBM Rational SW Architect"*. Finally, we tried to produce a model containing the diagrams and constructs of interest for our study (for each tool we produced the same model with the same diagrams and the same UML constructs).

Courses. We started carrying out a search using Google. The combinations of strings used were: "UML course", "UML lecture" and "UML university course". We found several university courses satisfying the inclusion criteria stated above, but in several cases it was difficult, if not impossible, to recover the slides of the lectures, and in general the material. Often, the material was not publicly available; only the content of the lessons was present on the website. For this reason, we resort also to convenience sampling, asking to our colleagues the slides of UML courses they teach. At the end, we collected and analysed 22 different University courses. The complete list of lectures is not reported here for space reason and can be found on the online technical report [10]. It includes, for instance, courses from Canada, UK, USA, Hungary, Germany, Italy, France, Spain, Argentina, Australia. Convenience sampling was also useful to balance a little the geographic origin of the UML courses (e.g., before convenience sampling we had three France courses and zero USA courses).

Tutorials. We started with the tutorials lists present in the following three websites: www.uml.org[8], stackoverflow.com[9] and www.jeckle.de[10]. Then, we integrated the obtained results with other tutorials recovered using Google (the research was conducted using the strings: "UML tutorials" and "UML guide"). Finally, we collected and analysed 18 tutorials. The complete list of tutorials is not reported here for space reason and can be found on the online technical report [10].

4 Survey Results

We preliminarily decided to interpret the results of our survey assuming that a diagram is *widely used* if it is present in the $\geq 60\%$ of the sources, similarly it is *scarcely used* if it is present in $\leq 40\%$ of the sources, having also some non-defined cases (*grey zone*). In the following subsections we present the results concerning UML diagrams.

The level of usage of the various UML diagrams in books, courses, tutorials, tools, and in the totality of the sources respectively is summarized in Fig. 1. If we consider the totality of the sources, disregarding their kind, we have that the scarcely used diagrams are timing, interaction overview and profile, listed starting from the most used; all of them were not present in UML 1.x, and the profile diagram appeared only in version 2.2. The last position of the profile

[8] http://www.uml.org/#Links-Tutorials

[9] http://stackoverflow.com/questions/1661961/recommended-uml-tutorials

[10] http://www.jeckle.de/umllinks.htm#tutorials

UML Diagram	Book Guide	Book Spec	Book Tot	Tool	Course	Tutorial	All Sources
Class	100%	100%	100%	100%	100%	100%	**100%**
Composite Structure	87%	60%	73%	80%	14%	33%	**52%**
Component	93%	80%	87%	85%	59%	89%	**80%**
Deployment	93%	80%	87%	90%	55%	89%	**80%**
Object	93%	80%	87%	70%	55%	67%	**71%**
Package	100%	79%	89%	65%	52%	67%	**70%**
Activity	100%	93%	97%	100%	95%	100%	**98%**
Sequence	100%	93%	97%	100%	100%	89%	**97%**
Communication	100%	80%	90%	90%	59%	89%	**82%**
Interaction Overview	80%	53%	67%	45%	5%	28%	**39%**
Timing	87%	53%	70%	40%	5%	33%	**40%**
Use Case	100%	93%	97%	100%	95%	89%	**96%**
State Machine	100%	93%	97%	100%	95%	89%	**96%**
Profile	7%	13%	10%	30%	0%	6%	**11%**

Fig. 1. Usage levels of UML diagrams - ("Book Tot" means all the books)

diagram is not very surprising due both to the late appearance and to the fact that this kind of diagram has a very restrict scope (indeed it is used only to present a profile) and that, it is essentially a variant of the package diagram. Also timing diagrams, see an example in Fig. 2, have a restrict scope, and UML offers other ways to model time related aspects (e.g., timed events may be used in state machines and activity diagrams; durations and time intervals may appear in sequence diagrams), and this may be the motivation for their low usage. Finally, interaction overview diagrams are quite complex and in many cases may be replaced by sequence diagrams and/or a combination of sequence and activity diagrams, and perhaps this is the reason for not being so considered.

The widely used diagrams, when considering the totality of the sources, are instead, listed again starting from the most used ones, class (100% in any kind of sources), activity, sequence, state machine and use cases, communication, deployment, component, object and package diagrams. The first position of class diagrams is not surprising, it is indeed the fundamental diagram of the UML, while the fact that activity diagrams are the second is relevant and is due, in our opinion, to the fact that they are used also for business process modelling and for SOA based systems. All the widely used diagrams, except the package diagram, were already present in the UML 1.x versions (also if the communication diagrams were before called collaboration diagrams). The only diagram in the grey area (i.e., above 40% and below 60%) is the composite structure, that allows to represent both structured classes and collaborations; again it is a new diagram appearing in the UML 2.0 and this may be a reason for its low usage. How-

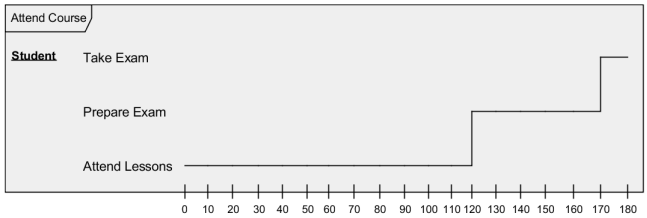

Fig. 2. Timing Diagrams, usage level 40% in the totality of the sources

ever, the result is surprising because structured classes were completely absent in UML 1.x, and this was a perceived problem, and the new collaborations are truly useful (see for example the big role that they have in representing service oriented architectures in the SoaML OMG standard profile [9]).

The results are different, see Fig. 1, if we distinguish the various kind of sources. In the case of books we have considered two kinds of books: 15 UML guides (Book Guide) and 15 books that make usage of UML but where it is not the primary subject (Book Spec). In the case of the UML guides, all diagrams except the profile diagrams are widely used. This result may perhaps be due to the fact that we count as present in a book a diagram also if it is only mentioned. Unfortunately, it is really difficult to devise a better metrics; for example trying to distinguish if a diagram is just mentioned, shortly presented, presented, or presented with all the details may to be too much depending on the personal judgement of who examines the books; also counting the occurrence of the name of a construct/diagram is in our opinion too dependent on the way the books were written, e.g., more or less verbose. We have also tried to distinguish the case of a simple mention of a construct in the text and the presence of an example of such construct, without detecting a relevant difference. Similar results, even if slightly reduced in magnitude, came from the other category of books (Book Spec).

For the tools the only scarcely used diagrams, see Fig. 1, are timing and profile diagrams, whereas the interaction overview diagram is the only one in the grey area. This result is a little surprising, since one can expect that due to the effort required to add a feature to a tool, the less relevant diagrams are omitted in many cases. In our opinion, since once you have available the functionalities to draw a class diagram it is quite easy to add the possibility to handle composite structure, component, deployment, object, package, communication and profile diagrams, whereas timing and overview requires new graphical functionalities.

Considering only the courses, we have a striking distinction between the diagrams; indeed, class, sequence, activity, state machine and use case diagrams are widely used with percentage over 90%, component diagram, deployment, object package, communication are in the grey area (all above or equal to 50%), whereas composite structure (14%), interaction overview (5%) and timing (5%) are really scarcely used. A lecturer preparing in a course has to decide which are the most relevant diagrams to present to the students in the allowed time slots, and it seems that this decision is quite homogeneous: there are five fundamental diagrams in the UML for the lecturers (class, sequence, activity, state machine and use case).

For the tutorials, we have a neat distinction between the widely used diagrams (class, activity, component, deployment, sequence, communication, use case and state machine) and the scarcely used (composite structure, timing, interaction overview and profile diagrams), but the differences in the usage level are lower than for courses, e.g., composite structure, timing, interaction overview percentages are more than 25%. In this case the tutorial writer has to decide

what is relevant, but s(he) usually has no strong constraints on the size of the tutorial itself.

5 Threats to Validity

To avoid to bias the results of our survey, we have considered only sources concerned with the use of the UML, avoiding those with different aims, for example drawing tools suitable to produce pictures of UML diagrams, or books presenting a survey on the current visual notations have been excluded; whereas instead books covering specific use of the UML or courses about software engineering where the UML was taught were included.

For the tools, instead, we are quite confident to have examined almost all the available ones; we think that a UML tool cannot exist without being presented somewhere on the Web. Notice that Argo UML, one of the most known UML tool was not included in our survey since it does support only UML 1.x.

We have considered here only four kinds of sources (books, courses, tutorials, sources) and we are aware that these are not the only ones; indeed there are also the UML users, and we are ready to launch a personal survey to investigate which constructs they know and which they use.

Finally, we have decided to define widely used (scarcely used) when a diagram was considered in the $\geq 60\%$ ($\leq 40\%$) of the sources, resulting also in a grey area. We think that this a sensible choice, using a threshold lower than 60%, e.g., 50%, should have led to have that a construct is either widely used or scarcely used without any doubt cases, and this does not sounds realistic. On the other hand, a higher threshold, e.g., 80%, should have led to a quite large number of inconclusive cases. We have also computed the widely/scarcely used on the totality of the sources, disregarding the fact that they are of very different kinds, e.g. books and courses, and so assigning to them different weights would have been more realistic. Again, we had the problem to compute these weighs in an unbiased way: is it sensible to say that a book is three times more relevant than a course, or that a tool is two times more relevant than a tutorial? To avoid to make our result too dependent on our personal judgement we have preferred to assume that all the sources have the same weight.

6 Conclusions

We have investigated, by means of a survey, how much the UML diagrams are used, considering in this paper four kinds of sources: books, tools, courses, and tutorials concerning the UML itself. The results of our survey show that the usage of the various diagrams is different . An "essential" UML seems to emerge also if its boundaries are not exactly defined. For what concerns the diagrams, class, activity, sequence, use case and state machine diagrams are widely used without any doubts, whereas interaction overview, timing and profile diagrams are really scarcely used.

In this paper we have considered only "objective" sources and examined them for checking if UML diagrams are used in an objective way (e.g., can a tool

produce a model including such diagrams?, is a course/tutorial teaching the fact that UML has such diagrams?), so these results are not biased by any personal opinion (neither ours nor of any human being taking part in the examination of the sources). We are now investigating the usage of the constructs used in the UML diagrams and launching a personal survey to investigate which UML diagram/constructs are known and used by UML users trying to cover different categories of them, and different applicative fields. The combined results of the two surveys should lead to finally determine an "essential" UML.

References

1. D. Budgen, A. J. Burn, O. P. Brereton, B. A. Kitchenham, and R. Pretorius. Empirical evidence about the UML: a systematic literature review. *Software Practice and Experience*, 41(4):363–392, Apr. 2011.
2. B. Dobing and J. Parsons. How UML is used. *Communications of the ACM*, 49(5):109–113, May 2006.
3. J. Erickson and K. Siau. Can UML be simplified? practitioner use of UML in separate domains, (white paper).
4. M. Grossman, J. E. Aronson, and R. V. McCarthy. Does UML make the grade? insights from the software development community. *Information and Software Technology*, 47(6):383–397, Apr. 2005.
5. R. M. Groves, F. J. J. Fowler, M. P. Couper, J. M. Lepkowski, E. Singer, and R. Tourangeau. *Survey Methodology*. John Wiley and Sons, 2009.
6. I. Jacobson. Taking the temperature of UML. WEB site `blog.ivarjacobson.com/taking-the-temperature-of-uml/`, 2009.
7. B. Kitchenham and S. Pfleeger. Personal opinion surveys. In F. Shull and Singer, editors, *Guide to Advanced Empirical Software Engineering*, pages 63–92. Springer London, 2008.
8. P. Mohagheghi, V. Dehlen, and T. Neple. Definitions and approaches to model quality in model-based software development - a review of literature. *Information and Software Technology*, 51(12):1646–1669, Dec. 2009.
9. OMG. *Service oriented architecture Modeling Language (SoaML) Specification Version 1.0.1*, 2012. Available at `www.omg.org/spec/SoaML/1.0.1/PDF`.
10. G. Reggio, M. Leotta, F. Ricca, and D. Clerissi. What are the used UML diagrams? A preliminary survey - Technical Report. Technical report, DIBRIS, University of Genova, July 2013. Available at `http://softeng.disi.unige.it/TR/UsedUMLTechnicalReport.pdf`.
11. G. Scanniello, F. Fasano, A. D. Lucia, and G. Tortora. Software quality assessment and error/defect identification in the Italian industry preliminary results from a state of the practice survey. In *Proceedings of 6th International Conference on Software Engineering Advances (ICSEA 2011)*, pages 589–594, 2011.
12. E. Seidewitz. Uml 2.5: Specification simplification. Presented at "3rd Biannual Workshop on Eclipse Open Source Software and OMG Open Specifications", 2012.
13. M. Torchiano, M. D. Penta, F. Ricca, A. D. Lucia, and F. Lanubile. Migration of information systems in the italian industry: A state of the practice survey. *Information and Software Technology*, 53(1):71 – 86, 2011.
14. M. Torchiano, F. Tomassetti, F. Ricca, A. Tiso, and G. Reggio. Relevance, benefits, and problems of software modelling and model driven techniques a survey in the italian industry. *Journal of Systems and Software*, 86(8):2110 – 2126, 2013.
15. UML Revision Task Force. *OMG UML, Superstructure, V2.4.1*, 2011.