# A Comparative User Evaluation on Visual Ontology Modeling Using Node-Link Diagrams

Muhammad Rohan Ali Asmat[1], Vitalis Wiens[1,2], and Steffen Lohmann[1]

[1] Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS, Sankt Augustin, Germany
muhammad.rohan.ali.asmat@iais.fraunhofer.de
vitalis.wiens@gmail.com
steffen.lohmann@iais.fraunhofer.de
[2] TIB Leibniz Information Center for Science and Technology, Hannover, Germany

**Abstract.** The emergence of several ontology modeling tools is motivated by the growing attention ontologies receive in scientific and industrial contexts. The available tools implement different ontology modeling paradigms, including text-based editors, graphical user interfaces with hierarchical trees and form widgets, and visual modeling approaches based on node-link diagrams. In this paper, we present an empirical user study comparing a visual ontology modeling approach, based on node-link diagrams, with a modeling paradigm that uses hierarchical trees and form widgets. In particular, the user study compares the two ontology modeling tools Protégé and WebVOWL Editor, each implementing one of the modeling paradigms. The involved participants were given tasks of ontology modeling and also answered reflective questions for the individual tools. We recorded the completion times of the modeling tasks and the errors made as well as the users' understanding of the conceptual spaces. The study indicates that visual ontology modeling, based on node-link diagrams, is comparatively easy to learn and is recommended especially for users with little experience in ontology modeling and its formalization. For more experienced users, no clear performance differences are found between the two modeling paradigms; both seem to have their pros and cons depending on the type of ontology and modeling context.

**Keywords:** Ontology Engineering, Visual Modeling, Visualization, OWL, VOWL, WebVOWL, User Study, Comparative Analysis.

## 1 Introduction

A fundamental aspect of the Semantic Web is to create and communicate conceptualizations of information and data in certain domains. Ontologies serve this purpose by providing a formal representation of domain knowledge which is shareable across the web in a machine readable format [6]. The development of various ontology modeling tools with different modeling paradigms is triggered by the growing attention ontologies receive in scientific and industrial contexts. Ontologies are used in tasks such as exploring and studying a new subject domain, automated information retrieval, and learning management [3].

This work will be published as part of the book "Emerging Topics in Semantic Technologies. ISWC 2018 Satellite Events. E. Demidova, A.J. Zaveri, E. Simperl (Eds.), ISBN: 978-3-89838-736-1, 2018, AKA Verlag Berlin".

Modeling of ontologies has not remained limited to ontology engineers, as nowadays different communities are pursuing towards formal representation of domain knowledge. Thus, modeling of ontologies is often done collaboratively in joint efforts of knowledge engineers and domain experts. On the one hand, domain experts, providing the conceptualization of the knowledge domain, are typically not familiar with semantic formalism and conceptual modeling techniques. They often find it hard to follow logical notations in OWL representation. On the other hand, ontology engineers, who provide the necessary know-how for ontology modeling and logical notations in OWL, usually lack the expertise in the domain to create ontologies of sufficient quality [5].

Several ontology engineering tools implementing different modeling paradigms have been developed in the last years. However, they are mostly designed for ontology engineers with profound knowledge in ontology modeling. The different modeling paradigms range from direct text input, UML-based graphs [17], widget and hierarchical based GUIs [16], node-link diagrams [7,8], to hybrid solutions like Turtle Editor [14].

In this paper, we evaluate the visual ontology modeling paradigm using node-link diagrams with WebVOWL Editor[3]. WebVOWL Editor exploits the VOWL notation, which is a well-defined visual notation for OWL ontologies. It is designed for the user-oriented representation of ontologies that is easy to understand [12]. This visual ontology modeling tool allows us to conduct our evaluation on different target groups including non-expert users. The current implementation of this tool does not (yet) support all OWL constructs, however, it covers all required ones for our comparative evaluation.

We have conducted a comparative, empirical user study over two different ontology modeling tools (teh well-known Protégé and the aforementioned WebVOWL Editor). The study involved 12 participants, comprising of master students, PhD students, and post docs in the field of computer science. The participants modeled ontologies with the individual tools and also answered reflective questions respectively. The results indicate that the visual ontology modeling paradigms are easier to learn and use for non-expert users and that these require less time for the creation of *small* ontologies. The scores for expert users were not that significant due to a high variance in their prior experience with different ontology modeling tools. Therefore, we purpose a follow up study with a controlled prior experience of tools and increased number of participants.

After introducing the related work in Section 2, we describe our pretest in Section 3. The pretest defines the concept spaces used in the final user evaluation. The design for the evaluation is specified in Section 4. After presenting the results of the user study in Section 5, we continue by drawing conclusions in Section 6.

## 2 Related Work

The diversity of ontology modeling paradigms and tools increased also the interest in their benefits and drawbacks. Thus, several evaluations have been conducted, investigating users' understanding of ontology representations and the effectiveness of the different tools. An overview of the different ontology visualization tools can be found in the work of Anikin et al. [3]. An evaluation on visual modeling was conducted by García et al. [13], investigating the effectiveness and usability of the tool OWL-VisMod.

[3] The tool and GitHub repository can be found at `https://w3id.org/webvowl/editor`

Katifori et al. [11] conducted a comparative user study of four ontology visualization tools. Users were asked to perform information retrieval tasks with each tool, such as finding a value of some property. Time to accomplish each task was measured and the users were interviewed afterwards. Based on the answers, the effectiveness of each tool was measured. According to that, Protégé Entity Browser is the most effective, then Jambalaya, TGViz and OntoViz is the least effective.

Fu et al. [9] compared the representation of ontologies with indented lists and node-link diagrams. The users were asked to evaluate and create new mappings between ontologies using the two modeling paradigms. In this work, Fu et al. found out that indented lists are more suitable for the evaluation of the mappings, whereas node-link diagrams are better suited for creating new mappings and for showing an overview of the ontology. In a follow-up study, Fu et al. [10] used eye-tracking technology to investigate the differences between indented lists and graphs in more detail.

Most of the existing evaluations however are focusing on information retrieval tasks and on investigating how the comprised information of an ontology is communicated to the users. In contrast to the comparison of different representations of ontologies, this paper aims to fill the research gap by investigating the potential benefits and drawbacks of different modeling paradigms for ontology generation. A pretest defines concept spaces that are used as modeling task in our evaluation. Participants had to model small ontologies using two ontology modeling tools Protégé and WebVOWL Editor. Modeling completion times were measured and additional questionnaires were used in order to determine the potential benefits and drawbacks of the individual tools.

## 3 Pretest

In advance to the user evaluation, we conducted a pretest. It includes 1) the definition of concept spaces and 2) the identification of their individual difficulty levels respectively. The results of the pretest are used for the comparative user evaluation for visual ontology modeling using node-link diagrams and a hierarchical tree based approach.

### 3.1 Concept Spaces for the User Study

Prerequisite to the pretest, we introduce five small concept spaces. These are defined with an idea of having a small generalized set of domain knowledge in order to evaluate different ontology modeling tools. The concept spaces define common, every-day knowledge, whereas each comprises of thirteen concepts. In this paper, concept are associated with classes, subclasses, object properties, or datatype properties. Our defined set of domain knowledge includes the following concept spaces: University Space, Zoo Space, Media Space, Family Tree Space, and City Traffic Space (cf. Appendix A). The cognitive complexity of all concept spaces is balanced by:

1. Asking a person to define the concept spaces that are equal in hierarchical and graphical representations while created using any ontology modeling tool or even realized on paper. Repetitive iterations were performed on paper, defining the concepts for each individual domain knowledge.
2. Evaluating the difficulty levels for our defined set of domain knowledge through measuring the time required for modeling a concept space.

### 3.2 Evaluating the Cognitive Complexity of the Concept Spaces

The difficulty level for each of the defined concepts spaces was measured by recording the time which is required to perform the modeling task with Protégé. In total, four male participants without any visual, physical, or color blind impairment were involved in this evaluation. The participation in the pretest was voluntary and the users age was restricted to the range of $33 \pm 6$ years. This restriction assures that human motor performance is not effecting the modeling completion time. All participants had profound experience with ontology modeling tools as they were affiliated with the field of Semantic Web, working as scientific researchers employed at Fraunhofer IAIS.

**Method:** The participants had to model the individual concept spaces which were presented to them in a tabular format (cf. Appendix A). The University Space was used as a training example, thus created by all the participants in their first modeling task. The purpose of the training example was to make them comfortable using Protégé and allowing them to configure the tool to their needs. The remaining modeling tasks were performed in an alternating order as shown in the Table 1. This alternation was applied in order to avoid carry-over effects during the modeling task with the passage of time. The completion time for each of the modeling task was recorded in seconds and rounded off to the next smaller integer. All participants performed the experiment using a standard English (US) keyboard layout and an external mouse. The screen size was also kept same to 16"9 inches with a resolution of $1920 \times 1080$ pixels.

**Results:** The completion times for the individual concept spaces are presented in the Table 1. Additionally, the average completion times are also shown in this table. The mean difference between each concept space were calculated and evaluated. The results of this pretest indicate that the modeling task of the four concept spaces have on average the same completion time. The mean difference time between Family Tree Space and City Traffic Space was 7.5 seconds, between Family Tree Space and Media Space was 85.75 seconds, between Family Space and Zoo space was 17 seconds, between City Traffic Space and Media Space was 78.25 seconds and between City Traffic Space and Zoo Space was 9.5 seconds. The qualitative findings from the pretest are:

1. During the modeling, two participants have crossed out the concepts in the table.
2. In general we have noticed that the participants modeled classes, subclass hierarchies, and datatype properties in a similar fashion.
3. The participants varied in the way they have modeled object properties.

Table 1: Modeling completion times and the varying order of concept spaces.

| Participant | Modeling Completion Times | | | | Order of Concept Spaces |
|---|---|---|---|---|---|
| | **Family Tree** | **City Traffic** | **Media** | **Zoo** | |
| **A** | 237 | 302 | 349 | 362 | Zoo, City Traffic, Media, Family Tree |
| **B** | 330 | 428 | 429 | 403 | City Traffic, Zoo, Family Tree, Media |
| **C** | 389 | 183 | 361 | 270 | Family Tree, Media, City Traffic, Zoo |
| **D** | 343 | 416 | 503 | 332 | Family Tree, City Traffic, Media, Zoo |
| **Sum** | 1367 | 1329 | 1642 | 1299 | |
| **Mean** | 341.75 | 332.25 | 410.50 | 324.75 | |

## 4 Experimental Design

The evaluation design is based on the results we obtained from the pretest. We selected two concept spaces with the lowest mean difference between each other (i.e. *Family Tree Space* and *City Traffic Space*). In order to perform an empirical, comparative user study over visual modeling paradigm and hierarchical trees, participants were presented with the following nine tasks T1–T9:

T1: The participants had to fill out a demographic questionnaire, stating their *name, age, profession, experience in ontology modeling, experience with Protégé and WebVOWL, and any sort of visual, physical, or color blind impairment.*
T2: Using Protégé, the participants had to model an ontology for the Family Tree Space or the City Traffic Space.
T3: Based on the modeled concept space of the task T2, the participants had to fill out an After-Scenario Questionnaire (ASQ)[4] as a post task.
T4: As a cued recall process [1], the participants had to highlight the concepts in a $6 \times 4$ table which they thought they modeled using Protégé.
T5: Based on the modeled concept space of the task T2, the participants had to fill out a Computer System Usability Questionnaire (CSUQ)[5] as a post study task
T6: Using WebVOWL Editor, the participants had to model an ontology for the Family Tree Space or the City Traffic Space.
T7: Based on the modeled concept space of the task T6, the participants had to fill out an ASQ questionnaire as a post task.
T8: As a cued recall process, the participants had to highlight the concepts in a $6 \times 4$ table which they thought they modeled using WebVOWL Editor.
T9: Based on the modeled concept space of the task T6, the participants had to fill out a CSUQ questionnaire as a post task.

### 4.1 Participants

The user study comprised of 12 voluntary participants. Based on the answers in the demographic questionnaire, the participants were divided into two groups of experienced and non-experienced participants ($P_{G1}$ and $P_{G2}$). The first user group $P_{G1}$ contained six participants who had experience with ontology modeling. The second user group $P_{G2}$ contained six participants without prior experience in the ontology modeling domain. All the participants were male. The age of the participants was in the range of 25–36 years. In order to ensure that the participants human motor performance does not vary too much among the participants, the sample size was restricted to the age range of 31 $\pm$ 6 years. None of the participants had any sort of visual or physical impairment. One of the participants was color blind. The participants included employees and students of Fraunhofer IAIS, University of Bonn, and RWTH Aachen. All participants had a profound background in computer science as they were masters students, PhD students, or post docs in the field of computer science.

[4] http://garyperlman.com/quest/quest.cgi?form=ASQ
[5] http://garyperlman.com/quest/quest.cgi?form=CSUQ

### 4.2 Setup

In order to provide a homogeneous evaluation setup, all experiments were performed on a Dell Precision 3520 laptop with a standard English (US) keyboard layout and a screen size of 16"9 inches having a resolution of $1920 \times 1080$ pixels. An external mouse was used for navigation. The experiments were performed using Protégé (5.2.0) running on Ubuntu 18.04 and WebVOWL Editor using Mozilla Firefox or Google Chrome web browser. The study was conducted at the daily working environment of the participants.

### 4.3 Procedure

The experiments were always supervised by the same person and performed using the setup that was provided by the conductor of the evaluation. All participants were given approximately ten minutes of training on both tools. In the training sessions *Media Space* and *Zoo Space* were used. These were selected due to significantly larger mean differences (cf. Section 3.2), meaning different difficulty levels. The remaining two concept spaces *Family Tree Space* and *City Traffic Space* were used in the actual ontology modeling tasks of the user study. The results of our pretest indicate that these had significantly closer mean differences, meaning approximately same difficulty levels.

All participants started the user study by answering the demographic questionnaire. We categorise the remaining eight tasks into two groups, $T_{G1}$ and $T_{G_2}$. Tasks T2–T5 are related to Protégé ($T_{G1}$), whereas the tasks T6–T9 refer to WebVOWL Editor ($T_{G2}$). After finishing the demographic questionnaire, each participant was asked to perform the tasks corresponding to one group first and then continue with the other group. The completion time for the modeling task was recorded in seconds and rounded off to the next smaller integer. As post study questionnaires we have chosen ASQ and CSUQ because of their high global reliability degree [4]. The ASQ measures *ease of task completion, satisfaction with completion time*, and *support of information*. The CSUQ comprised of 19 questions, measuring *effectiveness, efficiency*, and *satisfaction* based on the ISO-9421-11 criteria [4]. Additionally, it measures *discriminability* based on the ISO/WD-9421-112 criteria [4]. *Guidance, workload*, and *error management* are measured w.r.t the Scapin and Bastien criteria [15]. Both questionnaires were answered using a Likert scale of 1 to 7, where 1 refers to strong disagreement and 7 refers to strong agreement.

The 12 participants were divided into two groups ($U_1$ and $U_2$), the first group containing three experienced and four non-experienced participants and the second group containing two non experienced and three experienced participants. This in-balanced assignment is a result of two invalid participations in the user group ($U_2$). However, the concept spaces are still counterbalanced as illustrated in the Table 2. The first user group ($U_1$) was asked to perform the Protégé specific tasks ($T_{G1}$) first and then continue with WebVOWL Editor specific tasks ($T_{G2}$). The second user group ($U_2$) was asked to perform the group tasks in an inverse order. This was done in order to avoid increasing or decreasing performance measures with the passage of time. The exact order of tool specific tasks and the order of the concept spaces is shown in the Table 2. The duration of the experiments for each participant was approximately 45–60 minutes.

Table 2: Order of tools and concept spaces presented to the participants.

| Participant | Tool A | (Concept Space) | Tool B | (Concept Space) |
|---|---|---|---|---|
| 1 | Protégé | (Family Tree) | WebVOWL Editor | (City Traffic) |
| 2 | Protégé | (City Traffic) | WebVOWL Editor | (Family Tree) |
| 3 | Protégé | (City Traffic) | WebVOWL Editor | (Family Tree) |
| 4 | WebVOWL Editor | (Family Tree) | Protégé | (City Traffic) |
| 5 | WebVOWL Editor | (City Traffic) | Protégé | (Family Tree) |
| 6 | Protégé | (Family Tree) | WebVOWL Editor | (City Traffic) |
| 7 | Protégé | (Family Tree) | WebVOWL Editor | (City Traffic) |
| 8 | WebVOWL Editor | (City Traffic) | Protégé | (Family Tree) |
| 9 | WebVOWL Editor | (City Traffic) | Protégé | (Family Tree) |
| 10 | Protégé | (Family Tree) | WebVOWL Editor | (City Traffic) |
| 11 | WebVOWL Editor | (Family Tree) | Protégé | (City Traffic) |
| 12 | Protégé | (City Traffic) | WebVOWL Editor | (Family Tree) |

## 5 Results and Discussion

The results of our user study comprise of 1) performance scores for the ontology modeling tasks, 2) scores for the participants recalling the concepts of the modeled ontology, and 3) the scores for user satisfaction for the two ontology modeling tools. The study manual, definition of concept spaces, and the evaluation data are available on GitHub[6].

### 5.1 Performance Scores for Ontology Modeling

The performance scores for tasks T2 and T6 were calculated based on the required time to model an ontology (cf. Section 4). The completion times are illustrated in Table 3 and indicate that WebVOWL Editor performed better in comparison to Protégé. On average, all 12 participants completed the ontology modeling task 18.7 seconds faster using WebVOWL Editor. The experienced ($P_{G1}$) and non-experienced ($P_{G2}$) participants performed respectively 26.2 and 11.4 seconds faster using WebVOWL Editor. The average difference between the completion times for the individual tools $M_{avg} = \frac{1}{12}\sum_{i=1}^{12} T2(i) - T6(i)$ was 23.4 seconds. Where $T2(i)$ and $T6(i)$ denote the time participant $i$ required to model an ontology for the individual task respectively. The standard deviation of differences was 80.54 seconds.

We used a paired t-test calculator[7] to analyze the results. The Shapiro-Wilk test ($\alpha = 0.05$) was used for normality validation. The normality p-value resulted as $0.0610$, thus, signifying that the required modeling completion time was normally distributed. For the paired t-test, the test statistic $t$ was $1.007164$ and the p-value was $0.335510$. As the p-value is larger than $\alpha$, it implies that the difference between the population means was not statistically significant. Consequently, users had a similar performance time using WebVOWL Editor and Protégé. The modeling completion times for all participants and grouped based on experience are illustrated in Figure 1a) and 1b) respectively.

[6] `https://github.com/RohanAsmat/VisualOntologyModelingEvaluationData`

[7] Paired t-test calculator can be accessed online at `http://www.statskingdom.com/160MeanT2pair.html`

Table 3: Average time required to model an ontology for both tools.

| | Mean Scores | | Standard Deviation | |
|---|---|---|---|---|
| **Participant Type** | **Protégé** | **WebVOWL Editor** | **Protégé** | **WebVOWL Editor** |
| All Participants (12) | 386.5 | 367.8 | 89.84 | 149.06 |
| Experienced Participants (6) | 364.5 | 338.3 | 105.00 | 136.76 |
| Non Experienced Participants (6) | 408.5 | 397.1 | 74.63 | 167.64 |

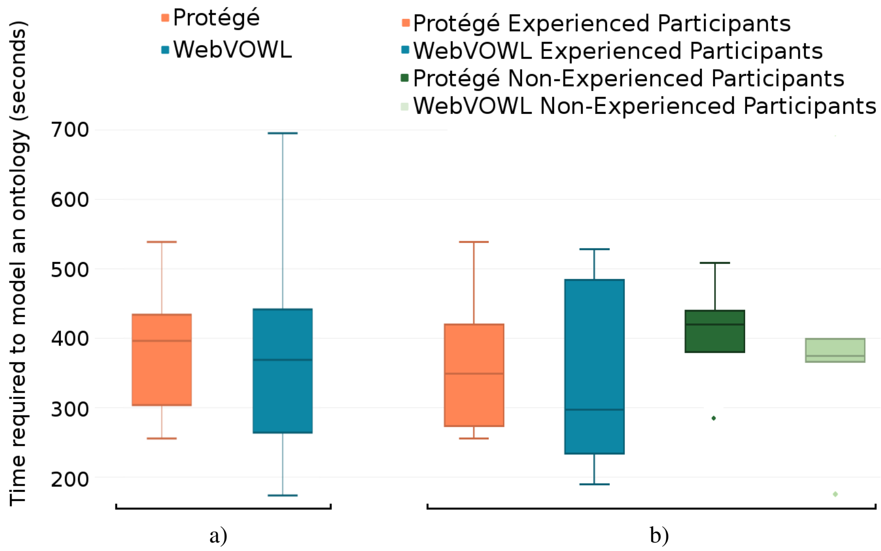


Fig. 1: Required modeling time as a box plot diagram. a) Modeling time for all participants. b) Modeling time for different participant groups based on experience.

Figure 1a) indicates that the Protégé had less variance, whereas WebVOWL Editor had more variance in results. For two participants the experiment was repeated, as their modeling task was interrupted, thus the outliers shown in Figure 1b) can be a result of the experiment repetition. Additionally, Figure 1b) indicates that the results for experienced participants had a higher variance with a wider spread of the central box while using WebVOWL Editor and Protégé, that is 250 and 146 seconds. In contrast, for non-experienced participants the spread of the central box for WebVOWL Editor and Protégé is 60 and 33 seconds respectively. Therefore, we can infer that the wider spread in case of experienced participants is due to their diversified experience of using the tools. For the non-experienced participants, a much lower spread denotes that the performance of participants was similar. A lower central box for non-experienced participants while performing on WebVOWL Editor than Protégé, reveals that users with no prior experience tend to perform much better using WebVOWL Editor than Protégé.

### 5.2 Cued Recall Scores

The cued recall scores were measured by the number of concepts that were correctly highlighted for tasks T4 and T8 (cf. Section 4). While measuring the correctness, *is a* and *has* concepts were not considered. These were allowed to be used repetitively or not at all. Figure 2 shows the number of incorrectly highlighted concepts for the individual participants. In total, for each tool respectively, the number of incorrectly highlighted concepts was eight. With respect to highlighting concepts, seven participants were incorrect for task T4, where as five were incorrect for task T8. This results indicate that fewer participants were incorrect with WebVOWL Editor than with Protégé.

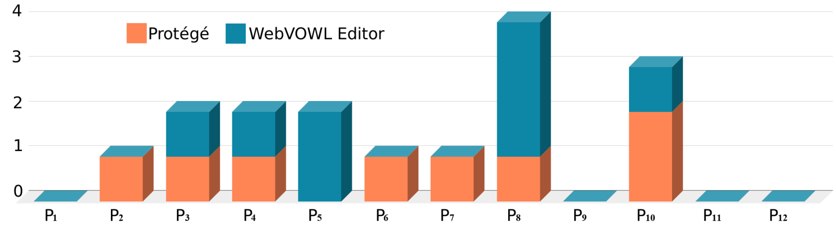


Fig. 2: Incorrectly highlighted concepts per participant ($P_i$) for the two tools.

### 5.3 User Satisfaction Scores

**ASQ —** Figure 3a) indicates that the participants were more satisfied with the *ease of completing the task* and the *time it takes to complete a task* when using WebVOWL Editor. The participants were equally satisfied in using the two tools for the *support information provided by the tool*. Figure 3b) indicates that the experienced group ($P_{G1}$) had a higher score for *ease of completing the task* and *time it takes to complete a task* using WebVOWL Editor, however, this result also indicate that the *support information provided by the tool* for WebVOWL Editor requires improvement. The results for the non-experienced group ($P_{G2}$) show that WebVOWL Editor was perceived requiring less *time to complete a task* and it provided better *support information*.

**CSUQ —** WebVOWL Editor scored better in 16 of 19 CSUQ questions. Protégé scored better in questions related to *number of system capabilities, information provided by the system*, and *if they can effectively complete their work using the system*. Protégé scored 5.4, 3.75, and 5.9, whereas WebVOWL Editor scored 4.9, 3.5, and 5.75 respectively. Based on the different participants groups ($P_{G1}$ and $P_{G2}$), the scores show that $P_{G2}$ still rated WebVOWL Editor better for *effectively completing their work* and for the *number of system capabilities* with a score of 5.3 and 5.5, whereas Protégé scored with 5.2 and 5.2. Six questions for which the results had significant difference between the two tools are shown in Figure 4. The CSUQ results indicate that both participant groups $P_{G1}$ and $P_{G2}$ rated WebVOWL Editor better in terms of usability.

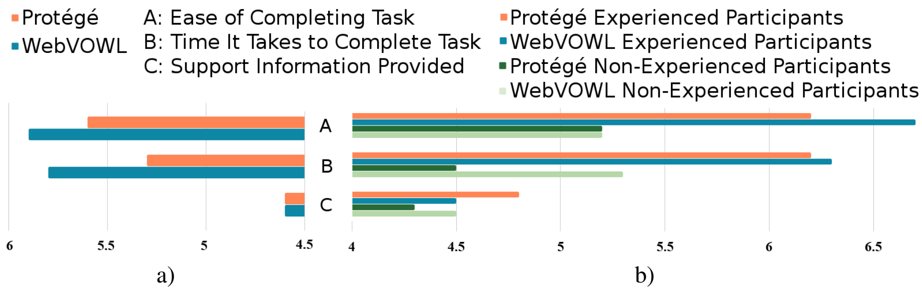


Fig. 3: ASQ: a) Scores for all participants. b) Scores for different participant groups.

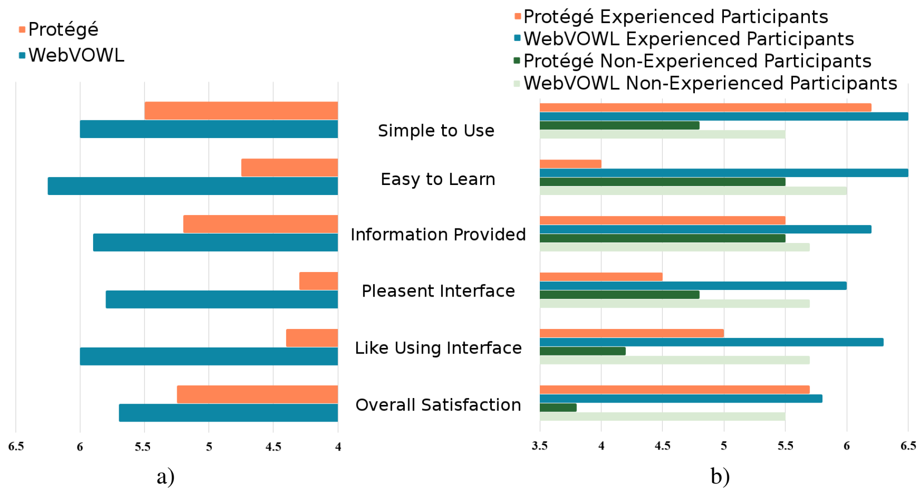


Fig. 4: CSUQ: a) Scores for all participants. b) Scores for different participant groups.

## 6 Conclusions

In this paper we compared a visual ontology modeling approach using WebVOWL Editor with a hierarchical tree, GUI-based modeling using Protégé. Visual ontology modeling approaches, particularly in the form of node-link diagrams, help non-expert users to get directly involved in ontology modeling without any prior experience. We introduced five small concept spaces (cf. Appendix A) and determined their cognitive complexity using a pretest. The results of the pretest indicate similar difficulty levels for *City Traffic Space* and *Family Tree Space*, thus these two were used in the ontology modeling tasks. Participants had to perform ontology modeling task, reflective question answering tasks and filled out additional ASQ and CSUQ post task questionnaires.

The results of the experiment (cf. Section 5) indicate that overall the participants were more efficient, they had a better understanding of the model, and they were more satisfied using WebVOWL Editor than Protégé. The mean performance measures for both tools had a minor difference with WebVOWL Editor having a better performance.

For the non-expert user group ($P_{G2}$) the performance was much better, highlighting a low learning curve with a good performance rate for novice users. For the expert user group ($P_{G1}$), the results were not significant and had high variance due to their prior experience with both tools as shown in the Figure 1b). In the following usability areas WebVOWL Editor scored better: *ease of task completion, time taken for task completion, ease of learning the system, simplicity of using the system, pleasantness of the interface, likeability to interface,* and *overall user satisfaction for the system*.

The VOWL notation is designed for a user-oriented representation of OWL ontologies for different user groups. WebVOWL Editor is designed for device independent ontology modeling and thus realizes minimalistic user interactions, allowing it to be used on touch devices. Visual modeling paradigms which allow for better mental map preservation, the VOWL notation, and the minimalistic user interactions are beneficial for the performance of WebVOWL Editor. However, due to the small sample size, the results indicate only a small increase in performance, thus we suggest a follow up study with an increased number of participants to atleast twenty as suggested by Nielsen [2], it improves the confidence interval and reduces the margin of error. We also purpose to control the prior experience with modeling tools, thus, reducing the variance and improving comparison of results between the two tools.

## Acknowledgments

This work has partly been funded by the EU project GRACeFUL (grant no. 640954) as well as a scholarship from the German National Library of Science and Technology (TIB). In addition, parts of it evolved in the context of the Fraunhofer Cluster of Excellence "Cognitive Internet Technologies".

## References

1. Memory recall/retrieval. `http://www.human-memory.net/processes_recall.html`, online Accessed: 2018-07-17
2. Quantitative studies: How many users to test? `https://www.nngroup.com/articles/quantitative-studies-how-many-users/`, online Accessed: 2018-07-17
3. Anikin, A., Litovkin, D., Kultsova, M., Sarkisova, E., Petrova, T.: Ontology visualization: Approaches and software tools for visual representation of large ontologies in learning. In: Proceedings of the 2nd Conference on Creativity in Intelligent Technologies and Data Science. pp. 133–149 (2017)
4. Assila, A., de Oliveira, K.M., Ezzedine, H.: Standardized usability questionnaires: Features and quality focus. electronic Journal of Computer Science and Information Technology **6**(1) (2016)
5. Dimitrova, V., Denaux, R., Hart, G., Dolbear, C., Holt, I., Cohn, A.G.: Involving domain experts in authoring OWL ontologies. In: The Semantic Web - ISWC 2008, 7th International Semantic Web Conference, ISWC 2008, Karlsruhe, Germany, October 26-30, 2008. Proceedings. pp. 1–16 (2008)
6. Duineveld, A.J., Stoter, R., Weiden, M.R., Kenepa, B., Benjamins, V.R.: Wondertools? A comparative study of ontological engineering tools. Int. J. Hum.-Comput. Stud. **52**(6), 1111–1133 (2000)

7. Falco, R., Gangemi, A., Peroni, S., Shotton, D.M., Vitali, F.: Modelling OWL ontologies with graffoo. In: The Semantic Web: ESWC 2014 Satellite Events - ESWC 2014 Satellite Events, Anissaras, Crete, Greece, May 25-29, 2014, Revised Selected Papers. pp. 320–325 (2014)
8. Falconer, S.: Ontograf. `http://protegewiki.stanford.edu/wiki/OntoGraf` (2010)
9. Fu, B., Noy, N.F., Storey, M.D.: Indented tree or graph? A usability study of ontology visualization techniques in the context of class mapping evaluation. In: The Semantic Web - ISWC 2013 - 12th International Semantic Web Conference, Sydney, NSW, Australia, October 21-25, 2013, Proceedings, Part I. pp. 117–134 (2013)
10. Fu, B., Noy, N.F., Storey, M.D.: Eye tracking the user experience - an evaluation of ontology visualization techniques. Semantic Web **8**(1), 23–41 (2017)
11. Katifori, A., Torou, E., Halatsis, C., Lepouras, G., Vassilakis, C.: A comparative study of four ontology visualization techniques in protege: Experiment setup and preliminary results. In: 10th International Conference on Information Visualisation, IV 2006, 5-7 July 2006, London, UK. pp. 417–423 (2006)
12. Lohmann, S., Negru, S., Haag, F., Ertl, T.: Visualizing ontologies with VOWL. Semantic Web **7**(4), 399–419 (2016)
13. Navarro, J.F.G., García-Peñalvo, F.J., Therón, R., de Pablos, P.O.: Usability evaluation of a visual modelling tool for OWL ontologies. J. UCS **17**(9), 1299–1313 (2011)
14. Petersen, N., Similea, A., Lange, C., Lohmann, S.: TurtleEditor: A web-based RDF editor to support distributed ontology development on repository hosting platforms. Int. J. Sem. Comp. **11**(3), 311–324 (2017)
15. Scapin, D.L., Bastien, J.M.C.: Ergonomic criteria for evaluating the ergonomic quality of interactive systems. Behaviour & IT **16**(4-5), 220–231 (1997)
16. Stanford: Protégé. `https://protege.stanford.edu/`
17. TopQuadrant: Topbraid composer(tm). `https://www.topquadrant.com/tools/modeling-topbraid-composer-standard-edition/`

## Appendix A Concept Spaces for the Study

The five concept spaces that were defined for the pretest and used in the study are shown in the Figure 5. The concepts indicated with * could be used zero or multiple times.

| University Space | Zoo Space | City Traffic Space | Media Space | Family Space |
|---|---|---|---|---|
| *Staff Member* | *Animal* | *Vehicle* | *Media Network* | *Child* |
| *Person* | *Herbivores* | *Bus* | *News Channel* | *childs birthplace* |
| *Professor* | *Snow Leopard* | *model name* | *Daily News Channel* | *Family Tree* |
| *University* | *eats* | *Manufacturer* | *channel name* | *Female* |
| *Student* | *Carnivores* | *City Traffic* | *Sports Channel* | *Mother* |
| *Graduate Student* | *Cheetah* | *Public Vehicle* | *shows* | *Grandmother* |
| *has name* | *Zoo* | *Car* | *Local Sports Channel* | *Male* |
| *teaches* | *Goat* | *manufactured by* | *News Program* | *Gives Birth* |
| *Course* | *Grass* | *Private Vehicle* | *Channel* | *Father* |
| *Undergraduate Student* | *grass type* | *Train* | *International Sports Channel* | *Person* |
| *course name* | *has legs* | *manufacturing date* | *Airing Time* | *person name* |
| *has** | *has** | *has** | *has** | *has** |
| *Is a** | *is a** | *is a** | *is a** | *is a** |

Fig. 5: Classes and properties defined for each concept space respectively.