Indented Tree or Graph? A Usability Study of Ontology Visualization Techniques in the Context of Class Mapping Evaluation

Bo Fu[1], Natalya F. Noy[2], and Margaret-Anne Storey[1]

[1] Department of Computer Science, University of Victoria, BC, Canada
[2] Stanford Center for Biomedical Informatics Research, Stanford University, CA, US
{bofu,mstorey}@uvic.ca, noy@stanford.edu

Abstract. Research effort in ontology visualization has largely focused on developing new visualization techniques. At the same time, researchers have paid less attention to investigating the usability of common visualization techniques that many practitioners regularly use to visualize ontological data. In this paper, we focus on two popular ontology visualization techniques: indented tree and graph. We conduct a controlled usability study with an emphasis on the effectiveness, efficiency, workload and satisfaction of these visualization techniques in the context of assisting users during evaluation of ontology mappings. Findings from this study have revealed both strengths and weaknesses of each visualization technique. In particular, while the indented tree visualization is more organized and familiar to novice users, subjects found the graph visualization to be more controllable and intuitive without visual redundancy, particularly for ontologies with multiple inheritance.

Keywords: Ontology visualization, indented tree, graph, usability study.

1 Introduction

Information visualization (InfoVis) is a well-established research field. The goal of InfoVis is to transform information into visual representations that enable viewers to offload cognition to their perceptual systems in the process of better observing and understanding the information at hand. On the semantic web, researchers have applied visualization techniques to a range of topics such as semantic search [1], linked open data [2] and most notably, ontology design and management [3]. In recent years, ontology visualization has attracted much attention from the research community with a focus on providing the necessary support to enable users to create new and browse existing ontological resources. This research trend is reflected in the various visualization plugins developed for the Protégé[1] ontology editor [4], and visual support designed for querying and browsing ontology libraries [5, 6].

A commonly used technique in ontology visualization is indented tree where indentation is used to illustrate super/sub-class relationships and there is one path and

---

[1] http://protege.stanford.edu

one path only between any pair of nodes. Another observation from the literature is that several ontology visualization tools have built upon node–link diagrams (i.e., graphs), which are essentially nodes with connecting edges that illustrate ontological entities and the relationships that exist among them. While researchers have devoted significant effort to develop new tools and techniques, they have paid less attention to investigating the usability of existing ontology visualization techniques that many practitioners already use on a regular basis.

Motivated by this research opportunity, we evaluated two frequently used ontology visualization techniques in the state of the art: indented tree and graph visualization. The goal of our study is to investigate the effectiveness and efficiency of the support that these visualization techniques provide. Specifically, we are interested in comparing their support to users during manual mapping evaluation tasks. We used a controlled experimental approach and present quantitative and qualitative analysis of the usability issues associated with these visualization techniques. The results from this research have uncovered useful information on the suitability of these visualization techniques in knowledge representation and mapping evaluation. In particular, we identified and highlighted perceived benefits and drawbacks of these techniques.

2   Related Work

In recent years, researchers have developed a variety of techniques to visualize ontologies. In this section, we present a brief overview on notable advances in this area. For extended discussions and classifications, see [3, 7].

Ontology development is one key activity that routinely relies on visualization. Visualizations assist users monitoring changes during ontology evolution [8], provide alternative development platforms by enabling UML-based editing [9] and rule-based authoring[2] of ontologies. Ontology editors such as Protégé, WebProtégé [10], OBO-Edit[3] and structOntology[4]; ontology browsers such as VectorBase[5]; ontology libraries such as BioPortal[6]; as well as ontology mapping tools such as OntoLink[7] all use indented tree visualization to present hierarchical structures that are typically associated with ontological entities. Others have applied treemaps [11] in ontology visualization to make use of all available screen space and to maximize the information displayed. Plaisant and colleagues [12] explored SpaceTree for ontology visualization, which extends the conventional node–link diagrams with dynamic rescaling to utilize screen space. Parsia and colleagues [13] proposed CropCircle, which illustrates parent–child and sibling relationships simultaneously. Protégé visualization plugins such as

---

[2] http://oogis.ru/component/option,com_remository/
  Itemid,34/func,fileinfo/id,15/lang,en see DroolsTab
[3] http://oboedit.org
[4] http://openstructs.org/structontology
[5] https://www.vectorbase.org/content/ontology-browser
[6] http://bioportal.bioontology.org
[7] http://www.mindswap.org/2004/OntoLink

OwlViz[8], NavigOWL[9], TGVizTab [19] and OWLPropViz[10] use graphs to illustrate classes and relationships in ontologies. Other web-based tools using similar node–link diagrams to visualize ontologies include FlexViz [20], BioMixer [21] and OLSVis [6]. In addition, 3D techniques have been applied to add more space on the screen by introducing a third dimension to node–link diagrams, such as OntoSphere [14]. Other research has focused on reducing information overload in node–link visualizations by presenting only classes above a calculated importance score [15], while several authors [16, 17, 18] have argued for the benefits of multiple visualizations with the goal of adapting to user preference and style.

The vast majority of these tools and techniques use indented tree or graph visualizations. Researchers in the field of InfoVis have extensively studied both techniques [22, 23] and proposed a range of evaluation approaches depending on the stage of the visualization software [24], including empirical studies [25] and insight-based methodologies [26]. In contrast, evaluation of visualization techniques in the context of ontology-focused tasks has been limited. Existing studies have compared Protégé plugins [27] and visualizations with built-in query support [28], focusing on evaluating their ability to support users seeking specific ontological information through controlled experiments. This paper aims to fill this research gap by presenting a comparative usability study of the commonly used indented tree and graph visualizations focusing on how well they illustrate ontological semantics.

3　Usability Study Overview

The goal of our usability study is to investigate the extent to which indented trees and graphs can support users in the process of understanding the semantics in ontologies.

Specifically, we asked the study participants to evaluate a given set of mappings between pairs of ontologies by interacting with the visualizations of these ontologies. To evaluate a mapping successfully, a participant must understand the semantics of the mapped entities in their respective ontologies and must use this knowledge to determine whether a mapping relation exists. Hence this task setup ensures the study focuses on examining the interactions between the participants and visualizations. Note that we did not explicitly specify that the participant must generate an overview of each ontology as we believe exploratory activities are inevitable in the given tasks. To generate a mapping correctly or to identify an incorrect one, the user typically must understand the semantics of the entities in their respective ontologies. This understanding is often a result of exploring the semantics and gaining an overview of the structures.

Fig. 1. Sample Task Screen. In this example, graphs are used to visualize two biomedical ontologies. Mappings to be evaluated are presented in a spreadsheet. Interacting with the visualizations, participants must use drop-down lists containing either "yes" or "no" responses to evaluate the correctness of existing mappings (in part 1) and add missing mappings by typing class names (into part 2 of the spreadsheet).

3.1   Tasks

We presented the participants with a set of mappings and asked them to identify correct and incorrect mappings as well as add missing mappings. The participants were assisted by visualizations of the ontology pair. Participants essentially engaged in two types of activities: identification activities and creation activities. The former involves the identification of correct and incorrect results among an existing set of mappings (i.e., determining the correctness of the given mappings). The latter involves the creation of new mappings that are absent from the existing set (i.e., determining the completeness of the given mappings). Figure 1 shows an example of what a participant saw on her (or his) screen.

3.2   Datasets

We used two pairs of ontologies, each accompanied by a set of mapping standards, taken from the Ontology Alignment Evaluation Initiative (OAEI) 2012 conference[11]

---

[11] http://oaei.ontologymatching.org/2012/conference/index.html

and the BioMed[12] tracks. Table 1 presents an overview of the ontologies used in this study.[13]

The conference ontologies describe the organization of conferences with a total of 74 and 100 classes respectively, at most 3 or 6 classes on the longest path to root, at most 8 or 9 subclasses for a class, without any multiple inheritance. The conference task represents a less difficult scenario, where the ontologies involved have fewer classes, the number of subclasses per class is fewer and the paths to root are shorter.

The BioMed task involves ontologies describing concepts in the organism domain. We reduced the size of the original ontologies and gold standards. In our study, the BioMed ontologies have a total of 89 and 181 entities respectively, at most 11 or 12 classes on the longest path to root, at most 6 or 10 subclasses for a class, with at most 4 occurrences of multiple inheritance. The BioMed task illustrates a more difficult scenario as the ontologies contain more entities, the number of children per entity is increased, the paths to root are longer and multiple inheritance is present.

Based on the OAEI gold standards, for each ontology pair, we randomly removed correct mappings from its gold standard and added incorrect mappings in order to create two mapping sets to present to the participants. The conference task and the BioMed task both required the participants to identify 13 correct results, 3 incorrect results and add 7 missing mappings in each task scenario. This setup thus ensures that the study outcome (in particular, time on task) is not affected by the number of mappings to be evaluated, but rather a result of ontology and visualization complexity.

Table 1. Characteristics of the Ontologies Used in the Study

|	Conference Ontologies		BioMed Ontologies	
	O_1	O_2	O_3	O_4
Class Count	38	77	89	181
Object Property Count	13	33	-	-
Data Type Property Count	23	-	-	-
Multiple Inheritance Occurrences	-	-	2	4

3.3   Visualization Support

We presented indented tree visualizations to the participants by loading ontologies into Protégé and asked participants to interact with the trees but not with any other features in Protégé. We implemented graph visualizations[14] in force directed layouts

---

[12] `http://www.cs.ox.ac.uk/isg/projects/SEALS/oaei/2012/`

[13] The ontologies and gold standard used in this study can be found at the prefix
`http://webhome.csc.uvic.ca/~bofu/study/datasets/`
followed by file name: `o1.owl, o2.owl, o3.owl, o4.owl,`
`o1-o2%20gold%20standard.rdf or o3-o4%20gold%20standard.rdf`

[14] The graph visualization of the ontologies used in the study can be found at the prefix
`http://webhome.csc.uvic.ca/~bofu/study/` followed by file
name `o1.html, o2.html, o3.html, or o4.html`
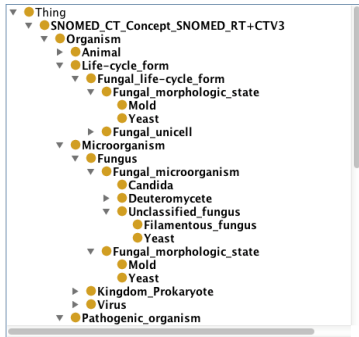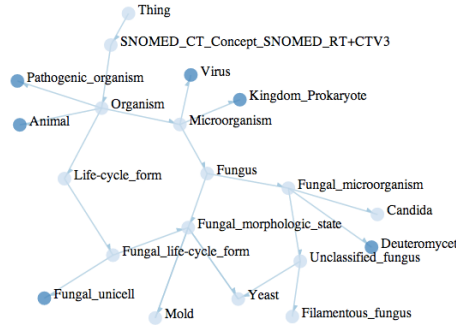
(a) Indented Tree Visualization   (b) Graph Visualization

Fig. 2. Visualization Techniques Investigated in the Study

using the D3 JavaScript library[15]. This implementation is representative of current graph techniques as it is composed of nodes and connecting edges, which are key characteristics of graphs as shown from the literature review.

Figure 2 presents visualization snippets of the SNOMED ontology using indented tree (Figure 2-a) and graph (Figure 2-b). In the indented tree visualization, *is-a* relationships are illustrated by indentation and the expanders allow users to toggle children of a node. Participants can use horizontal and vertical scroll bars to adjust the viewing area. In the graph visualization, classes are illustrated by vertices and *is-a* relationships are illustrated by directional edges with arrowheads pointing to the subclasses. The coloring of the vertices denotes whether a node is expandable (i.e., dark-colored vertices illustrate the existence of subclasses whereas light-colored vertices illustrate nonexpandable vertices). Clickable vertices allow users to toggle children of a particular node. In addition to using scroll bars to adjust the viewing area, the graph visualization is also editable: users can customize and manipulate the visualization by dragging and dropping nodes to any location on the screen. In both visualization techniques, we presented only the ontology root initially and participants must expand the root to view other classes.

3.4   Participants

We recruited volunteers via engineering departmental mailing lists at the University of Victoria. Each participant who successfully completed a study session received a $20 gift certificate. A total of 36 participants took part in our study. The participants were undergraduate and graduate students enrolled in disciplines including computer science, biomedical, biochemistry, and mechanical, electrical, software engineering. All participants were novice users of semantic technologies and they were new to ontologies and ontology mapping. As users of ontologies and visualizations

---

[15] http://d3js.org

increasingly include people with little knowledge of semantic technologies (e.g., Bio-Portal users are clinical and biomedical researchers who are new to ontologies and mappings), we were interested in studying the visualization support for novices. This group of users is of particular interest to us since a true expert should be able to successfully and accurately evaluate mappings regardless of the tool support. We do, however, recognize the opportunity to include expert users in future studies (discussed in section 6).

3.5   Protocol

We carried out one-on-one sessions with participants, where a session lasted approximately two hours. In each study session, we asked the participant to first complete an online tutorial on ontologies and ontology mapping.[16] The participant was then given instructions on the goal of her (or his) tasks: evaluate a set of exact mappings between a pair of ontologies.[17] We asked each participant to complete two tasks. Each task involved one ontology pair and one type of visualization. Each participant was asked to complete a video tutorial on how to interact with a given visualization before they began a task. We varied the ordering of the ontologies and visualization support between participants. For example, in one session, we asked Alice to complete the conference task using the indented tree, and then we asked her to complete the BioMed task using the graph. In another session, we asked Bob to complete the BioMed task using the indented tree, and then we asked him to complete the conference task using the graph. We randomly assigned tasks, ensuring equal distribution of tasks in the population as well as counterbalancing the order of tasks overall. This protocol ensured that a participant did not become overly familiar with a particular visualization, nor did the participant learn about the domain of interest over time, thus minimizing the impact of task order on the study outcome. Our protocol ensures that the only independent variable in the experiment is the visualization type, since we are interested in how two visualizations differ in their support to the same user group conducting the same set of tasks. However, we recognize a potential research opportunity to compare behaviors of different user groups in the future (discussed in section 6).

3.6   Metrics

To investigate the extent to which the indented tree and graph visualization can assist novice users in evaluating mappings *effectively* and *efficiently*, we measured task success and time on task as follows.

We calculated success scores for a participant to reflect *identification* success (i.e., the activity focusing on evaluating the correctness), *creation* success (i.e., the activity focusing on evaluating the completeness) and *overall* success (i.e., combing both type

---

[16] Materials used in this study can be found at
   https://sites.google.com/site/uvicstudy
[17] Since the evaluation process does not differ regardless of the type of entity or mapping relation, it is thus sufficient to use exact mappings as examples for the purpose of this study.

of activities). For example, suppose a task presents a set of existing mappings between ontology O and O', among which are n_1 number of correct mappings, n_2 number of incorrect mappings and n_3 number of missing mappings. If a participant successfully identifies x number of correct mappings and y number of incorrect mappings, then her (or his) identification success = $(x+y)/(n_1+n_2)$. If a participant correctly creates z number of new mappings, then her (or his) creation success = z/n_3. Her (or his) overall success = $(x+y+z)/(n_1+n_2+n_3)$. Her (or his) error rate is recorded as the number of incorrect answers divided by her (or his) total number of answers. Success scores range between 0 and 1; the higher the score the more successful the participant was at the task. Error rates also range between 0 and 1; the lower it is, the fewer mistakes the participant made in the task.

We asked participants to raise any questions before they began a task as we did not allow any interactions during the task. This restriction ensured a clear end state in the tasks, whereby time on task is the length of time it took a participant to complete the spreadsheet (which included both identification and creation activities).

3.7   Participant Feedback

After each task, we collected user feedback through computerized surveys based on the NASA-task load index (NASA-TLX) [29], the System Usability Scale (SUS) [30], the Usefulness, Satisfaction and Ease of Use (USE) questionnaire [31], and reaction cards [32]. We used 7-point Likert scales for all questionnaires.

Workload is "the cost of accomplishing mission requirements for the human operator" [33]. The NASA-TLX is specifically designed to measure workload through six dimensions, namely mental demand, physical demand, temporal demand, effort, performance and frustration level. Each dimension is measured through a question that asks the participant to rate the demand level on scales with endpoints being low-high and poor-good. In this study, we used raw NASA-TLX [34], which eliminates weightings between paired dimensions. Raw NASA-TLX is shown to be of no particular accuracy loss compared to the original, weighting NASA-TLX [33]. For each participant, we calculated a single workload score by averaging normalized scores of the six dimensions. The workload rating for a dimension ranges between 0 and n-1 given an n-point Likert scale. The workload score for a dimension is calculated as $(n-1)\times100/6$. The overall workload is the mean of six ratings. The workload rating ranges between 0 (low workload) and 100 (high workload).

The SUS is a questionnaire that contains 10 statements collecting feedback on agreement scales. Five statements are positively worded and the other five are negatively worded. Example statements include "I thought the visualization was easy to use" and "I found the visualization unnecessarily complex". Using an n-point Likert scale, the score contribution for a statement ranges between 0 and n-1. For a positively worded statement, the score contribution is the scale position minus 1. For a negatively worded statement, the score contribution is n minus the scale position. Multiply the sum of ten score contributions by $10/(n-1)$ to obtain the overall usability score. An aggregated usability score can be calculated for a visualization, which ranges between 0 and 100. The higher it is, the more usable the visualization.

In addition to SUS, we used the USE questionnaire, which expresses usability in four dimensions: usefulness, ease of use, ease of learning and satisfaction. We collected levels of user agreement to 30 statements. Our goals was to gain a further understanding in the variations between the two visualization techniques by breaking usability down to four dimensions. We calculated a mean rating that ranges between 0 and 6 to indicate the average rating for each of the four usability dimensions.

Finally, we presented 118 reaction cards containing adjectives (e.g., "engaging", "powerful", "rigid", "dated", etc.) to participants after the completion of each task. We asked the participants to pick out top five cards that best described the specific visualization in the given task and explain their choices. This technique aimed to elicit commentary and collect qualitative feedback.

4   Findings

We present the results of the measures discussed in section 3.6 and 3.7 below.

4.1   Effectiveness

Figure 3 presents mean overall success. Table 2 presents further details on the various success scores. In the conference task, the user group that was assisted by graphs yielded a slightly higher mean overall success score. Both visualization techniques generated the same median overall success score. We carried out independent sample t tests (with an alpha level equal to 0.05) with the null hypothesis being there is no difference between the two user groups. P-values from these independent t tests indicate that there is no significant difference between the user groups.

(a) Conference Task   (b) BioMed Task
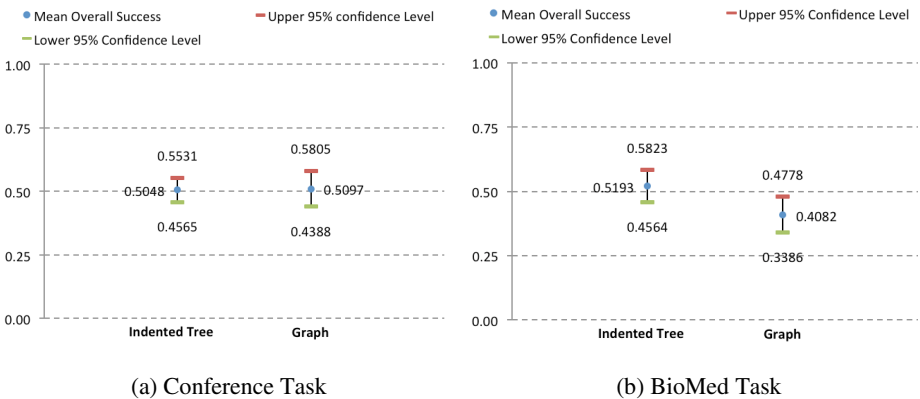
Fig. 3. Visualization Effectiveness. The vertical axis illustrates mean overall success and the horizontal axis represents the user groups using different visualizations. Error bars show 95% confidence intervals, i.e., how far from the reported value the true (error free) value might be.

In the BioMed task, the user group who used indented trees generated higher mean, median and lower standard deviation in identification and overall success scores. They also made fewer mistakes, as suggested by lower and less dispersed error rates. With the exception of creation success scores, p-values generated from all other scores are equal to or less than the alpha level. This finding suggests that there are significant differences between the identification score, overall success and error rates between the two user groups, indicating that indented trees were more effective.

Table 2. Task Success. Statistically significant results are bolded.

Visualization	Task Success	Conference Task			BioMed Task		
		Mean	Median	StDev	Mean	Median	StDev
Indented Tree	Identification	0.6458	0.6250	0.1134	**0.6944**	**0.6875**	**0.1514**
	Creation	0.1825	0.1429	0.2068	0.1190	0	0.1715
	Overall	0.5048	0.5000	0.1045	**0.5193**	**0.5000**	**0.1362**
	Error	0.3951	0.4045	0.1322	**0.3668**	**0.3640**	**0.1433**
Graph	Identification	0.6563	0.6563	0.1458	**0.5382**	**0.5625**	**0.2013**
	Creation	0.1746	0.0714	0.2525	0.1111	0.0714	0.1433
	Overall	0.5097	0.5000	0.1534	**0.4082**	**0.4130**	**0.1507**
	Error	0.3794	0.3787	0.1339	**0.4747**	**0.5147**	**0.1747**

4.2 Efficiency

Figure 4 presents an overview of the average time spent completing each task using different visualizations. Further details are shown in Table 3. It is consistently shown in both tasks that user groups assisted by the indented tree visualization were faster at completing their tasks than those who used graphs. However, p-values do not provide sufficient evidence to indicate a statistically significant difference between the user groups, suggesting comparable completion time regardless of the visualization used.

(a) Conference Task   (b) BioMed Task
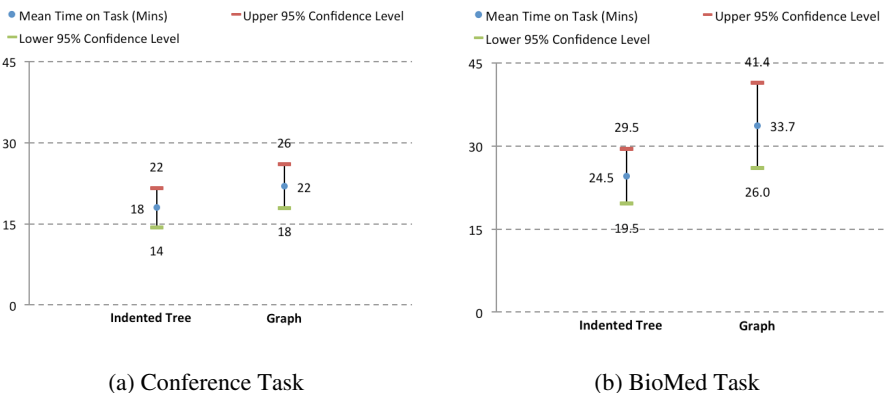
Fig. 4. Visualization Efficiency. The vertical axis represents mean time-on-task, and the horizontal axis illustrates the user group. Error bars show 95% confidence intervals.

Table 3. Time on Task

Visualization	Conference Task			BioMed Task		
	Mean	Median	StDev	Mean	Median	StDev
Indented Tree	17.9	15.5	7.8	24.5	25	10.7
Graph	21.9	21	8.9	33.7	30	16.6

4.3   Workload

Figure 5 presents an overview of the workload scores. Further details are presented in Table 4. Mean values indicate that the user group assisted by the graph visualization found the task more demanding that those who used the indented tree visualization in both tasks. In the conference task, the workload scores for graphs are particularly disperse, which consequently led to higher mean and median even though the most common rating is much lower (see mode) compared to the indented tree visualization. However, p-values in both tasks indicate that the differences between the two user groups are not statistically significant, i.e., there was no particular increase in workload regardless of the type of visualization used.

Table 4. Workload Scores

Visualization	Conference Task				BioMed Task			
	Mean	Median	Mode	StDev	Mean	Median	Mode	StDev
Indented Tree	39.97	41.67	50.00	13.31	52.47	52.78	52.78	12.31
Graph	47.99	45.83	30.56	18.17	57.87	56.97	61.11	10.05

(a) Conference Task   (b) BioMed Task
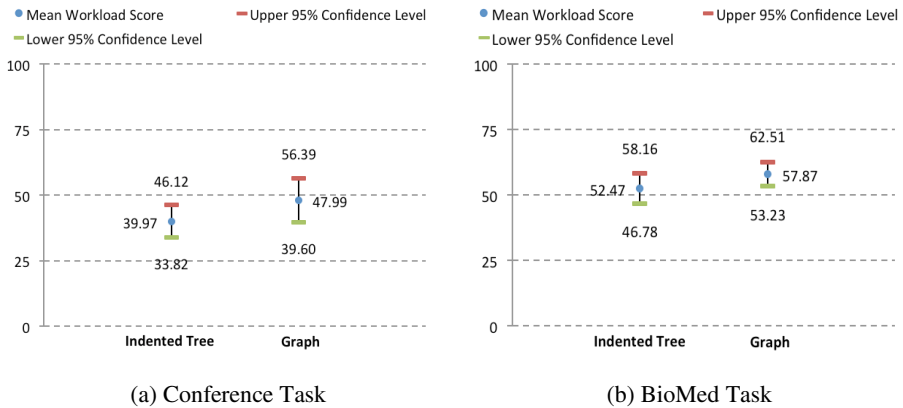
Fig. 5. Task Workload. The vertical axis represents mean workload scores, and the horizontal axis illustrates the user group. Error bars show 95% confidence intervals.

4.4   Usability and Qualitative Feedback

Figure 6 presents an overview of the SUS scores. Further details are shown in Table 5. In both tasks, the usability scores indicate that participants found the indented tree more usable than graph as the average, mid-point and most commonly occurred values are always higher. However, this difference is not statistically significant.

(a) Conference Task   (b) BioMed Task
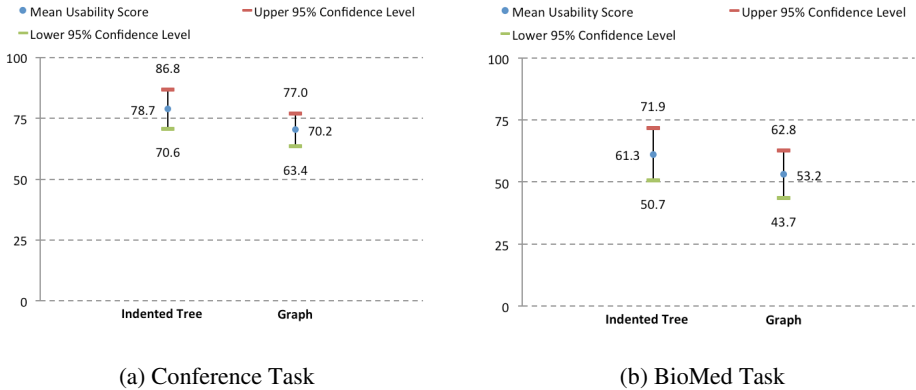
Fig. 6. Visualization Usability. The vertical axis represents mean usability score, and the horizontal axis illustrates the user group. Error bars show 95% confidence intervals.

Figure 7 shows USE ratings generated for the visualizations. In the conference task, similar ratings are generated for both visualization techniques with the indented tree having slightly higher mean and median ratings in all four dimensions. However, p-values suggest that the differences shown in this task are not statistically significant. In the BioMed task, there is a decrease in all ratings for both visualization techniques, although higher mean and median values are found in indented tree. P-values indicate a statistically significant difference between the two visualization techniques in terms of usefulness (note that statistical significance was not found in ease of use, ease of learning and satisfaction ratings). The results suggest that as the evaluation task becomes more difficult, visualization support appears to be less helpful regardless of the specific technique. Overall, the USE results indicate that all usability dimensions of the two visualization techniques are in fact very comparable.

Table 5. Usability Scores

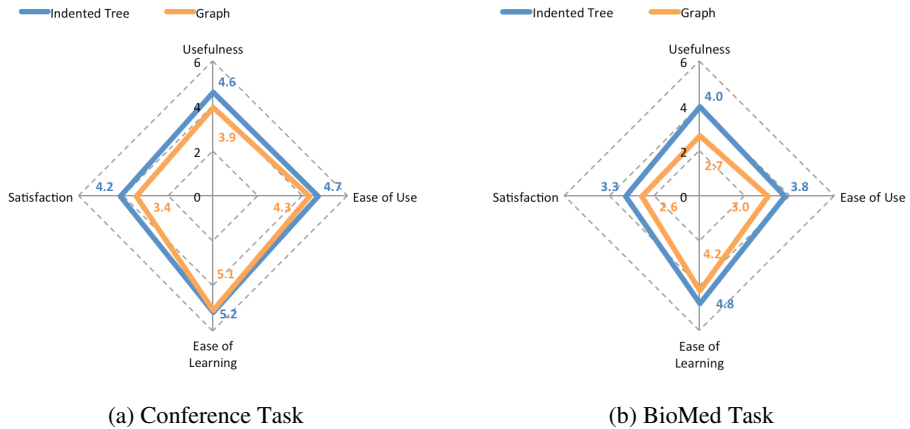| Visualization	Conference Task				BioMed Task			
	Mean	Median	Mode	StDev	Mean	Median	Mode	StDev
Indented Tree	78.70	80.00	71.67	17.54	61.30	60.00	70.00	22.96
Graph	70.19	73.33	50.00	14.72	53.24	50.83	50.00	20.67

Fig. 7. Visualization Usability Breakdown. Each axis presents a usability dimension (usefulness, ease of use, ease of learning and satisfaction). The radar chart presents a mean rating for each dimension.
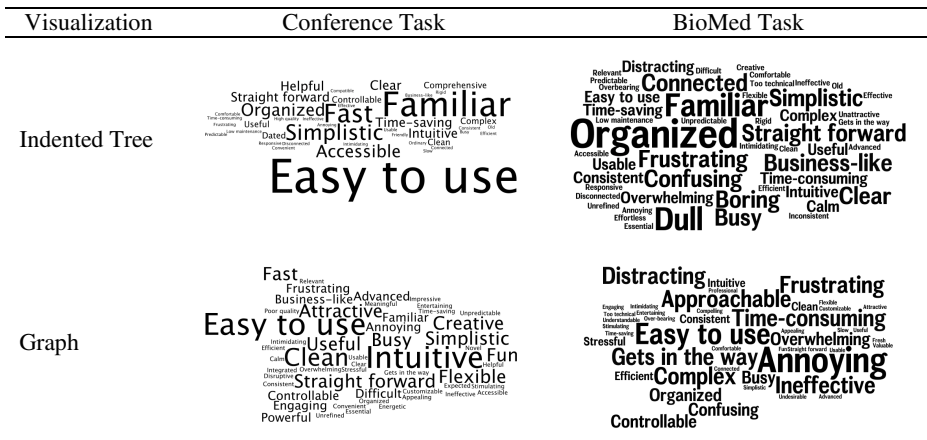
Visualization	Conference Task	BioMed Task
Indented Tree		
Graph		

Fig. 8. Reaction Card Responses. Font sizes illustrate the frequency of use for a particular card; the bigger the font, the more frequently the card was used to describe a visualization.

Figure 8 presents tag clouds of reaction card responses. In the conference task, participants found both visualization techniques easy to use. They found the indented tree *familiar* and the graph *intuitive*. In the case of the BioMed task, as the task becomes more difficult, more diverse reaction cards are used and an increased number of negative cards are present for both visualization techniques. For instance, the participants described both visualization techniques as *distracting*, *frustrating* and *confusing*. They characterized indented trees as *organized*, *straightforward* and *simplistic*, although *dull*, *boring* and *busy*. They found the graph visualization to be *approachable* and *controllable* in the conference task, however, it became *annoying* and *complex* in the BioMed task. Overall, participants consistently used *simplistic* to describe the

indented tree in both tasks. They also consistently used *easy to use* to describe the graph visualization in both tasks, although this phrase is used much less frequently to describe the indented tree in the BioMed task. Furthermore, several participants mentioned that they particularly liked how multiple inheritance is visualized in graphs.

5   Discussion

We present correlation results and key observations drawn from this study next.
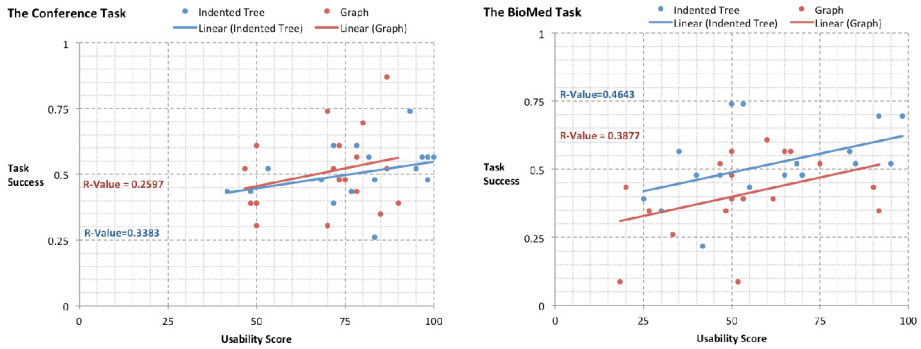
5.1   Correlation Tests

Given the range of variables (i.e., effectiveness, efficiency, workload, SUS and USE scores) associated with each visualization, we conducted correlation tests to determine whether dependable relationships exist. If a strong correlation coefficient exists between a variable pair, then knowing the value of one variable, we could predict the likely value of the other variable for a given visualization. The degree of correlation between two variables is represented by the R-value, which ranges between -1 and 1. The stronger the correlation, the closer the R-value is towards -1 (negative correlation) or 1 (positive correlation). Overall, results show that R-values indicate mostly weak or non-existent associations between variables. An example is presented in Figure 9. In Figure 9-a, task success is correlated with usability scores. R-values indicate that visualization usability did not impact task success. In Figure 9-b, error rates are correlated with task completion time. Notably in the BioMed task, we found a stronger R-value suggesting that if more time is spent to complete a task, users using graphs are likely to make fewer mistakes.

5.2   Summary of Findings

The effectiveness results suggest that when ontologies are smaller and have a simpler structure, users are likely to achieve the same level of success regardless of the specific visualization used, such as the case with the conference task. However, given more complex ontologies, the indented tree is more effective. More specifically, users are likely to be more successful at activities that concern the evaluation of existing mappings using indented trees, but more successful at activities that involve creating new mappings using the graphs. This finding suggests that the indented tree visualization is more suitable for list-checking activities, and the graph visualization is more suitable for overviews.

The efficiency results suggest that the task completion time is more likely to be a result of domain familiarity (the majority of participants being engineering students) rather than a direct cause of the specific visualization used, since both tasks had comparable completion times and the differences are not statistically significant. Similar findings are shown in the workload ratings, where participants did not feel a particular visualization is more demanding than the other.

(a) Correlating Task Success and Usability Score
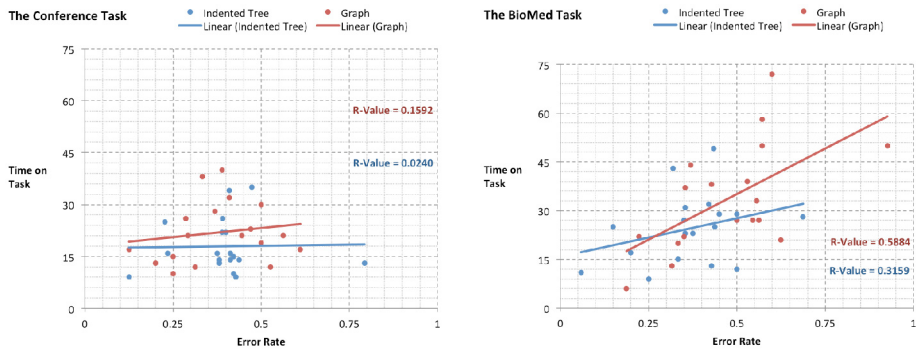
(b) Correlating Error Rate and Time on Task

Fig. 9. Correlation Results. The axes represent the variables being tested for correlation. The scattered plots illustrate individual participant results. Trend lines indicate linear regressions between the variables.

Another notable finding is that although most participants were interacting with graphs for the first time, they did not feel that it was more difficult to learn as suggested by the ease of learning ratings. Some mentioned that graphs held their attention better. However, it is clear that graphs can become difficult to manage once they exceeded a certain threshold of nodes. This finding suggests that the graph size should not be overlooked when determining the suitability of its application.

Multiple inheritance is inevitable in certain domains. Visualization techniques that can seamlessly incorporate such conceptual models are essential to users, and as such, graphs are more suitable than indented trees in these scenarios as noted by several participants. Take the SNOMED visualization snippets shown in Figure 2 as an example. *Yeast* has two parents: *Fungal_morphologic_state* and *Unclassified_fungus*; *Fungal_morphologic_state* also has two parents: *Fungal_life-cycle_form* and *Fungus*. This semantic structure is illustrated with ease using directional edges in the graph visualization (see Figure 2-a). However, *Fungal_morphologic_state* is shown twice and *Yeast* appears three times (see Figure 1-b) in the indented tree. This visual duplication in indented trees requires users to make additional efforts when understanding the data at hand and can potentially add to confusion.

Another disadvantage of the indented tree is that given a fixed screen space, it is not always possible to view the entire tree structure. It is particularly challenging given ontologies with greater depth and a large number of descendants per node. The sheer amount of expanders can be overwhelming and this makes it difficult for the user to preserve a mental model of the ontological hierarchy. While the indented tree offers little adaptation to the user, the graph visualization is much more customizable and adaptive. For example, users can simply place previously explored nodes on the far side of the screen to make room for nodes that are of current interest. Users stated that the flexibility offered by graphs helped them to better hold their attention during the tasks. A disadvantage of the graph is that it can quickly get busy on a fixed screen size providing ineffective visualization given a large number of nodes. Overall, the advantage of the indented tree is that it is familiar and predictable, as most participants are already accustomed to this visualization technique given its similarity with computer file directories. However, we attempted to minimize this bias by presenting visualizations in Protégé (given it is representative of state-of-the-art indented tree techniques and none of the participants have encountered it before), as it is unlikely for one to find participants who have never seen a computer directory before participating in our experiment.

6   Conclusions, Limitations and Future Work

Given the different strengths and weaknesses associated with graphs and indented trees, their applications should thus be determined upon specific ontology characteristics, visualization needs and user goals. Tool designers should consider combining multiple ontology visualization techniques that can engage users from different viewpoints yet are complementary to one another. In addition, ontology visualization should aim to empower users by providing customizable visualizations that are not only in manageable segments but are also adaptive to diverse personal preferences and styles.

The results of our study are dependent upon the visualization implementation, datasets used and participants involved. Although the graph visualization is representative of current techniques, some behaviors are unique to this specific force directed implementation. For instance, class names can overlap in graphs, and although participants can easily drag and rearrange nodes for a better view of the text, this process can increase frustration for users. Nevertheless, we have uncovered some motivating results from this study.

Our study suggests several future research directions. First, it would be useful to conduct studies with larger participant groups, as increased sample sizes could potentially lead to more statistically significant findings. Feedback regarding the controllable nature of graphs is specific to the implementation used in this study. Future experiments could explore non-editable graph layouts as well as other visualization techniques, such as treemaps and SpaceTrees. In addition, although it is relevant to investigate usability issues that arise among novice users, it can be even more informative for the study to recruit true ontology and mapping experts. Secondly, the datasets used in this study involve a limited set of ontologies and mappings. Future studies including larger ontologies from other domains and an increased number of mappings may uncover additional scalability issues. However, it may be challenging to recruit volunteers given tasks that could take hours or days to complete. Moreover,

the ontologies used in this study mostly contain hierarchical relationships among classes. Other object properties (e.g., transitive relationships, inverse relationships, etc.) associated with ontological entities can be the focus of further studies. For instance, future experiments could investigate whether graphs are more suitable to visualize object properties. Lastly, it may be beneficial to apply other evaluation approaches discussed in section 2 such as identifying usability issues based on observations of users over a long period of time.

References

[1] Stab, C., Nazemi, K., Breyer, M., Burkhardt, D., Kohlhammer, J.: Semantics Visualization for Fostering Search Result Comprehension. In: Simperl, E., Cimiano, P., Polleres, A., Corcho, O., Presutti, V. (eds.) ESWC 2012. LNCS, vol. 7295, pp. 633–646. Springer, Heidelberg (2012)

[2] Dadzie, A.S., Rowe, M.: Approaches to Visualising Linked Data: A Survey. Semantic Web 2, 89–124 (2011)

[3] Katifori, A., Halatsis, C., Lepouras, G., Vassilakis, C., Giannopoulou, E.: Ontology Visualization Methods - A Survey. ACM Computing Surveys 39(4), Article 10 (2007)

[4] Sivakumar, R., Arivoli, P.V.: Ontology Visualization Protégé Tools – A Review. International Journal of Advanced Information Technology 1(4) (2011)

[5] Noy, N.F., Shah, N.H., Whetzel, P.L., Dai, B., Dorf, M., Griffith, N., Jonquet, C., Rubin, D.L., Storey, M.-A., Chute, C.G., Musen, M.A.: BioPortal: Ontologies and Integrated Data Resources at the Click of A Mouse. Nucleic Acids Research 37(2), 170–173 (2009)

[6] Vercruysse, S., Venkatesan, A., Kuiper, M.: OLSVis: An Animated, Interactive Visual Browser for Bio-ontologies. BMC Bioinformatics 13(1), 116 (2012)

[7] Lanzenberger, M., Sampson, J., Rester, M.: Ontology Visualization: Tools and Techniques for Visual Representation of Semi-Structured Meta-Data. Journal of Universal Computer Science 16(7), 1036–1054 (2010)

[8] Khattak, A.M., Latif, K., Khan, S., Ahmed, N.: Ontology Recovery and Visualization. In: The 4th International Conference on Next Generation Web Services Practices, pp. 90–96 (2008)

[9] Brockmans, S., Volz, R., Eberhart, A., Löffler, P.: Visual Modeling of OWL DL Ontologies using UML. In: McIlraith, S.A., Plexousakis, D., van Harmelen, F. (eds.) ISWC 2004. LNCS, vol. 3298, pp. 198–213. Springer, Heidelberg (2004)

[10] Tudorache, T., Nyulas, C., Noy, N.F., Musen, M.A.: WebProtege: A Collaborative Ontology Editor and Knowledge Acquisition Tool for the Web. Semantic Web Journal 4(1), 89–99 (2013)

[11] Baehrecke, E.H., Dang, N., Babaria, K., Shneiderman, B.: Visualization and Analysis of Microarray and Gene Ontology Data with Treemaps. BMC Bioinformatics 5, 84 (2004)

[12] Plaisant, C., Grosjean, J., Bederson, B.B.: SpaceTree: Supporting Exploration in Large Node Link Tree, Design Evolution and Empirical Evaluation. In: The IEEE Symposium on Information Visualization, vol. 57 (2002)

[13] Parsia, B., Wang, T., Goldbeck, J.: Visualizing Web Ontologies with CropCircles. In: The 4th International Semantic Web Conference. LNCS, vol. 3729, pp. 6–10 (2005)

[14] Bosca, A., Bomino, D., Pellegrino, P.: OntoSphere: More than A 3D Ontology Visualization Tool. In: The 2nd Italian Semantic Web Workshop. CEUR-WS, vol. 166 (2005)

[15] Motta, E., Mulholland, P., Peroni, S., d'Aquin, M., Gomez-Perez, J.M., Mendez, V., Zablith, F.: A Novel Approach to Visualizing and Navigating Ontologies. In: Aroyo, L., Welty, C., Alani, H., Taylor, J., Bernstein, A., Kagal, L., Noy, N., Blomqvist, E. (eds.) ISWC 2011, Part I. LNCS, vol. 7031, pp. 470–486. Springer, Heidelberg (2011)

[16] Storey, M.-A., Musen, M., Silva, J., Best, C., Ernst, N., Fergerson, R., Noy, N.: Jambalaya: Interactive Visualization to enhance Ontology Authoring and Knowledge Acquisition in Protégé. In: Workshop on Interactive Tools for Knowledge Capture (2001)

[17] Petrelli, D., Mazumdar, S., Dadzie, A.S., Ciravegna, F.: Multi Visualization and Dynamic Query for Effective Exploration of Semantic Data. In: Bernstein, A., Karger, D.R., Heath, T., Feigenbaum, L., Maynard, D., Motta, E., Thirunarayan, K. (eds.) ISWC 2009. LNCS, vol. 5823, pp. 505–520. Springer, Heidelberg (2009)

[18] Kuhar, S., Podgorelec, V.: Ontology Visualization for Domain Experts: A New Solution. In: The 16th International Conference on Information Visualisation, pp. 363–369 (2012)

[19] Alani, H.: TGVizTab: An Ontology Visualisation Extension for Protégé. In: Workshop on Visualization Information in Knowledge Engineering (2003)

[20] Falconer, S.M., Callendar, C., Storey, M.-A.: A Visualization Service for the Semantic Web. In: Cimiano, P., Pinto, H.S. (eds.) EKAW 2010. LNCS, vol. 6317, pp. 554–564. Springer, Heidelberg (2010)

[21] Fu, B., Grammel, L., Storey, M.-A.: BioMixer: A Web-based Collaborative Ontology Visualization Tool. In: The 3rd International Conference on Biomedical Ontology. CEUR-WS, vol. 897 (2012) ISSN 1613-0073

[22] Herman, I., Melançon, G., Marshall, M.S.: Graph Visualization and Navigation in Information Visualization: A Survey. IEEE Transactions on Visualization and Computer Graphics 6(1), 24–43 (2000)

[23] Graham, M., Kennedy, J.: A Survey of Multiple Tree Visualisation. Information Visualization 9(4), 235–252 (2009)

[24] Lam, H., Bertini, E., Isenberg, P., Plaisant, C., Carpendale, S.: Seven Guiding Scenarios for Information Visualization Evaluation. Technical Report (2010)

[25] Chen, C., Czerwinski, M.: Empirical Evaluation of Information Visualizations: An Introduction. International Journal on Human-Computer Studies 53, 631–635 (2000)

[26] Saraiya, P., North, C., Duca, K.: An Insight-Based Methodology for Evaluating Bioinformatics Visualization. IEEE Trans. on Visualization and Computer Graphics 11(4) (2005)

[27] Akrivi, K., Elena, T., Constantin, H., Georgios, L., Costas, V.: A Comparative Study of Four Ontology Visualization Techniques in Protégé: Experiment Setup and Preliminary Results. In: The 10th International Conference on Information Visualization, pp. 417–423 (2006)

[28] Swaminathan, V., Sivakumar, R.: A Comparative Study of Recent Ontology Visualization Tools with a Case of Diabetes Data. International Journal of Research in Computer Science 2(3), 31–36 (2012)

[29] Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. Human Mental Workload. North Holland Press, Amsterdam (1988)

[30] Brooke, J.: SUS: A Quick and Dirty Usability Scale. Usability Evaluation in Industry. Taylor & Francis, London (1996)

[31] Lund, A.: Measuring Usability with the USE Questionnaire. Usability and User Experience Newsletter of the STC Usability SIG (2001)

[32] Benedek, J., Miner, T.: Measuring Desirability: New Methods for Evaluating Desirability in a Usability Lab Setting. In: Usability Professionals Association Conference (2002)

[33] Hart, S.G.: NASA-Task Load Index (NASA-TLX); 20 Years Later. In: The Human Factors and Ergonomics Society 50th Annual Meeting, pp. 904–908 (2006)

[34] Likert, R.: A Technique for the Measurement of Attitudes. Archives of Psychology 140, 1–55 (1932)