Eye Tracking the User Experience - An Evaluation of Ontology Visualization Techniques

Bo Fu[a, *], Natalya F. Noy[b], Margaret-Anne Storey[a]
[a] *Department of Computer Science, University of Victoria, Canada*
[b] *Google Inc.*

Abstract. Various ontology visualization techniques have been developed over the years, offering essential interfaces to users for browsing and interacting with ontologies, in an effort to assist with ontology understanding. Yet, few studies have focused on evaluating the usability of existing ontology visualization techniques. This paper presents an eye-tracking user study that evaluates two commonly used ontology visualization techniques, namely, indented list and graph. The eye-tracking experiment and analysis presented in this paper complements the set of existing evaluation protocols for ontology visualization. In addition, the results found from this study contribute to a greater understanding of the strengths and weaknesses of the two visualization techniques, and in particular, how and why one is more effective than the other. Based on approximately 500 MB of eye-tracking data containing around 30 million rows of data, we found indented lists to be more efficient at supporting information searches, and graphs are more efficient at supporting information processing.

Keywords: eye tracking, ontology visualization evaluation, usability study, statistical analysis

1. Introduction

Information visualizations have long become rooted in various aspects of our daily lives, from reading subway maps, to illustrating trends in weather forecasts, to monitoring fluctuations in stock markets. The benefits of information visualizations (InfoViz), undoubtedly, are immense, as they offer clarity, promote insight and understanding of a dataset [59]. The evaluation of visualizations is central to the advancement of existing techniques, as it is only through evaluation that we are able to demonstrate the effectiveness of a technique, identify potential shortcomings and design further improvements to better support and amplify cognition.

Over the years, various ontology visualization techniques have been developed and applied in a range of applications including ontology editors such as Protégé [52], NeOn Toolkit [53], and ontology libraries such as BioPortal [22] and OLS [56]. However, despite a tremendous amount of effort to develop new layouts and improve drawing algorithms, few visualization techniques are accompanied by evaluations of their actual usability. A lack of scientific evaluations of existing ontology visualization techniques could be potentially damning to the advancement of this field as a whole, as we may fail to recognize and adopt good designs, or to identify and reject bad practices.

In this paper, we identify two frequently used ontology visualization techniques including indented lists and graphs, and present an evaluation comparing their support in ontology understanding. In addition to traditional evaluation techniques that measure ontology visualizations indirectly (i.e. through comparison of users' speed and accuracy when using different visualizations to complete the same tasks), we apply eye tracking to complement the usability analysis of the two visualization techniques, in order to gain a deeper understanding of why one visualization technique is more efficient or effective than the other. In particular, we aim to answer the following research questions:

- Is one visualization technique more efficient at supporting *information search* than the other?
- Is one visualization technique more efficient at supporting *information processing* than the other?
- Is one visualization technique better at *reducing cognitive workload* than the other?
- Do participants using one particular visualization technique are faster and/or more successful?

---

* Corresponding author. E-mail: bofu@uvic.ca.

2. Related Work

In this section, we discuss the rationale of including indented lists and graphs in our evaluation, and our motivation for combing eye tracking with indirect evaluation measures such as speed and accuracy.

2.1. Ontology Visualization Techniques

Katifori & Halatsis [3] review 34 ontology visualization tools and classify them into six categories including *indented list, node-link and tree, zoomable, space-filling, focus with context or distortion*, and *3D information landscapes*, based on the presentation, interaction, functionality and dimensions used. The predominant functionalities of the tools are used in the categorization when some tools contain overlapping features of other categories, e.g., when some space-filling techniques are also zoomable. A key observation from the review is that among the surveyed visualizations, 24 techniques[2] use indented lists and graphs to present ontological classes and their relationships. In this paper, *indented list* visualizations refer to those that use indentation to illustrate super/sub-class relationships and there is one single path between any pair of ontological nodes, and *graph* visualizations refer to node–link diagrams that have nodes with connecting edges. *Zoomable* visualization techniques[3] allow users to enlarge ontological nodes at specific levels, where classes are presented as nested nodes in the visualization. *Space-filling* techniques[4] are designed to make use of the entire screen, with larger classes (i.e. classes that have more children) taking up more space. *Context with focus or distortion* refers to a group of techniques that distort graphs in order to present a focused view of a node and its context (i.e. surrounding super and sub-classes) to the user. The underlying technique of this category is essentially node-link diagram, which can be classified as graph visualization for the purpose of this paper. *3D Information landscapes* refer to techniques[5] that use color and size coded 3D objects to map out the hierarchy of an ontology.

Given a readily available set of visualization techniques, recent research has focused on improving the underlying technical algorithms and platforms. For instance, reducing the number of nodes displayed at once in an effort to avoid cognitive overload in KC-Viz [25]; supporting networked ontology engineering in the NeOn toolkit [53]; improving real-time interaction with force-based layouts in OLSVis [65]; presenting multiple visualization views as discussed in Kuhar & Podgorelec [61], BioPortal [22], BioMixer [8], and ALViz [48]; moving the ontology editing and navigation process to an online environment in WebProtégé [70] and FlexViz [63]; as well as supporting web-based collaborative visualization in BioMixer [8]. Notably, the majority of recent advancement in ontology visualization has incorporated or is building upon indented lists and graphs, which makes these two most commonly used visualization techniques relevant and interesting subjects for our evaluation.

2.2. Evaluation Approaches in InfoViz

Evaluating visualizations is a challenging task as discussed in [20], as ideally we need to use realistic tasks that echo real-world problems, cope with incomplete datasets, and report on long term use in natural settings. Over the years, a range of evaluation approaches have been developed to improve our understanding of usability issues associated with different visualization types. Komlodi et al. [5] identify four themes of evaluation approaches including i) controlled experiments that compare design elements, ii) usability evaluation of tools, iii) controlled experiments that compare tools, and iv) case studies in realistic settings. Lam et al. [30] reviewed over 800 papers with 345 evaluations and identified seven guiding scenarios based on the goal of the evaluation, including i) evaluating environments and work practices in order to elicit formal requirements for design, ii) evaluating visual data analysis and reasoning in order to assess how a visualization tool supports users in generating actionable and relevant knowledge in specific domains, iii) evaluating communication through visualization in order to assess how messages are conveyed to relevant audiences, iv) evaluating collaborative data analysis in order to study how a visualization tool support joint conclusion and/or decision-making, v) evaluating user performance in order to investigate how specific features affect measurable user performance objectively, vi) evaluating user experience in order to understanding how people react to visualizations either in a short or a long time span, and vii) automated evaluation of visualizations to test certain aspects that can be measured automatically through a computerized procedure. Carpendale [55] groups visualization evaluation techniques into quantitative and qualitative approaches. Quantitative approaches typically include controlled experiments that measure the effect of (a set of) independent variable(s) on (a set of) dependent variable(s), and the application of statistical tests. Examples

---

[2] These 2D and 3D tools include Protégé Class Brower [52], OntoEdit [74], OntoViz [50], GOBar [38], IsAviz [26], SpaceTree [21], TreePlus [10], OntoTrack [68], GOSurfer [66], GOMiner [11], Cone Tree [28], fsviz [31], Reconfigurable Disk Tree [18], Tree Viewer [24], OntoSphere [1], Information Cube [37], HyperTree Visualization [45], OntoRama [54], StarTree [35], MoireGraphs [44], Protégé TGVizTab [29], Ozone [14], BiFocal Tree [15], and 3D Hyperbolic Tree [69].
[3] These include 2D and 3D techniques such as Grokker [73], Jambalaya [46], CropCircles [71], Information Pyramids [39], and Gopher VR [41].
[4] This category includes 2D and 3D techniques such as TreeMaps [12], SequoiaView [58], Information Slices [40], and BeamTrees [27].
[5] Examples include the File System Navigation [62] and the Harmony Information Landscape [47].

of qualitative evaluations include interviews, observations, and think-aloud protocol. Regardless of the specific evaluation approach applied, there is always a trade-off between the generalizability, precision and realism of the results [36]. For instance, controlled experiments can produce precise results, but may lose realism when tasks are overly simplified to suit the needs of the study and fail to mirror real-world problems. Interviews and observations may collect in-depth results from realistic tasks in natural settings, but lack precision and are difficult to generalize when the sample size is too small to minimise bias in individual subjectivity.

User studies or user-centered evaluations have become the predominant method in recent years and are well documented in terms of experimental design, data collection and analysis [17, 19, 51]. Using eye tracking as a method to infer cognitive processes in user-centered evaluation has long been practiced in InfoViz, and early eye movement studies date back many decades in cognitive and perceptual psychology [23]. As eye tracking hardware and algorithms mature, various reviews on tracking systems [2, 7], as well as guidelines on experimental designs and metrics have been proposed [6, 34], which have been applied in a range of research fields such as neuroscience, psychology, marketing and computer science [2]. Compared to the aforementioned evaluation approaches that indirectly measure visualizations (i.e. through interviews and controlled experiments), eye tracking is able to provide direct information on the eye movements of a participant during an experiment. In other words, while speed and accuracy evaluation tells us *what* has happened, eye tracking results help us understand *why* certain design elements lead to an increase/decrease in speed and accuracy. This makes eye tracking a suitable method for the evaluation of visualizations, as it offers a "more focused tool to investigate micro-level design issues" [32].

2.3. Evaluation of Ontology Visualizations

Compared to the advances in developing ontology visualization techniques, efforts to evaluate existing and new visualizations have been relatively scarce. One evaluation approach is to benchmark the proposed visualization against one or a set of existing tools in terms of supported features and use cases [64, 65]. The evaluation typically outlines the extra features and functions that are supported by the proposed tool but not by other existing tools. It is often difficult to argue against bias in the choice of the benchmarks and scenarios. For instance, what makes some tools reliable benchmarks but not others, or how to ensure that the use cases are selected objectively and are not chosen to simply make the proposed tool appear superior.

Another evaluation approach is to elicit qualitative feedback through a range of techniques such as interviews, observations and think-aloud protocols [60]. As discussed previously, it is often challenging to argue for representativeness and generalizability of the results due to individual preferences and subjectivity. This is especially difficult when the sample size is very small.

A third evaluation approach that is most often practiced is controlled experiments that typically measure the speed and accuracy such as [9, 72, 25, 4, 67]. While typically, these experiments can evidently show better results of some visualizations in terms of their effectiveness and efficiency, but they do not necessarily provide insight to why that is happening. Building upon these existing evaluation strategies, we aim to provide a micro-level understanding of ontology visualizations through combining eye tracking in controlled experiments.

3. Eye Tracking User Study

The goal of our study is to investigate the extent to which indented lists and graphs can support users in the process of understanding the semantics in ontologies. The rationale for choosing the tasks and the physical setup of the experiments are discussed next.

3.1. Research Questions & Study Setup

We presented each participant with a set of mappings between two ontologies, 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. The participant needs to: 1) identify correct and incorrect results among an existing set of mappings (i.e., determining the correctness of the given mappings); and ii) create 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.

We used a Tobii 2150 [6] eye tracker to track eye movements during each task session, and the ClearView 2.7.1 [7] software to export gaze data. Figure 2 shows a picture of the physical setup. The eye-tracking monitor (sitting on the table in Figure 2) integrates a 21.3″ TFT (thin-film-transistor) display that supports a maximum of 1600×1200 pixels. High-resolution cameras are built into the monitor with a field-of-view of 26×20×32cm at 73cm away from the monitor. The Tobii eye tracker uses near infrared diodes to generate and collect reflection patterns on the corneas of the eyes of the user. It uses sophisticated processing and validation algorithms to calculate the positions of the eyes and gaze points on the display. For further details on eye tracking methodologies, see [2]. As shown in Figure 2, the Tobii monitor has a similar appearance with a regular computer screen as it does not have any other hardware

---

[6]
http://chiselgroup.files.wordpress.com/2012/07/product_descriptio n_tobii_50_series.pdf

[7] http://chiselgroup.files.wordpress.com/2012/07/tobii-clearview-user-manual_cv2-6.pdf

elements visible to the user. In addition, the participants are not required to wear helmets or other restrictive gears. These conditions thus minimise distractions to the user and create an ordinary interaction between the user and the computer during the study.

To successfully evaluate a mapping, the 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. Since it is only possible to complete the tasks after exploring and interacting with the ontology visualizations, this setup ensures the evaluation focuses on how well the indented lists and graphs support ontology understanding. In particular, the participant must first *search* for relevant information in the visualizations, and after *processing* this information can she (or he) use this knowledge to complete the tasks. This experimental design is thus well suited to answer our research questions (discussed in section 1).
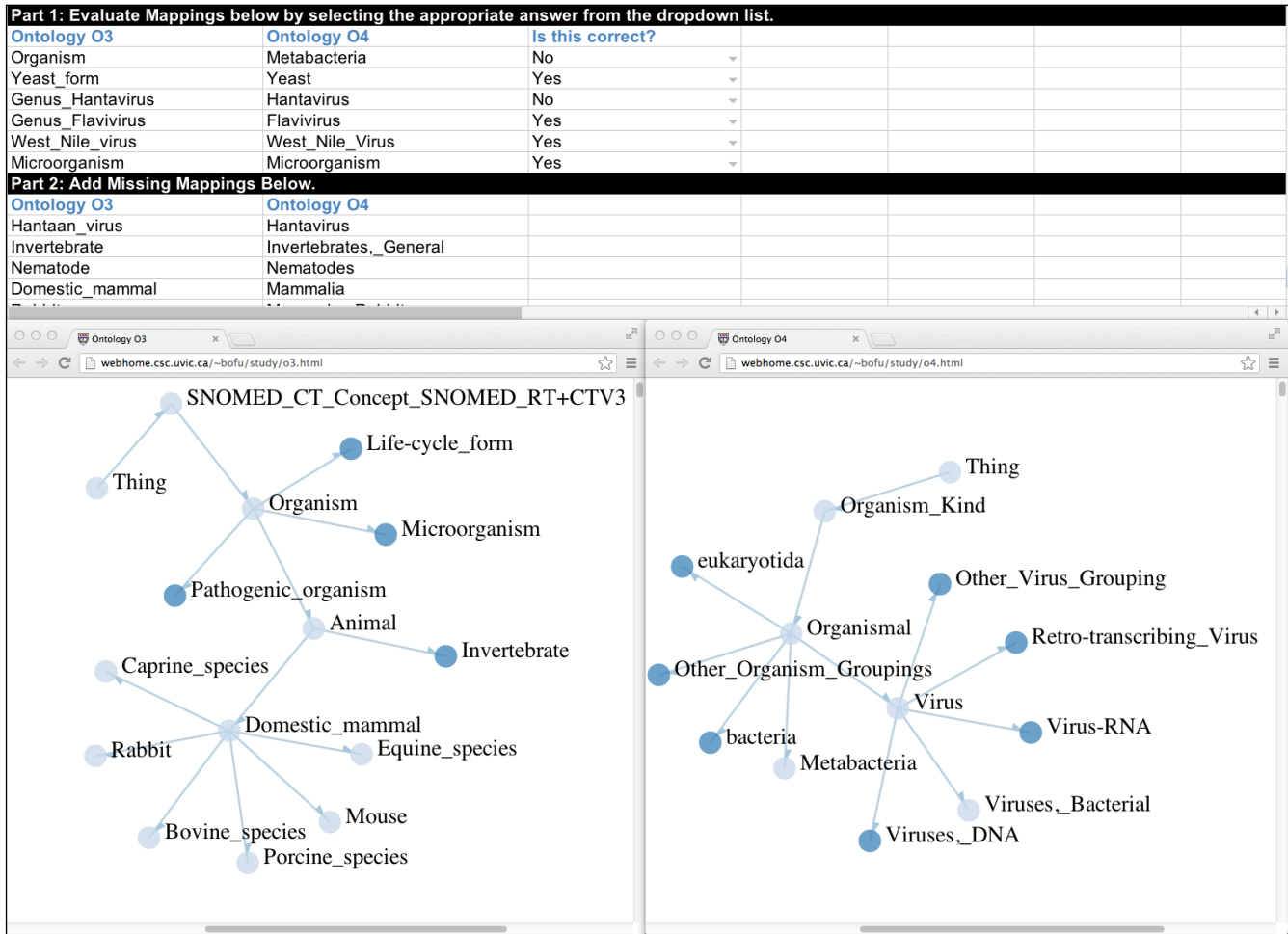

Fig. 1. Sample Task Screen. In this example, graphs are used to visualize two biomedical ontologies. Mappings to be evaluated are presented in a spreadsheet. Interacting with the visualizations, participants must use drop-down lists containing either "yes" or "no" responses to evaluate the correctness of existing mappings (in part 1) and add missing mappings by typing class names (into part 2 of the spreadsheet).
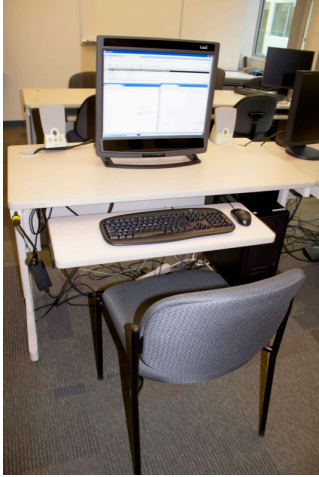
Fig. 2. Tobii 2150 Eye Tracker used in the Study. The eye tracker was set up in an office environment. There were indoor lighting but no direct sunlight or other high intensity near-infrared light.

3.2. Datasets & Visualizations

We used two pairs of ontologies in the study. Each accompanied by a set of mapping standards, taken from the Ontology Alignment Evaluation Initiative (OAEI) 2012 conference[8] and the BioMed[9] tracks. Table 1 presents an overview of the ontologies used in this study.[10]

The conference ontologies describe the organization of conferences with a total of 74 and 100 classes respectively, with at most 3 or 6 classes on the longest path to root, at most 8 or 9 subclasses for a class, and without any multiple inheritance. The conference task represents a less difficult scenario, because 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 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

We presented indented list visualizations to the participants by loading ontologies into Protégé and asked participants to interact with the lists exclusively but not with any other features in Protégé. We implemented graph visualizations[11] in force directed layouts using the D3 JavaScript library[12].

Figure 3 presents visualization snippets of indented list (Figure 3-a) and graph (Figure 3-b). In the indented list 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. The presence of the expander itself illustrates whether a node is expandable.

In the graph visualization, classes are illustrated by nodes and is-a relationships are illustrated by directional edges with arrowheads pointing to the subclasses. The coloring of a node denotes whether it is expandable (i.e., dark-colored nodes illustrate the existence of subclasses whereas light-colored nodes illustrate nonexpandable classes). Clickable nodes allow users to toggle children of a particular class. In addition to using scroll bars to adjust the viewing area, users can 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 the rest of the classes.

The visualizations used in our study are representative of current indented list and graph visualization techniques, as they contain key characteristics of their respective visualization type shown from the literature review (e.g. 52, 53, 70, 22, 56]). Also, as Shneiderman point out [13], visualizations should provide high-level overviews, and support zoom-ins on specific parts and filter-outs of irrelevant parts. In the indented list visualization, collapsed

---

[8] http://oaei.ontologymatching.org/2012/conference/index.html
[9] http://www.cs.ox.ac.uk/isg/projects/SEALS/oaei/2012/
[10] The ontologies 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, or o4.owl. The gold standards used in this study can be found at the same prefix followed by file name: o1-o2%20gold%20standard.rdf or o3-o4%20gold%20standard.rdf.

[11] The graph visualizations 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.
[12] http://d3js.org

expanders allow an overview of the ontological structure with irrelevant information remaining hidden, while further expansion allows the user to focus on specific parts of the ontology. Similarly in the graph visualization, collapsed color-coded nodes present an overview of the ontology at a high level while keeping irrelevant details hidden, and expanded nodes allow the user to concentrate on specific parts of the ontology. Since the two visualizations used in our study each satisfy all of the above, we believe that the evaluation setup is reasonable and fair, and that bias is minimized since the indented list and the graph visualization are highly comparable in terms of their functionalities.

(a) Indented list Visualization

(b) Graph Visualization
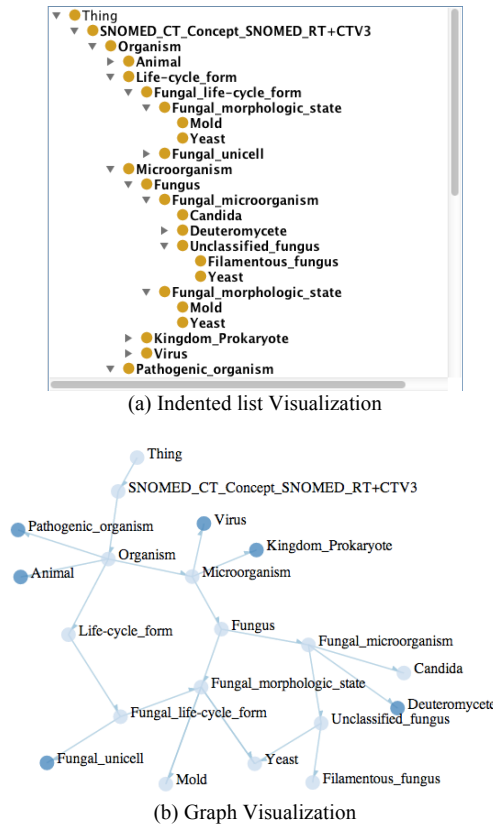
Fig. 3. Visualization Techniques Investigated in the Study. The example shown here is taken from the SNOMED ontology.

3.3. Participants & Protocol

We recruited volunteers via mailing lists at the University of Victoria. 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 increasingly include people with little knowledge of semantic technologies (e.g., BioPortal [22] 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 especially when a true expert should be able to successfully and accurately evaluate mappings regardless of the visualization support.

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.[13] The participant was then given instructions on the goal of her (or his) tasks, namely, evaluating a set of exact class mappings between a pair of ontologies.[14] 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 indented lists, and then we asked her to complete the BioMed task using graphs. In another session, we asked Bob to complete the BioMed task using indented lists, and then we asked him to complete the conference task using graphs. We randomly assigned tasks, ensuring equal distribution of tasks in the population as well as counterbalancing the order of tasks and visualizations 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 and visualization order on the study outcome.

Each participant began a study session by first calibrating her (or his) eyes. The calibration lasted 6 seconds and required the participant to look at a dot that appeared in different locations on the display. This process allows the eye tracker to accurately track a subject's eye movements in the subsequent task recording, based on the calibration data that was collected while the fixation locations were known to the eye tracker. In addition, we asked each participant to look at an empty task screen (i.e., Figure 1 without any visualization or data in the spreadsheet) for 10 seconds to collect her (or his) baseline pupil data, which was later compared to the pupil data generated during the task recording for pupil dilation analysis (discussed in section 4.3). After calibration and baseline data collection, the participant was then given a task to complete, where the eye tracker recorded her (or his) eye movements from the beginning to the end of the task.

---

[13] Tutorials, videos and other materials used in this study can be found at https://sites.google.com/site/uvicstudy

[14] Since the human evaluation process is likely to contain a similar set of procedures regardless of the type of entity (e.g. class or property) or mapping relation (e.g. exact match, narrow or broad match), it is thus sufficient to use exact matches as examples for the purpose of this study.

Each participant sat on a stationary chair, i.e. without wheeling, leaning or swirling functions, during the calibration and throughout the task recording. This ensured that there was no sudden distance change from the eyes of a subject to the eye tracker. Participants were told to use the display as how they normally would with a computer screen. They could blink as usual and look down on their keyboards if necessary during a recording. Although the eye tracker was able to tolerate small changes in head positions, the participants were encouraged to reduce sudden large head movements to improve tracking accuracy. Furthermore, participants were asked to raise any questions before they began a task and no interactions were allowed with the researcher during a task.

4. Eye Tracking Metrics

We used a set of eye tracking metrics to answer the research questions outlined in section 1. A brief tutorial on the metrics used in our study is presented in this section, for in-depth reviews on various eye movement metrics, see [34, 49, 6].

The raw eye movement data generated by ClearView includes:

- each fixation's duration (in ms), associated with an ID and a timestamp;
- each fixation's position on the display (X and Y coordinates measured in pixels);
- for each fixation, the pupil size (in mm) for each eye, associated with its validity code (illustrating the confidence of the eye tracker).

Based on the above, measures of information search, information processing, and cognitive workload can be calculated, which are discussed next.

4.1. Measures of Search

Measures of search include fixation and saccade count, saccade length, convex hull and scanpath length.

Fixations are those moments when the eyes are relatively stationary. Saccades occur between fixations, which are the quick eye movements from one fixation to the next. A scanpath is the complete saccade-fixate-saccade sequence. Figure 4 illustrates these concepts with a diagram. In good visualization designs, we would expect users to find relevant components quickly, where they only need to sample a small amount of objects to complete the given task, hence leading to a lower fixation count [34]. As the saccade count is essentially the number of fixations minus one, naturally, we would also expect to see small saccade counts given low fixations. In contrast, poor designs would be expected to produce more extensive search behaviours, i.e. higher number of fixations and saccades.
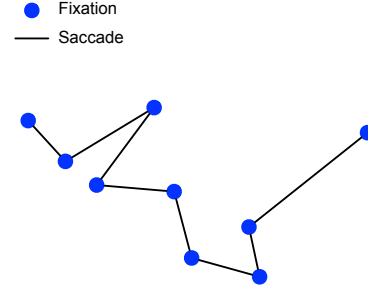

Fig. 4. Fixation, Saccade and Scanpath. Dots illustrate fixations, and lines illustrate saccades. The saccade-connected fixations illustrate the scanpath.

Given the X and Y coordinates of each fixation, we can determine the saccade length, the absolute angle that is associated with each saccade as well as the relative angles between saccades. An absolute angle (measured in degree) indicates the inclination of the saccade with respect to the X-axis. A relative angle indicates the gradient between two saccades. An example is presented in Figure 5. Given the positions of fixation points n and k, coupled with the absolute degree β, we can determine the length L of the saccade, as shown in Figure 5. A well-designed visualization should provide useful cues to direct users' scanning to desired targets rapidly with few interim fixation, meaning that we would expect to see larger saccadic amplitudes, i.e. longer average saccade length for a participant [34]. In the case of poorly designed visualizations that provide misleading or unhelpful cues, we would expect the distances between saccades to be smaller as the user continues to search near or around a component until a meaningful cue is found.
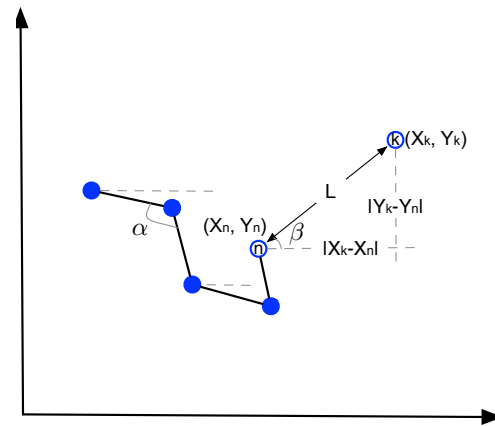

Fig. 5. Saccade Length, Absolute and Relative Angle. L illustrates the saccade length between fixations n and k. α illustrates an example of relative angle, and β illustrates an example of absolute angle.

Knowing each saccade length, the scanpath length therefore can be determined as the sum lengths of all saccades. It is important to note that lengthy scanpaths is an indication of inefficient scanning behaviour, but do not directly measure search efficiency since scanning does not distinguish between searching activities and processing

activities. Thus, scanpath length should be understood in the context of search area, or convex hull. Figure 6 shows an example of the convex hull area. In a well-designed visualization, we would expect to find a smaller search area coupled with a shorter scanpath. In contrast, a wildly distributed search pattern in poorly designed visualizations would lead to a larger search area with longer scanpath [34].
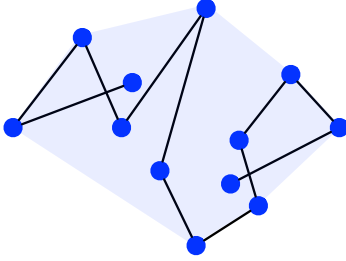

Fig. 6. The highlighted area shows an example of convex hull. The convex hull area is determined by the bounding fixations in the scanpath.

4.2. Measures of Information Processing

Measures of information processing include average fixation duration and saccade-to-fixation ratio.

Longer fixations indicate the user is spending more time interpreting or processing the objects presented [34]. In a well-designed visualization, we would expect the user to have a relatively short average fixation duration, i.e. faster processing time. Knowing each fixation duration of a participant, her (or his) average fixation duration is simply the sum of all fixation durations divided by the fixation count. Poorly designed visualizations would yield longer average fixation durations, as the cues are not as helpful.

Given the timestamps of fixations, the duration of each saccade can be determined, as the time between two succeeding fixation timestamps. Knowing the time a participant spent searching for information (saccade durations) and the time she (or he) spent processing information (fixation durations), we can determine the search-to-process ratio for the participant, as the sum of saccade durations divided by the sum of fixation durations. The higher the ratio, the more time the participant spent processing information, or the less time the participant spent searching for information. When comparing two visualizations' support for information processing, the more effective visualization should have faster processing time combined with a low search-to-process ratio, as we would expect users to not only quickly find relevant information but also spend less time processing them, i.e. lower total fixation durations coupled with lower saccade-to-fixation ratio.

4.3. Measures of Cognitive Workload

Measures of cognitive workload include pupil dilation, absolute and relative saccade angles.

The changes in pupil sizes and more specifically, the increase of pupils are an indication of more cognitive effort [49], i.e. the more difficult or demanding it is to complete the task using a particular visualization, the larger the pupils will be. In a well-designed visualization, we would expect to see no increase or very small increase of the pupil sizes compared to a poorly designed visualization. To determine the pupil size for a participant, we used the average of both eyes. To analyze pupil dilation, for each participant, we first calculated her (or his) baseline average pupil size of both eyes s while looking at an empty task screen. We then calculated her (or his) average pupil size of both eyes s' during a task. Finally, her (or his) pupil dilation is calculated as a percentage: $(s'-s)/s$.

Recall that relative angles (see Figure 5) express the successive eye movements of a participant. Within a scanpath, the sum of absolute values of relative angles indicates the directness of the scanning [32]. In a well-designed visualization, we would expect the participant to have a certain and confident scanpath throughout, as opposed to randomly scanning around with a great deal of uncertainty. Hence, when comparing two visualizations, the better-designed one should lead to smaller sum relative angles. Absolute saccade angles (see Figure 5) express the inclination of scan directions, and similar to relative angles, a confident scanpath with certainty should lead to consistent and less dispersed absolute angles. We expect that when the visualization is more demanding to work with (i.e. require more cognitive effort), the sum of relative angles would be higher and the standard deviation of absolute angles would be higher too.

4.4. Speed & Accuracy

Measures of a participant's speed and accuracy include calculating her (or his) task success, error rate and time on task.

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 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. Finally, time on task is the length of time it took a participant to complete the task (i.e. after the participant has finished both identification and creation activities).

4.4. Data Cleaning, Validation & Statistical Tests

As discussed earlier, raw data is associated with validity codes, which are an indication of how confident the eye tracker was that it had recorded the correct data. In an effort to reduce noisy data, we filtered out the raw data that were associated with poor validity codes. For example, when the eye tracker could not detect both eyes (e.g., data of one eye is associated with a poor validity code), we would then discard that entire row of raw data in our analysis.

Eye trackers can be sensitive to participants with eye-wear, such as glasses and contact lenses. They can also have difficulty tracking eyes with certain characteristics such as large pupils, or when eyelids obscure pupils. We thus discarded incomplete data sets that had lost traces of some participants' eyes during the recordings. We discarded 5 participants' data during the data cleaning procedure, which led to a total of 31 valid data sets. Among which, 16 participants used indented lists to solve the conference task and graphs to solve the BioMed task, and another 15 participants used graphs to solve the conference task and indented lists to solve the BioMed task.

When differences are found between the two visualizations, we used mixed-effects models to validate statistical significances of our findings, which present estimates of fixed effects in the presence of correlated data and unequal variances. For a comprehensive review on mixed-effects models, see [16]. In addition, we measured covariance between variables using the Pearson correlation coefficient [42] to determine the degree of associations among these variables. Following the guidelines presented by Taylor [57], in absolute values, we interpret weak correlations to be $R \leq 0.35$, moderate correlations to be $0.36 \leq R \leq 0.67$, high correlations to be $0.68 \leq R \leq 1.00$, and very high correlations to be $R \geq 0.09$.

5. Results

We had 36 participants in total, who generated recordings that ranged between ten minutes to over an hour. Typically, an approximate 20-minute recording would correspond to 65,000 rows of raw data. Overall, we collected approximately 500 MB of raw data, which contains approximately 30 million rows of eye movement data. To automatically process this large volume of raw data, we generated and ran a script[15] on them.

5.1. Information Search

Figure 7 presents box plots of the fixation counts that were generated for each participant. In both tasks, fixation counts for participants who used indented lists vary less than those who used graphs. The aggregated group mean fixation counts are shown in Figure 8, where the average number of fixations is always lower for participants who used indented lists. This difference is statistically significant. Similarly, as expected, we found smaller number of saccades for those who used indented lists (see Figure A-3). These results indicate that participants needed to sample a fewer number of visualization components when using indented lists. In other words, indented lists were more effective at assisting participants to efficiently search for information compared to graphs.

When comparing the average saccade length, we did not find statistically significant differences between the participants who used indented lists and those who used graphs. It is important to note that non-statistically significant differences do not imply a lack of difference between the two user groups (i.e. they are the same). It simply means that the results are inconclusive, in that a difference may exist although it was not shown in the present sample.
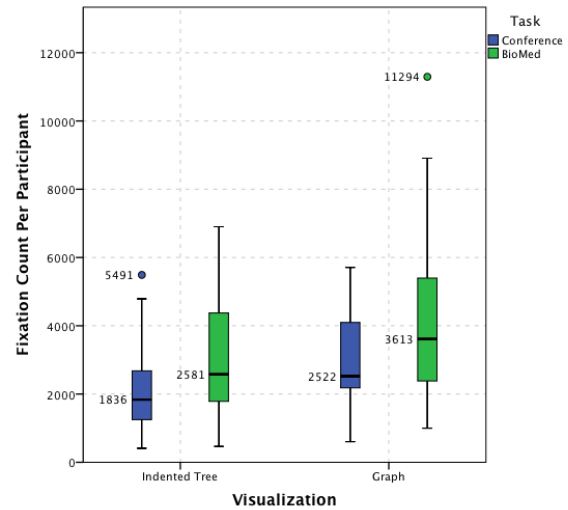

Fig. 7. Distribution of Fixation Counts. For each participant, her (or his) number of fixations during a task is generated. The horizontal lines within the boxes indicate the median values, which are labelled next to the boxes. The top of the box represents the upper quartile, which 75% of the cases fall. The bottom of the box represents the lower quartile, which 25% of the cases fall. The T-bars illustrate minimum and maximum values. The individual points represent outliers.

---

[15] The code used to process the raw data can be found at https://github.com/BoFu/Eye.Tracking.Data.Analysis
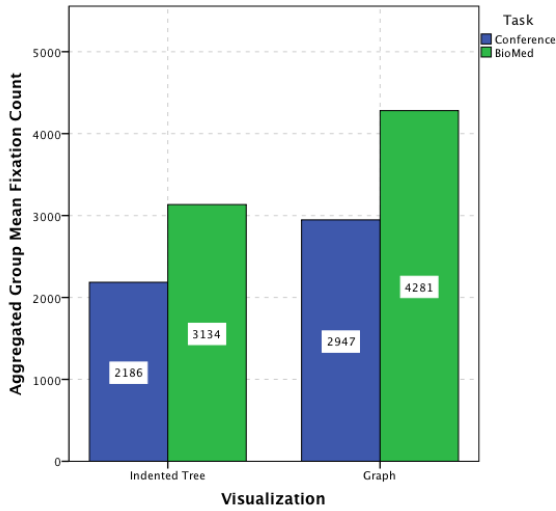
Fig. 8. Aggregated Group Mean Fixation Count. Mean values are labelled on each bar. The differences between the two visualization groups are statistically significant with P<0.05.

Figure 9 presents an overview of the convex hull areas found for each participant. The variations of convex hull were greater for participants who used indented lists. Figure 10 shows the aggregated group mean convex hull, where indented lists had a smaller area in the conference task, and graphs had a smaller area in the BioMed task. Convex hull areas should be understood in the context of scanpaths, as discussed previously (in section 4.1), i.e. more efficient search have smaller convex hull area companied by shorter scanpath, we thus compared the scanpath lengths of the two user groups.
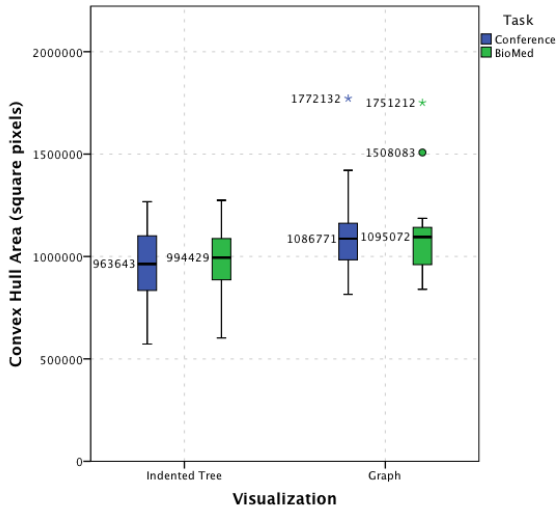
Fig. 9. Distribution of Convex Hull Areas. We calculated the convex hull area for each participant.
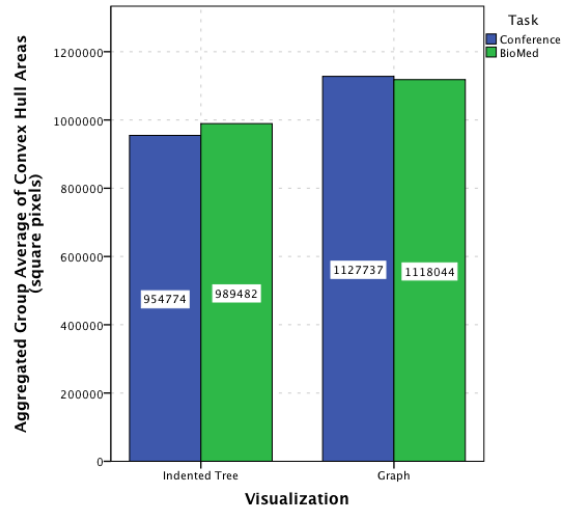

Fig. 10. Comparison of Aggregated Group Means found in Convex Hull Areas. The differences are statistically significant with P<0.02.

Figure 11 presents an overview of the data generated for scanpath lengths, and Figure 12 presents the aggregated group means. Participants who used indented lists had consistently shorter scanpaths with less variations compared to those who used graphs. These results indicate that indented lists were generally more effective at supporting participants searching for information, since participants generally searched within smaller areas with shorter scanpaths. In contrast, even though in the BioMed task, participants searched within a slightly smaller area using graphs, but since they had much lengthier scanpaths, it suggests they had inconsistent and random searches and felt lost during the task.
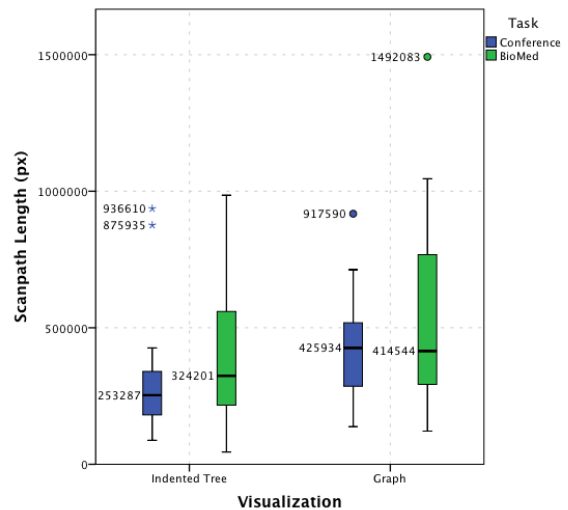

Fig. 11. Distribution of Scanpath Lengths. For each participant, we generated the length of individual saccade. The sum of all saccades' lengths is the scanpath length.
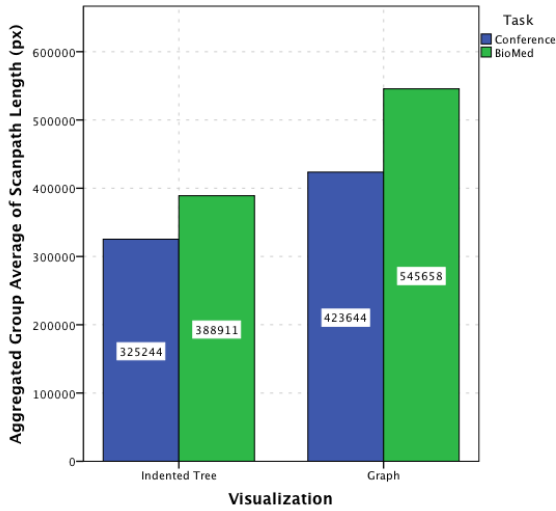
Fig. 12. Comparison of Aggregated Group Means of Scanpath Lengths. The differences are statistically significant with P<0.002.

Fig. 14. Aggregated Group Mean Fixation Duration. The differences between the two visualization groups are statistically significant with p<0.01.

5.2. Information Processing

Figure 13 presents an overview of the distribution of the average fixation durations that were generated for each participant. We found that graphs had less variation of average fixation durations. Figure 14 presents the aggregated group means, which shows that the group who used graphs had shorter fixation durations on average. Since participants who used graphs generally spent less time interpreting visualization components, this result indicates that participants were consistently more efficient at processing information with less variations using graphs.
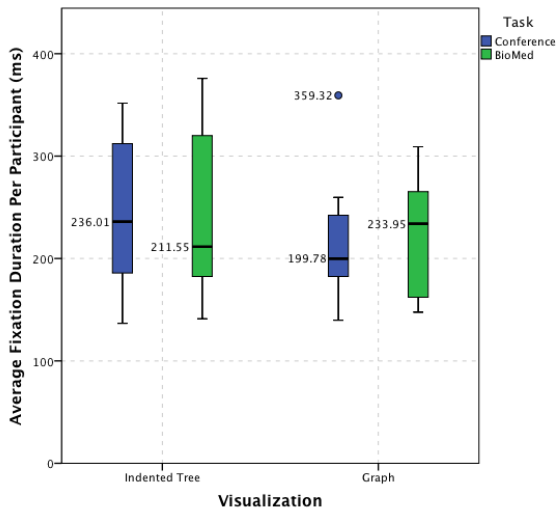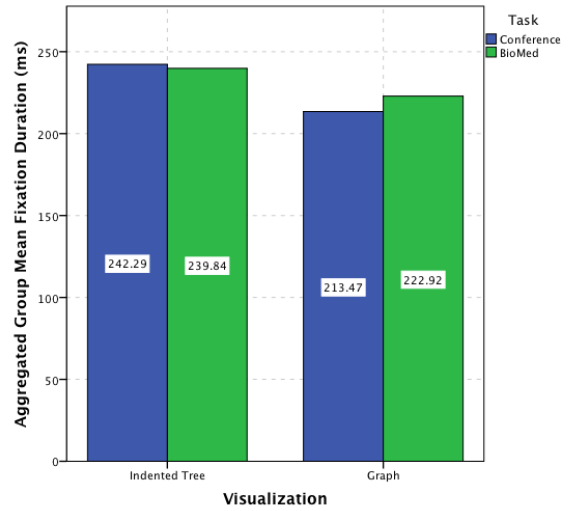
When comparing the sum of fixation durations, we did not find statistically significant differences between the user groups. An averaged search-to-process ratio of 1.88 was found for the participants who used indented lists, and an average ratio of 1.83 was found for those who used graphs, although this difference is not statistically significant either. These results show inconclusive findings, however, since both visualizations yield ratios larger than 1.0, we can conclude that regardless of the visualization used, all participants spent more time searching for information than processing information (as otherwise, the ratios would have been smaller than 1.0).

5.3. Cognitive Workload

When comparing pupil dilations, we found an average increase in pupil sizes from approximately 3.6077 mm to 3.6923 for participants who used indented lists, and to 3.6568 mm for participants who used graphs. However, the difference of these changes, 4.99% for indented list users and 4.93% for graph users, are not statistically significant. It was not clear whether one particular visualization technique was less demanding to use than the other.

When comparing absolute and relative angles, we found that those participants who used graphs consistently produced less varied absolute angles, as shown in Figure 15. In addition, the variations are always smaller for those who used graphs, as shown in Figure 16. However, we did not find statistically significant differences in the sums of relative angles. As discussed earlier (see section 4.3), sum of relative angles should be understood in the context of dispersions of absolute angles, these results are therefore inconclusive. It is not clear whether cognitive workload was reduced using one particular visualization technique.

Fig. 13. Distribution of Average Fixation Duration. The Y-axis illustrates the average fixation duration of a participant during a recording.
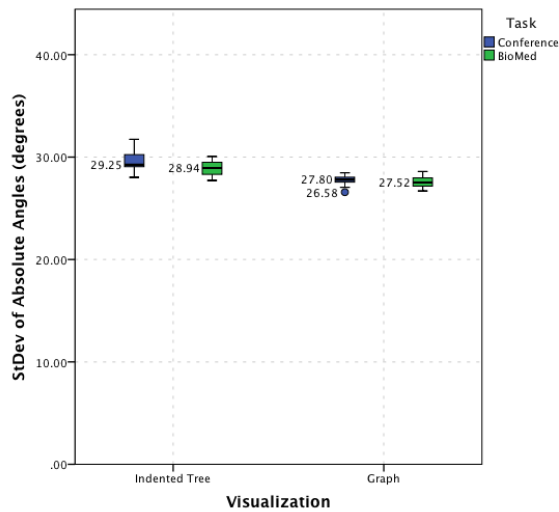
Fig. 15. Distribution of Standard Deviations found in Absolute Angles. For each participant, the standard deviation indicates how dispersed were her (or his) absolute angles.
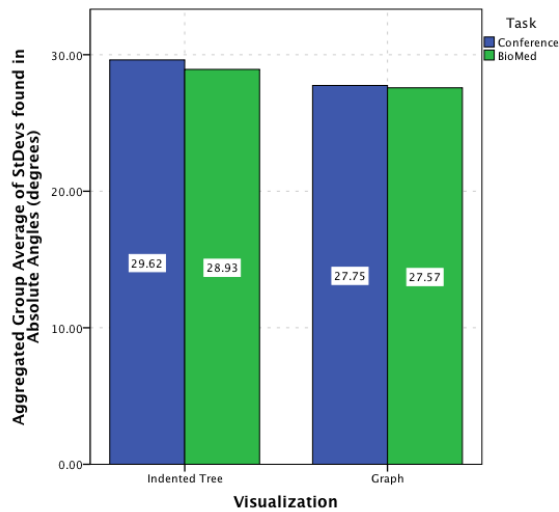

Fig. 16. Comparison of Aggregated Group Means of Standard Deviations found in Each Participant's Absolute Angles. The differences are statistically significant with P<0.0001.

5.4. Participant Speed & Accuracy

We found that participants who used indented lists were consistently faster when completing the tasks. As shown in Figure 17, there were fewer variations in task time for those who used indented lists, and the aggregated group means were always shorter, as shown in Figure 18.

However, participants were not necessarily more accurate with their tasks when using either visualization, as we did not find statistically significant differences in the correctness success (i.e. identifying an pre-existing incorrect mapping), completeness success (i.e. create a correct mapping that was previously missing in the spreadsheet), the overall success (i.e. taking both identification and creation success into account) or error rates. These results

indicate that although task accuracy was not improved, those who used indented lists were faster at completing the given tasks.
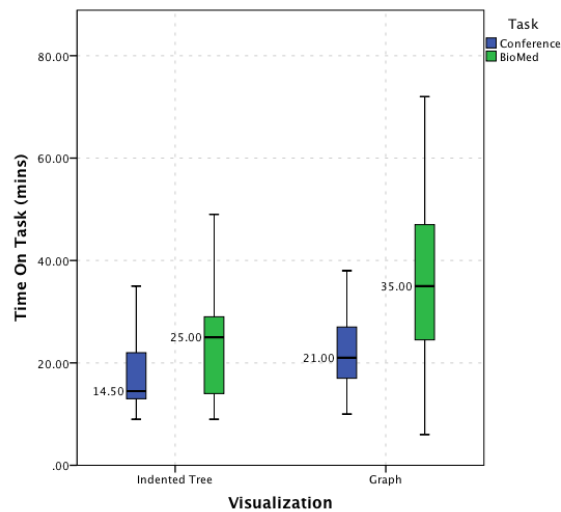

Fig. 17. Distribution of Time on Task. We recorded the time each participant needed in order to complete a given task.
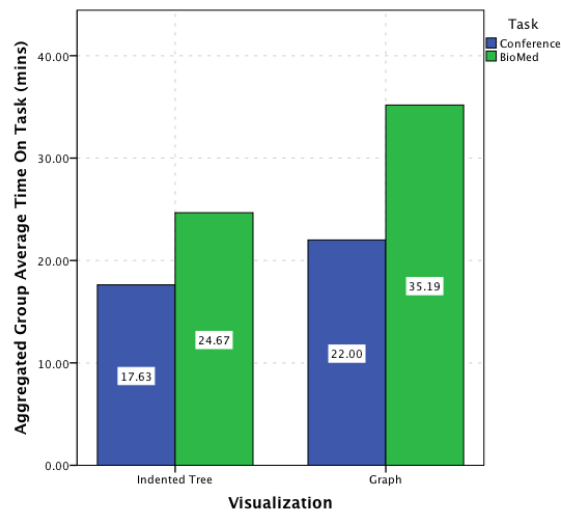

Fig. 18. Comparison of Aggregated Group Mean Task Time. These differences are statistically significant with P<0.02.

5.5. Correlations

We found several moderate to high correlations, some of which are to be expected. For instance, fixation count is highly correlated with scanpath length (with R=0.967, P<0.0001) and the sum of fixation duration (with R=0.898, P<0.0001). Theses results are not surprising, considering more fixations will most likely increase their duration sum and add to the length of a scanpath. In this section, we present two most interesting correlations next.

We found a weak negative correlation between the average saccade length and the average fixation duration. As shown in Figure 19, as the participant fixated on

visualization components that were further apart from one another, she (or he) also spent less time fixating on these points of interest. As longer saccade lengths indicate more useful cues that direct one's attention with few interim fixations (discussed in section 4.1), and since Figure 19 shows the correlation to be stronger for graphs, it indicates that graphs were more effective at providing useful cues and supporting information processing than indented lists.
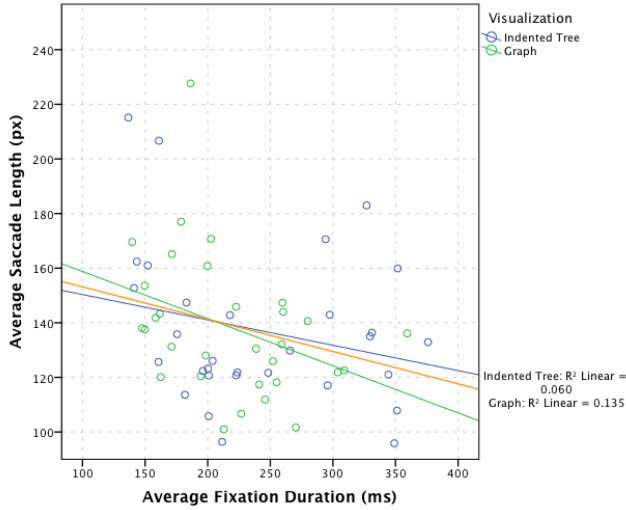

Fig. 19. Correlation between Average Saccade Length and Average Fixation Duration. Each scatter illustrates the average saccade length and the average fixation duration found for each participant. The orange line indicates the overall trend of the correlation between these two variables irrespective of the visualization, where R=-0.291 and P<0.03. The blue and green lines illustrate the correlation of the two variables found in the intended list user group and the graph user group respectively.
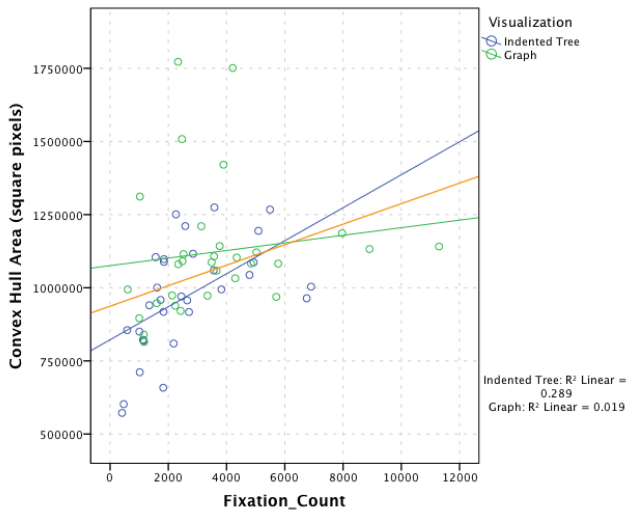

Fig. 20. Correlation between Convex Hull Area and Fixation Count. Each scatter illustrates the convex hull area and the total number of fixations for each participant. The orange line indicates the overall trend of the correlation between these two variables irrespective of the visualization, where R=0.343 and P<0.01. The blue and green lines illustrate the correlation of the two variables found in the intended list user group and the graph user group respectively.

We found that as the number of fixations increased, the convex hull area also increased, as shown in Figure 20. This correlation is stronger for participants who used indented lists. Note that the trend lines of the two visualizations cross at approximately (6000, 1200000), which suggest that for a total fixation count less than 6000, participants focused on a smaller area of interest when using indented lists. For those who had fixation counts greater than 6000, graphs generated smaller areas of interest. These results indicate that typically, graphs led to more widespread and less efficient searches and indented lists were more effective at assisting participants with concentrated and localized searches.

6. Summary & Discussion

The limitations of eye tracking should be taken into account when interpreting the results. For instance, eye tracking does not capture peripheral vision. When the user has obtained information from fixating on a particular point on the screen, it does not mean she (he) had not notice other points of interest nearby through peripheral vision. Also, eye tracking data provides evidence of eye movement patterns, but these results alone do not tell us whether the user is liking what they see or is having trouble processing what they see. Therefore, the results should be understood in the context of the given tasks, and coupled with how participants did in the tasks and what they said about the visualizations. A summary of the results found in our study is presented in Table 2.

As reported in [9], participants found indented lists to be familiar, fast, and easy to use although they can be confusing when redundant nodes are presented for multiple inheritance, and although graphs are connected, controllable and deal with multiple inheritance better, they were also time-consuming, distracting and frustrating to use. The eye tracking results shown in this paper provide further insight and explanation to potential causes for this feedback. For instance, we have evidence indicating that on one hand, it was easier for participants to search information in indented lists (i.e. fewer fixation counts) but harder to process information (i.e. longer average fixation duration) from them. On the other hand, we found that it was harder for participants to search information (i.e. more fixation counts) using graphs, but once found, it was easier for them to process that information (i.e. shorter average fixation duration, and stronger correlation with saccade length) from graphs. In addition, we have evidence suggesting that indented lists make better use of the display (i.e. smaller convex hull areas), which led to more efficient scanpath (i.e. shorter scanpaths). This has made a positive impact on the efficiency of searches in indented lists, which then led to faster task completion times. In comparison, graphs provide more effective mind maps to the participants, who were then more efficient at processing relevant information. Although searching for information was more difficult, which led to inefficient scanpaths and longer task completion times.

Table 2. Result Summary. Results generated from indented lists are illustrated by R(i), and results generated from graphs are illustrated by R(g).

	Measures	Findings	Indications
Information search	Fixation & saccade count	R(i) < R(g)	Participants who used indented lists sampled fewer visualization components during their tasks;
	Convex hull & scanpath	R(i) < R(g)	Participants who used indented lists searched for information within a smaller area and had shorter scanpaths;
	Average saccade length	Inconclusive	It was not clear which visualization provided more helpful cues during information search.
Information processing	Average fixation duration	R(i) > R(g)	On average, participants who used graphs generally spent less time interpreting and processing various visualization components;
	Scanpath duration & search-to-process ratio	Inconclusive	It was not clear which user group had more efficient scanpaths with lower search-to-process ratio.
Cognitive workload	Pupil dilation	Inconclusive	It was not clear if one visualization was less difficult or demanding to use than the other;
	Sum of relative angles & StDev of absolute angles		It was not clear if one user group was more certain and confident with their scans than the other.
Indirect measures	Participant speed	R(i) < R(g)	Participants who used indented lists were faster at completing their tasks;
	Participant accuracy	Inconclusive	It was not clear if one user group was more successful at the tasks than the other.

Overall, regardless of the visualization used, all participants spent more time searching for information than processing information during the tasks. This provides motivation for implementing appropriate search support in any ontology visualization. Within the context of the two specific visualizations included in the study, this is particularly true for graph visualizations.

7. Conclusions & Future Work

We found a key difference in the information search and information processing behaviours of the two commonly used ontology visualization techniques through tracking eye movements. This motivates further research to exclusively evaluate aspects of these visualizations in terms of their support for information search activities and information processing activities. For example, the tasks in the experiments can be designed to exclusively focus on either looking up specific information, or understanding a specific part of the ontology.

Recall that Figure 19 shows the trend lines of the two visualizations cross at approximately (200, 140), which indicates that if given visualization components located no further than 140px, graphs are more likely to be more effective at assisting users with processing information; and if given points of interest that are located further than 140px, indented lists are more likely to be more effective at processing information. These hypotheses can be further tested with fixed positions of relevant visualization components. The findings could potentially improve the physical layout in both visualization techniques for better support of information processing.

Similarly, recall Figure 20 shows that the majority of the scatters fell below 6000 fixation counts, which indicates that participants would be more efficient at information search using graphs if given a convex hull area that is less than 1.2

million px^2. Likewise, given a convex hull area of greater than 1.2 million px^2, indented lists should be more effective at information search. These hypotheses can be validated given restricted convex hull areas. The findings from these experiments could provide useful design guidelines to improve the usage of available screen space for both visualizations.

Future research can also expand to include additional types of ontology visualizations and distinct user groups (e.g. ontology authors vs. readers, novice vs. expert users), which may reveal differences in search and processing behaviours, and provide clarification for the requirements needed for different user groups and different visualization techniques. Furthermore, there were several inconclusive results (as shown in Table 2), which could be potentially clarified with larger sample sizes, additional domains, and larger ontologies.

Finally, eye-tracking experiments are sensitive to the specific technical and hardware configurations. The reliability and validity of the eye-tracking data remain crucial to the understanding and interpretation of the evaluation results. As eye-tracking equipment and software continue to improve, this form of evaluation will remain an important and objective instrument to the advances of ontology visualization techniques.
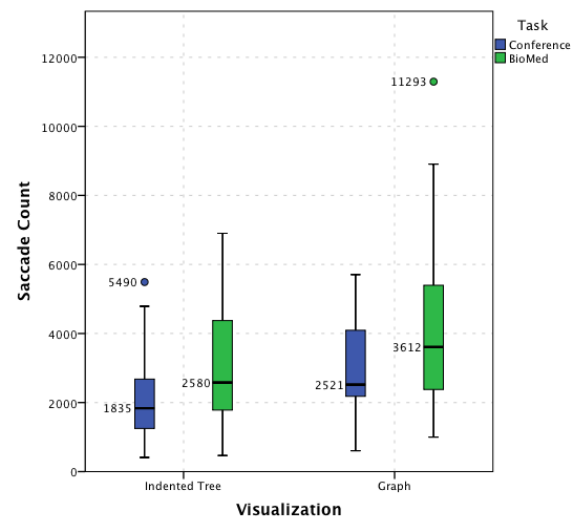
References

[1] A. Bosca, D. Bomino, P. Pellegrino, OntoSphere: more than a 3D ontology visualization tool, in Proceedings the 2nd Italian Semantic

Web Workshop, Trento, Italy, CEUR Workshop Proceedings, ISSN 1613-0073, 2005

[2] A. Duchowski, Eye Tracking Methodology: Theory and Practice, Springer-Verlag, 2007.

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

[4] A. Katifori, E. Torou, C. Halatsis, G. Lepouras, C. Vassilakis, A Comparative Study of Four Ontology Visualization Techniques in Protégé: Experiment Setup and Preliminary Results, in Processing of the 10th International Conference on Information Visualization, 417-423, 2006.

[5] A. Komlodi, A. Sears, E. Stanziola, Information Visualization Evaluation Review, ISRC Tech. Report, Dept. of Information Systems, UMBC. UMBC-ISRC-2004-1

[6] A. Poole, L. J. Ball, Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future, C. Ghaoui (Ed.): Encyclopedia of Human-Computer Interaction. Pennsylvania: Idea Group, Inc, 2005

[7] A. Villanueva, R. Cabeza, S. Porta, M. Böhme, D. Droege, F. Mulvey, D5.6 Report on New Approaches to Eye Tracking. Summary of new algorithms. Communication by Gaze Interaction (COGAIN), IST-2003-511598: Deliverable 5.6, 2008.

[8] B. Fu, L. Grammel, M.-A. Storey, BioMixer: A Web-based Collaborative Ontology Visualization Tool, in Proceedings of the 3rd International Conference on Biomedical Ontology (ICBO 2012), CEUR-WS Vol.897, ISSN 1613-0073, 2012

[9] B. Fu, N. F. Noy, M.-A. Storey, Indented Tree or Graph? A Usability Study of Ontology Visualization Techniques in the Context of Class Mapping Evaluation. In Proceedings of the 12th International Semantic Web Conference (ISWC 2013), LNCS 8218, 117-134, 2013

[10] B. Lee, C. Parr, C. Plaisant, B. B. Bederson, V. D. Veskler, W. D. Gray, C. Kotfila, TreePlus: Interactive exploration of networks with enhanced tree layouts, In IEEE TVCG Special Issue on Visual Analytics, 2006

[11] B. R. Zeeberg, W. Feng, G. Wang, M. D. Wang, A. T. Fojo, M. Sunshine, S. Narasimhan, D. W. Kane, W. C. Reinhold, S. Lababidi, K. J. Bussey, J. Riss, J. C. Barrett, J. N. Weinstein, GoMiner: a resource for biological interpretation of genomic and proteomic data, Genome Biol., 4: R28, 2003.

[12] B. Shneiderman, Tree visualization with tree-maps, A 2-d space-filling approach. ACMTrans. Graph.. Vol. 11, No. 1, September, 92–99, 1992

[13] B. Shneiderman, The Eyes Have It: A Task by Data Bype Taxonomy for Information Visualizations, in proceedings of the 1996 IEEE Symposium on Visual Languages. IEEE Computer Society, Washington, DC, US, 1996.

[14] B. Suh, B. B. Bederson, OZONE: A zoomable interface for navigating ontology information, in Proceedings of Advanced Visual Interfaces. ACM, 2002

[15] C. A. Ricardo, P. R. G. Luzzardi, C. M. D. S. Freitas, The Bifocal Tree: A technique for the visualization of hierarchical information structures, in Proceedings of Workshop on Human Factors in Computer Systems, Fortaleza, Brazil, 2002

[16] C. E. McCulloch, S. R. Searle, Generalized, Linear, and Mixed Models, John Wiley and Sons, 2000.

[17] C. Forsell, M. Cooper, An Introduction and Guide to Evaluation of Visualization Techniques Through User Studies, In Weidong Huang (Ed.) Handbook of Human Centric Visualization, 285-314, 2012

[18] C. Jeong, A. Pang, Reconfigurable disc trees for visualizing large hierarchical information space, in Proceedings of Information Visualization, 19–25, 1998.

[19] C. M. D. S. Freitas, M. S. Pimenta, D. L. Scapin, User-Centered Evaluation of Information Visualization Techniques: Making the HCI-InfoVis Connection Explicit, In Weidong Huang (Ed.) Handbook of Human Centric Visualization, 315-336, 2012

[20] C. Plaisant. The challenge of information visualization evaluation. in Proceedings of the working conference on Advanced visual interfaces (AVI '04). ACM, New York, NY, USA, 109-116, 2004.

[21] C. Plaisant, J. Grosjean, B. B. Bederson, SpaceTree: Supporting exploration in large node link tree, design evolution and empirical evaluation, in Proceedings of IEEE Symposium on Information Visualization, Boston, 57–64, 2002.

[22] D. L. Rubin, D. A. Moreira, P. P. Kanjamala, M. A. Musen, BioPortal: a web portal to biomedical ontologies, in Proceedings of the Symbiotic Relationships between Semantic Web and Knowledge Engineering, AAAI, 74–77, 2008.

[23] E. B. Huey, The psychology and pedagogy of reading. New York: Macmillan, 1908

[24] E. Kleiberg, H. Van De Wetering, J. J. Van Wijk, Botanical visualization of huge hierarchies, in Proceedings of the IEEE Symposium on Information Visualization, IEEE Computer Society Press, 2001

[25] E. Motta, P. Mulholland, S. Peroni, M. d'Aquin, J. M. Gomez-Perez, V. Mendez, F. Zablith. A novel approach to visualizing and navigating ontologies, in Proceedings of the 10th international conference on The semantic web - Volume Part I (ISWC'11), Lora Aroyo, Chris Welty, Harith Alani, Jamie Taylor, and Abraham Bernstein, Eds., Vol. Part I. Springer-Verlag, Berlin, Heidelberg, 470-486, 2011.

[26] E. Pietriga, IsaViz: A Visual Authoring Tool for RDF, http://www.w3.org/2001/11/IsaViz

[27] F. Van Ham, J. J. Van Wijk, Beamtrees: Compact visualization of large hierarchies, in Proceedings of the IEEE Conference on Information Visualization. IEEE CS Press, 93–100, 2002

[28] G. G. Robertson, J. D. Mackinlay, S. K. Card, Cone Trees: Animated 3D visualizations of hierarchical information, in Proceedings of the CHI'91 Human Factors in Computing Systems. ACM, New York, 189–202, 1991.

[29] H. Alani, TGVizTab: An ontology visualization extension for Protégé, in Proceedings of Knowledge Capture, Workshop on Visualization Information in Knowledge Engineering, Sanibel Island, Florida, 2003.

[30] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, S. Carpendale, Seven Guiding Scenarios for Information Visualization Evaluation. Technical Report 2011-992-04, University of Calgary, 2011

[31] J. Carriere, R. Kazman, Interacting with huge hierarchies: Beyond cone trees, in Proceedings of Information Visualization, 74-81, Oct. 1995

[32] J. H. Goldberg, J. I. Helfman, Comparing information graphics: a critical look at eye tracking. In Proceedings of the 3rd BELIV'10 Workshop: Beyond time and errors: novel evaluation methods for Information Visualization (BELIV '10). ACM, New York, NY, USA, 71-78, 2010.

[33] J. H. Goldberg, J. I. Helfman, Eye Tracking on Visualizations: Progressive Extraction of Scanning Strategies. W. Huang (ed.), Handbook of Human Centric Visualization, 337- 372, 2014

[34] J. H. Goldberg, X. P. Kotval, Computer interface evaluation using eye movements: methods and constructs, International Journal of Industrial Ergonomics, Volume 24, Issue 6, October 1999, 631-645, ISSN 0169-8141

[35] J. Lamping, R. Rao, The hyperbolic browser: A focus + context technique for visualizing large hierarchies, J. Visual Lang. Comput., vol. 7, 33–55, 1996

[36] J. McGrath, Methodology Matters: Doing Research in the Social and Behavioural Sciences. In: Readings in Human-Computer Interaction: Toward the Year 2000, Morgan Kaufmann, San Francisco, 1995.

[37] J. Rekimoto, M. Green, The Information Cube: Using transparency in 3D information visualization, in Proceedings of the Third Annual Workshop on Information Technologies and Systems, 125–132, 1993

[38] J. S. M. Lee, G. Katari, R. Sachidanandam, GObar: A Gene Ontology-Based Analysis and Visualization Tool for Gene Sets. BMC Bioinformatics, 2005

[39] K. Andrews, Visual exploration of large hierarchies with information pyramids, in Proceedings of the Sixth International Conference on Information Visualization (IV'02), IEEE Computer Society Press, London, England, 793–798, 2002.

[40] K. Andrews, H. Heidegger, Information slices: Visualizing and exploring large hierarchies using cascading, semicircular discs, in Proceedings of the IEEE Information Visualization Symposium, Carolina. 9–12, 1998.

[41] K. Andrews, J. Wolte, M. Pichler, Information Pyramids[TM]: A new approach to visualizing large hierarchies, in Proceedings of the IEEE Visualization '97, Phoenix, Arizona, 49–52, 1997.

[42] K. Pearson, Mathematical contributions to the theory of evolution. III. Regression, heredity and panmixia. Philos. Trans. Royal Soc. London Ser. A, 187, 253–318, 1896.

[43] K. Rayner, Eye movements in reading and information processing: 20 years of research. Psychological Bulletin 124, 372-422, 1998.

[44] K. T. J. Jankun, L. M. Kwan, MoireGraphs: Radial focus+context visualization and interaction for graphs with visual nodes, in Proceedings of IEEE Symposium on Information Visualization. Seattle, Washington, 20–21, 2003.

[45] K. X. S. Souza, A. D. Dos Santos, S. R. M. Evangeista, Visualization of ontologies through hypertrees, in Proceedings of the Latin American Conference on Human-Computer Interaction, Rio de Janeiro, Brazil. 251–255, 2003

[46] M.-A. Storey, M. Musen, J. Silva, C. Best, N. Ernst, R. Fergerson, N. Noy, Jambalaya: Interactive visualization to enhance ontology authoring and knowledge acquisition in Protégé, in Proceedings of Workshop on Interactive Tools for Knowledge Capture, K-CAP-2001, 2001

[47] M. Eyl, The Harmony Information Landscape: Interactive, Three Dimensional Navigation Through an Information Space, Master's thesis, Graz University of Technology, Austria, 1995

[48] M. Lanzenberger, J. Sampson, AlViz - A Tool for Visual Ontology Alignment, in Proceedings of the 10th International Conference on Information Visualisation, July 5-7, 2006, London, UK, IEEE Computer Science Society, 2006.

[49] M. Pomplu, S. Sunkara, Pupil dilation as an indicator of cognitive workload in human-computer interaction. In Proceedings of HCI International: Vol. 3, 542-546, Mahwah, NJ: Lawrence Erlbaum Associates, 2003.

[50] M. Sintek, Ontoviz tab: Visualizing Protégé ontologies, http://protegewiki.stanford.edu/wiki/OntoViz, 2003

[51] M. Tory, User Studies in Visualization: A Reflection on Methods, In Weidong Huang (Ed.) Handbook of Human Centric Visualization, 411-428, 2012

[52] N. F. Noy, R. W. Fergerson, M. A. Musen, The knowledge model of Protégé-2000: Combining interoperability and flexibility, in Proceedings of the 2nd International Conference on Knowledge Engineering and Knowledge Management, 2000.

[53] P. Hasse, H. Lewen, R. Studer, M. Erdmann, The NeOn Ontology Engineering Toolkit, WWW 2008, April 21–25, 2008.

[54] P. W. Eklund, N. Roberts, S. P. Green, OntoRama: Browsing an RDF ontology using a hyperbolic-like browser, in Proceedings of the First International Symposium on CyberWorlds, Theory and Practices, IEEE press, 405–411, 2002

[55] S. Carpendale, Evaluating Information Visualizations, A. Kerren et al. (Eds.): Information Visualization, LNCS 4950, 19-45, 2008.

[56] R. Côté, F. Reisinger, L. Martens, H. Barsnes, J. A. Vizcaino, H. Hermjakob, The Ontology Lookup Service: bigger and better. Nucleic Acids Res. 2010 Jul; 38 (Web Server issue): W155-60. Epub 2010 May 11.

[57] R. Taylor, Interpretation of the Correlation Coefficient: A Basic Review, JDMS 1:35-39, Jan./Feb. 1990.

[58] SequoiaView, http://w3.win.tue.nl/nl/onderzoek/onderzoek_informatica/visualization/sequoiaview/

[59] S. K. Card, J.D. Mackinlay, B. Shneiderman, Readings in information visualization: using vision to think, Morgan Kaufmann, 1999.

[60] S. Kriglstein, G. Wallner, Human Centered Design in Practice: A Case Study with the Ontology Visualization Tool Knoocks, In: international Joint Conference, VISIGRAPP - Computer Vision, Imaging and Computer Graphics. Theory and Applications Communications in Computer and Information Science Volume 274, 123-141, 2013.

[61] S. Kuhar, V. Podgorelec, Ontology Visualization for Domain Experts: A New Solution, in Proceedings of the 16th International Conference on Information Visualization, 363-369, 11-13 July 2012

[62] S. L. Strasnick, J. D. Tesler, Method and Apparatus for Displaying Data Within a Three-Dimensional Information Landscape, US Patent 5,528,735, Silicon Graphics, Inc., June. Filed 23rd March 1993, granted 18th June, 1996.

[63] S. M. Falconer, C. Callendar, M.-A. Story, A visualization service for the semantic web, in Proceedings of the 17th International Conference of Knowledge Engineering and Management by the Masses (EKAW 2010), LNCS 6317, 554–564, 2010.

[64] S. S. Guo, C. W. Chan., A Comparison and Analysis of Some Ontology Visualization Tools, SEKE, 357-362. Knowledge Systems Institute Graduate School, 2011.

[65] S. Vercruysse, A. Venkatesan, M. Kuiper, OLSVis: an animated, interactive visual browser for bio-ontologies, BMC Bioinformatics, 13:116, 2012

[66] S. Zhong. F. Storch, O. Lipan, M. J. Kao, C. Weitz, W. H. Wong, GoSurfer: A graphical interactive tool for comparative analysis of large gene sets in gene ontology space. Applied Bioinformatics, 3(4): 1–5, 2004.

[67] T. D. Wang, B. Parsia, CropCircles: Topology Sensitive Visualization of OWL Class Hierarchies, In Proc. of the 5th International Conference on Semantic Web, LNCS 4273, 695-708, 2006

[68] T. Liebig, O. Noppens, OntoTrack: Combining browsing and editing with reasoning and explaining for OWL lite ontologies, in Proceedings of the 3rd International Semantic Web Conference, 244-258, 2004.

[69] T. Munzner, Exploring large graphs in 3D hyperbolic space, IEEE Comput. Graph. Appl. Vol. 18, No. 4, 18–23, 1998

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

[71] T. Wang, B. Parsia, Cropcircles: topology sensitive visualization of owl class hierarchies, in Proceedings of the International Semantic Web Conference, 2006

[72] V. Swaminathan, R. Sivakumar, 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, April 2012

[73] W. Rivadeneira, B. B. Bederson, A Study of Search Result Clustering Interfaces: Comparing Textual and Zoomable Interfaces, University of Maryland HCIL Tech. Rep. HCIL-2003-36, 2003

[74] Y. Sure, J. Angele, S. Staab, OntoEdit: Guiding ontology development by methodology and inferencing, in Proceedings of International Conference on Ontologies, Databases and Applications of Semantics, 2002.

Appendix A. Additional Figures

(a) Distribution of Saccade Counts generated Per Participant
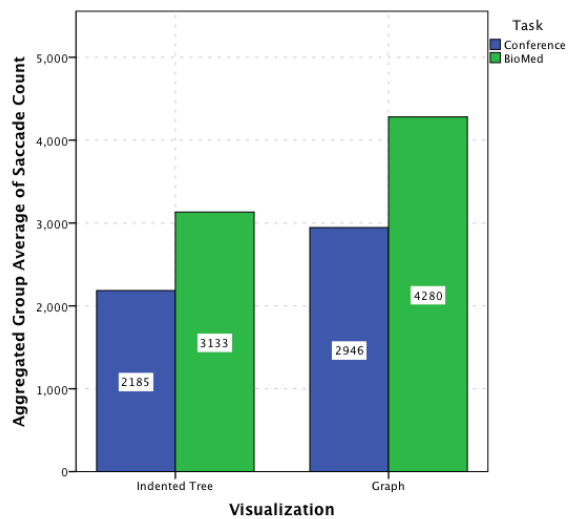
(b) Statistically Significant Differences in the Aggregated Group Means of Saccade Counts, where P<0.05

Fig. A-3. Comparison of Saccade Counts. For each participant, the number of saccades during a task is recorded. The aggregated group means are compared between the users grouped by visualization.