Optimize the user experience of Linked Data visualization

ELA YUDHANIRA
Optimize the user experience of Linked Data visualization

Ela Yudhanira
elayu@kth.se

DA221X Degree Project in Computer Science and Communication, Second Cycle, 30 credits
Master’s Programme in Interactive Media Technology, 120 credits
KTH Royal Institute of Technology year 2018
School of Electrical Engineering and Computer Science (EECS)
Supervisor at EECS was Anna Mavroudi
Examiner at EECS was Olle Bälter
Principal was Scania CV AB
2018-06-18

KTH EECS
Stockholm, Sweden
URL: www.kth.se/eecs
SAMMANFATTNING


NYCKELORD
Länkade Data; Användarupplevelse; Visualisering; Anvandbarhetstestning
Optimize the user experience of Linked Data visualization

Ela Yudhanira
KTH Royal Institute of Technology
Stockholm, Sweden
elayu@kth.se

ABSTRACT
The use of Linked Data to model and visualize complex information entails usability challenges and opportunities to improve the user experience. This study seeks to enhance the user experience of a product information tool which is developed with Linked Data approach. The research was carried out in an industrial setting and follows the case study paradigm. It consists of 1) user research and literature review to define design requirements, 2) prototyping, and 3) usability testing. The user research produced a list of user experience issues which were in turn translated into design requirements by reflecting on related research and following the user’s needs and goals. The design requirements formed the design elements which are embedded into the development of low- and high-fidelity prototypes. Next, usability evaluation of the final high-fidelity prototype examined the extent to which the design decisions could optimize the Linked Data visualization. The results show that several design decisions, such as adding interaction dynamics and the use of rich color representation, could indeed improve it. Also, in terms of general information and visual notation, the shift from UML diagram to node-links diagram received positive feedback from the users. But both node-links diagram and UML diagram received similar scores for the effectiveness and efficiency.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous;

Author Keywords
Linked Data; User experience; Visualization; Usability evaluation

1. INTRODUCTION
Embedded system product contains a lot of information that can be used by different stakeholders from multidisciplinary fields (for example embedded systems, software, industrial, mechanical, and electrical engineering). In a modern engineering environment, this information is saved in many engineering software tools. In embedded systems, all this information need to be integrated to support a consistent product lifecycle development. This integration is enabled by interoperability, a condition that facilitates large information exchange between software engineering tools [12]. To support better software interoperability and achieve enhanced information integration, information modelling is needed.

A Linked Data approach can support software interoperability modelling tool and provide a solution for data integration in software engineering environments [9]. In Linked Data technology, data is shared across applications providing an environment where applications can query the data, make inferences using vocabularies, etc [21]. Thus, Linked Data offers an approach to structure data from different sources, connect and linked them through semantic web technologies. In the semantic web, data is linked according to their semantic relations [4] and this technology supports information integration between different engineering software and can be used by different engineering disciplines. Information modelling for Linked Data requires a representation for describing web resources and metadata in order to explain large-scale information [20]. Visualization of information modelling can be used to help different stakeholders handling the complexity and heterogeneity of embedded systems information. To that end, visualization techniques have been used to contribute to well-integrated engineering environments [12].

This study is a case study on the use of Linked Data conducted at Scania company to model information of the software and the tools in the embedded systems development environment of the company. This research examined the User experience (UX) of Linked Data visualization on the existing modelling tool, Lyo Designer [1]. Lyo Designer is used to model information over the metadata of various tools used by the different Scania stakeholders. Lyo Designer helps tool integration and supports collaboration in the entire development lifecycle such as requirements analysis, design, verification, validation, etc. However, it has been shown that there are several usability problems [25]. For instance, the diagram tends to become too crowded when the toolchain becomes more complicated. Also, the existing visualization does not support detail interaction between different components of the interface [25].

In order to improve the UX of the existing visualization, a user study is conducted at Scania and a prototype is developed
based on the user study results. The main research question examined is:

How to optimize the visualization by improving the UX of a product information tool which is developed with Linked Data approach?

2. THEORY AND RELATED RESEARCH

2.1 Linked Data visualization

To visualize Linked Data it is important to understand its key concepts the most important of which is the semantic web. The definition of semantic web according to [3] is "an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation" (p.34). The idea is to define and link data in a way that it can be used across different applications for more effective integration. The interrelated datasets containing relationships among information in the semantic web is called Linked Data [24]. Linked Data is essential in the implementation of the semantic web that provides semantic annotations and descriptions that facilitate content relationships between information [20]. Visualization technology is contributing in organizing large-scale information resources and provides a way on how semantic web is shaped and represented. In an engineering environment, visualization technology can help people to understand patterns and flows; for example, to explain workflow and tool usage in the development environment [12].

To visualize information modelling of Linked Data, a set of visual representations is needed to describe web resources and metadata information. In the field of Linked Data and semantic web, it is common to take advantage of an ontology to visualize the complex information. In the semantic web, metadata is used to describe information. Metadata is a conceptual structure and can be represented by ontology. According to [14] "ontology" is a term used to represent a set of concepts and their interrelations in a specific domain. An ontology represents a conceptual structure of a domain knowledge and linked heterogeneous information sources [16]. Ontology visualization can represent complex relations between concepts as well as various attributes related to these concepts with the aim to facilitate user understanding of the relevant information and the underlying processes. Katifori [14] categorized six ontology visualization types based on the different characteristics of the presentation, interaction technique, functionality supported, and visualization dimensions. The six visualization types are: indented list, node-link network and tree, zoomable, space-filling, focus+context, and 3D information landscapes. A study [16] shows that graph visualization is the most common technique used to represent semi-structured data such as ontology.

There are a variety of tools for visualizing Linked Data. The Resource Description Framework (RDF) is the most common framework used in Linked Data visualization and provides a standard structure for data interchange on the web and information linking [24]. Graph representation in a form of node-link network is used in GViz [22], ISWIVE [7], OWL-VisMod [11] and VOWL [18]. They are described in the remainder of this section. The tool GViz [22] is a graph visualization tool to visualize large RDF model. ISWIVE [7] is an interactive visualization that presents semantic web in two distinctive frameworks: RDF and topic maps. In ISWIVE, the higher level of information is presented in a node-link network, and the more detail information is presented in a top-down tree by clicking on the corresponding node.

Various techniques for visualizing semantic web have been developed, but only few studies focused on evaluating the UX of existing visualization [16]. One study [10] evaluates the UX of two commonly used ontology visualization techniques, indented list and node-link network graph. The study shows that node-link network graph is better in controlling the diagram and easier to process information compared to the indented list model. On the other hand, node-link network graph is potentially time-consuming, distracting, and frustrating to use especially when the quantity of information is increased.

Another study [11] reveals the main problems in ontology visualization. For instance, the obstruction of visual elements, the lack of interaction techniques, overcrowded visualization, and the number of visual variables used for enriching visualizations. A solution for these problems is proposed in [23] with the use of a Visual Analytics technique. It is focused on the analytical reasoning complemented by the development of interactive visual interfaces. An implementation of this study is OWL-VisMod, a tool to model software engineering ontologies based on a Visual Analytics technique. The study presents better perceived UX and ontology modelling workflow due to the Visual Analytics usage.

In [17], a visual language for Web Ontology Language (OWL) called Visual Notation for OWL Ontologies (VOWL2) is developed to create ontologies. It has been implemented in an ontology editor in a responsive web application called WebVOWL. A qualitative user study reports that VOWL2 can be well understood by casual ontology users and performs better in terms of readability and usability. VOWL2 is developed based on the comparison insights of the previous VOWL version to the UML-based visualization. The components of VOWL2 is built on basic blocks consisting of shapes and colours that express specific aspects of the OWL elements (datatype or object properties, different class and property characteristics, etc.). The graphical representation of WebVOWL provides several interactive features such as zooming, colour highlight, search facility, different size for different levels of nodes, resulting in a usable visualization.

2.2 Evaluation approaches

To improve the UX in a visualization, it is important to evaluate the current visualization and evaluate the proposed prototype. Lam, Bertini, Isenberg, Plaisant, Carpendale [15] identified seven scenarios to evaluate information visualization based on extensive literature reviews of over 800 visualization publications. The scenarios can be used as a practical guideline to evaluate visualizations in different stages of visualization development including pre-design, design, prototype, deployment, and re-design. The study in [15] also provides a wide range of evaluation goals and questions that could be implemented in this master thesis research. OWL-VisMod [11]
presenting an ontology structure in a node-link network is evaluated by comparing the performance of a set of tasks by a set of subjects between the OWL-VisMod and the protégé tool. In the protégé ontology tool, the information is presented in a form of an indented tree. After those tasks, users were asked to answer a questionnaire measuring the user satisfaction using a Likert scale. This evaluation technique is similar to the usability testing method where the subjects are given a set of tasks and questionnaire during the test session [21]. The evaluation method in [11] and [21] presents statistical data for each task and shows an effective way to assess UX in the context of the software engineering ontologies visualization.

In a previous study at Scania [25], a modelling tool for software interoperability with Linked Data was developed to support collaboration between different engineering departments in the company. The modelling tool was developed in an Eclipse environment and is available as an open-source project under the Eclipse Lyo project currently known as Lyo Designer. The tool was developed using the Linked Data approach and was designed based on the prior interview with users. The data between toolchain are linked together with the same unified rules and generated as different diagrams. A usability test with a set of tasks related to the toolchain modelling was conducted accompanied with interview sessions. The evaluation provides a valuable insight on the visualization problems: although the modelling tool provides an easily understandable diagram, efficiency problems appear when the model becomes more and more complex. This was considered as one of the main issues to be solved in the proposed visualization and as a component that affects the UX of Lyo Designer.

3. METHOD

3.1 User research
The first phase was to work towards understanding the design problem. The methods used to collect data in this phase were user interviews and a literature review. The goal of the interviews was to understand users’ work environment and workflow and to identify a set of features for the visualization tool. The participants of the research interviews were the employees at Scania and KTH researchers outside Scania who use Lyo tool or had interaction with Scania development environment before. Two expert users and two non-expert users were selected for the interviews. The expert users were people who use Lyo Designer continuously for more than 3 months and strongly involved in the development of the tool. The non-expert users were people who use Lyo not more than 3 months or uses Lyo Designer not more than 3 times. The interviews were conducted at Scania office in Södertälje and at KTH main campus. As part of the interview, a consent information is presented before the interview. The interviews lasted around 30-45 minutes and were recorded. The interview recordings were translated into transcripts and analyzed manually. To guide the interview, a set of questions regarding visualization evaluation was used inspired by [15]. In addition, questions related to the complex visualization problem were added because this is one of the important issues in the existing visualization.

With respect to the literature review, it involves the prior research on what has been done before in the company related to the development of Lyo Designer. The main industrial papers collected during this phase were [9] and [25]. The notes taken from the literature review were also considered as useful resources to understand user’s work environment and the background of Lyo Designer.

3.2 Define design requirements
The second phase was to collect design requirements and collect ideas for the prototype based on the interview results. Using the same method as in [25], the UX issues found in the interview transcripts were grouped into several thematic categories. The problem findings from the literature reviews were also added to the list. These UX issues findings were used to define design requirements for the proposed design. The design requirements describe what information and capabilities are required by the users. Furthermore, they provide a clear picture of how the visualization should look like, how it should behave, and how it should perform [8].

3.3 Prototyping
After the design requirements were defined, they were translated into design elements following interaction principles and developed into prototypes [8]. I developed two forms of low-fidelity prototypes. The first one was a sketch prototype and the second one was literally a static image made using a graphical design tool. Informal feedback sessions were held towards each of the low-fidelity prototypes. The informal feedback sessions involved discussions with the project supervisors in Scania who work with Linked Data research.

From the low-fidelity prototype, I developed a high-fidelity prototype using web technologies such as HTML5, AngularJS, and D3js. The web technology with D3js library provides an interactive visualization that could address some of the design requirements that was defined before and as such, it was used. The high-fidelity prototype was evaluated regularly at least once in two weeks with the help of the project supervisor. An informal feedback session in the middle phase of the prototype development was also held in one of the weekly meetings in the department where about eight other employees attended the meeting.

3.4 Usability testing
To strengthen the research design, the evaluation involved data triangulation process where the combination of quantitative and qualitative methodologies was used. The quantitative methodology entailed a usability evaluation measuring effectiveness, efficiency, and user satisfaction [21]. Effectiveness was measured quantitatively with error rate to understand whether the product behaves in the way that users expect it to. Efficiency is a measure of time of the quickness with which the user’s goal can be accomplished accurately and completely. The data to measure efficiency was time to complete task. The data from both criteria was collected via observation by taking a note and screen recording when the participants performed a set of representative tasks. Satisfaction refers to the user’s subjective reactions, feelings, and opinions of the system and
was measured through the user feedback with System Usability Scale (SUS) survey [6] as one of the standardised research instrument. SUS survey consists of ten items using a five-point Likert-scale. It can give a global view of subjective assessment of usability. To get a qualitative feedback for the evaluation, interviews were conducted as a part of the usability testing. 

The participants from phase one were invited to the usability test. Besides, several more people from the system engineering department were invited to get a more comprehensive feedback. As a result, three people confirmed the appointment for the usability test. The test took place at Scania office in Södertälje. The testing room was a closed meeting room so there was no interruption for an unnecessary condition. At the beginning of the test, the objective of the study and how the test will be conducted was presented. Participants were presented with two different types of visualization with the same information and were asked to perform six different tasks for each visualization. The tasks were compiled based on the problem findings and the design requirements. The first visualization was a UML diagram presented in Lyo Designer tool and the second visualization was the one depicted in the web application prototype. The visualization was presented on a 27-inch monitor and the participants were asked to do 'think-aloud', a method where the participants were asked to verbalizing their thoughts while performing the task [21]. After the participants performed the tasks, a short interview was conducted. Inspired by the evaluation protocol from [15], a set of questions were prepared. The data recorded during the test were audio recordings, screen recordings, and observation such as time to complete task and number of errors. The time to complete task and number of errors were recorded manually in a paper. The other outputs were SUS survey results and interview transcripts.

4. RESULTS

4.1 Findings from the user research

4.1.1 User’s work practices with Lyo Designer

From the interviews and the literature review, it was concluded that Lyo Designer is used to model metadata information of a software or a system, and to describe properties that are allowed or required in a system. The information in represented in a form of schema diagram. The schema is a Linked Data diagram of software and tool’s properties in a form of UML diagram. The existing interface of Lyo Designer is shown in Figure 1. There are two different user groups using Lyo Designer. The first group refers to toolchain architects; the tool is used to design a schema and to review a schema from different departments. The second group refers to the developers who implement the schema model to develop products. Lyo Designer has embedded functionality that models the schema, but also it is used to view and edit the schema. The schema is used across different engineering roles and different people might interpret the visualization differently. Currently, people commonly discuss the schema either by accessing the Lyo Designer tool or by using the printed version of the Linked Data visualization. Several participants mentioned the importance of showing the information about which organisation units are responsible for a specific information model. A few excerpts from the interviews highlighting this issue are captured below: 

"Because some definition is defined by one group function, and it is important to keep the source as it is."

"...mainly use it to review schemas and vocabularies, different people from different part of the company can discuss common schema and vocabularies"

It is also worth mentioning that Lyo Designer is an open source software so the interviews also involved several people outside Scania who have experience with Linked Data and have used Lyo Designer. For people who are not in Scania, their role is similar to toolchain architect where they use the visualization to help them model the schema between different applications. In Lyo Designer, the information model is structured into domains where they can reflect the organisation units that are responsible to manage the information model [9].

4.1.2 List of UX issues

Several usability problems are mentioned in [25] such as the efficiency issues when the diagram becomes more complex and the need to highlight specific information in a better way. Confirming these issues, several participants mentioned the difficulties to find information especially in a large model. Two quotes below captured this problem:

"...Too many lines, the fact that all the things has relations"

"...(I) overwhelmed by the information on the classes, and I want to focus on the information of the requirement..."

It also emerged from the interviews that a feature that would be useful in a large complex diagram would be enable hiding elements. But unfortunately in the existing visualization, there is no way of showing the current hidden elements and it is difficult for the users to display the hidden elements again. In a printed format, a large diagram is usually printed in a huge paper which is less efficient and it make it less easier to find information. An excerpt below showing this issue:

"...and I want to hide class, lines disappear, links disappear. No indication what’s hidden..."

Other problems were related to the visual and color notation. One participant mentioned that the existing diagram presented in black and white make it less motivating to use the tool. For non-expert users, the representation of Linked Data in a UML diagram is understandable, but after training with the tool’s expert. There were also several opinions from two of the participants that it might be interesting to see another type of visual notation for the Linked Data. Especially the RDF which is the technology behind Linked Data usually presented with circle and arrows.

To summarize all the UX issues, a similar method of listing and grouping used in [25] was conducted. From this content analysis, the UX issues were categorised into four topics. The list is shown in Table 1.

4.2 Design decision

4.2.1 Visualizing Linked Data

To model information in software interoperability, it is important to adopt the right concept of Linked Data in the environment. The key elements that make up Linked Data visu-
Figure 1. Interface of Lyo Designer and the existing visualization in a UML diagram

alization are: resource shapes, properties, and relationships between them. The elements in Linked Data are similar in ontology visualization such as in VOWL [17] where two types of ontology visualization are presented in UML diagram and in a form of a node-link network graph. In [17] the key elements presented in the visualization are: concepts, instances, and relationships between them; similar but not the same with the key elements in Linked Data. The study in [17] presents the advantages and disadvantages of using UML diagram representation. The advantages of using UML diagram include that software engineers are familiarized with it, and that the names for properties and instances are clearly identified because of the rectangle container. Some of the disadvantages are that it is difficult to get an overview of the ontology, labels for the properties are redundant, more cluttered and seems chaotic because there are too many individual elements; it can be time consuming to find information, similar representations for concepts, properties, and instances make it difficult to distinguish different elements. On the other hand, the visualization using node-link graph has more advantages to represent ontologies. The node-link diagram provides a cleaner and more compact overview for structural conceptual data, because there is less redundant information with this type of diagram. The representation used in node-link network graph also offers different notation for concept and instances. It is also found in [17] that users easily can indicate the concept of the ontology in the node-link network compared to the UML diagram.

With respect to using UML diagram as a means of representation for Linked Data, it was also mentioned in the interviews of phase 1, that some participants prefer to see the representation in another kind of visual notation which is more closely representing Linked Data. Although there is no difficulty for the user to interpret and understand the UML representation in the existing tool, considerable time is required for the first-time user to understand it; users need guidance from the expert to explain it to them on their first attempt.

As mentioned in the UX issues findings in Table 1 points 1.1, 1.2, 1.3, 2.1, 2.5, and 2.6, there is a need to model information in a connected cluster by introducing a new notation that is not similar to a UML diagram, and a node-link network is seen as one of the possible representations. It is possible to visualize a node-link network diagram via web technologies by a forced-directed network layout using D3js [5]. The network layout is computed using a spring-embedding (force-directed) algorithm, in which nodes repel each other and edges act as springs [13]. The ability of forced directed network graph to represent RDF, [Table 1 points 1.3, 2.1], which is the central concept behind Linked Data, makes it a strong candidate for
1. General information
1.1. There should be a clear visual representation for the domain grouping
1.2. Domain grouping (see section 4.1.1) is important because it has the capability to represent responsible organisation
1.3. Information about different types of resources is important
1.4. Vocabularies and shapes are two different things
1.5. The feature to define vocabularies before modelling shapes and resources provide a good conceptual thinking
1.6. For the first-time user, it is difficult to understand where to start and what to do first in the tool

2. Visual notation for information
2.1. The elements represented in the diagram do not follow the Object-oriented paradigm (OOP), there should be no misinterpretation between Linked Data and OOP
2.2. Reuse graphical representations from common model reduce learning threshold
2.3. The existing visualization is not capable to display hidden elements
2.4. The importance to see imported and original vocabularies
2.5. Prefer to see shapes in different notation from the UML-like notation
2.6. Overwhelming information in a big model

3. Interaction dynamics
3.1. Show and hide elements feature is useful to reduce cluttered visualization
3.2. Flexibility of layout positioning, to keep the position and relation of elements when the diagram is moved or imported
3.3. Macro to microscopic perspective: provide domain labels when zoom out the screen interface
3.4. Interaction feature such as filtering, selecting, navigating, coordinating elements is useful in finding information
3.5. Drag and drop to create diagram is not intuitive (repeating action)
3.6. Rather than drag and drop, to create vocabulary and specification by filling a form and then automatically generate diagram might be more intuitive.
3.7. Ability to quickly access the vocabularies

4. Color visualisation
4.1. Existing visualization in plain black and white is proven to be less motivating
4.2. Generated random color in different domains to improve the uniformity of the visualization
4.3. Ability to preserve black and white in specific cases (printed materials, confusing interpretation)

Table 1. Findings from user interviews and system analysis

![Table 1](image)

Figure 3. The early visual representation developed in the low-fidelity prototype

As mentioned before, the key elements that make up Linked Data visualization are: resource shapes, properties, and the relationships between them. To identify the best way to represent the elements, a sketch and a static image designed in a graphical design tool were created and presented to the project supervisors.

The early visual representation developed in the low-fidelity prototype depicted in Figure 3. In this phase, several informal feedback sessions were conducted. From the meetings with the two project supervisors in Scania, several refinements were implemented in the high-fidelity prototype. This includes the decision to position literal properties around resource shapes with a different visual notation such as a square. It was also recommended to use a dashed-border for the literal properties because it is a metadata property, a concept, not representing an object. In an informal feedback meeting with eight employees from the systems engineering department (most of them having an engineering background and familiarity with UML diagram), the response from the audience was satisfactory and there were no arguments against the proposed model. It meant that the people are welcoming different notation rather than only use UML notation. Other feedback was around the visualization technology such as several suggestions on how the proximity is conveyed in the network graph. From the discussion with the project supervisors, proximity is conveyed from the relation between nodes. If it belongs in the same cluster then the proximity is closer. But if they belong to a different cluster, the proximity is further. There was a suggestion from the meeting to use the total number of nodes connected in one resource shape to convey the proximity aspect. For example, if the node has a lot of properties, then the line will be getting closer. But if the nodes have a few numbers of connected properties, the link distance is getting further.
Another thing added to the interaction is the ability to highlight the nodes when the mouse is over a particular node. This answer one of the need for more useful interaction features (Table 1 point 3.4). The rationale of providing a feature to highlight connected nodes, is that it tends to be easier for the user to find information and focus on the selected resource shapes or properties. The existing visualization does not provide this interaction feature and from the interviews it was discovered that it is hard to find information in a large-scale diagram. Thus, complexity in a large diagram interface (Table 1 point 2.6) can be alleviated with the implementation of dynamic interaction. The preview of the interactive prototype is shown in Figure 4.

The high-fidelity prototype also contained a feature to show and hide elements which is a feature that is problematic in the existing tool (Table 1 point 3.1). In the existing tool, when the user hides the elements, there is no information about the hidden elements. It is important to keep the user notified and aware of the hidden elements. In the prototype, the user can collapse and expand a domain cluster so it will reduce cluttered information in a large-scale diagram. In addition, there is a clear notation of how many elements were hidden in the selected domain, providing information to the user about the hidden elements. The ability to collapse and expand the domain cluster also provides an opportunity to the user to focus only in one specific domain while maintaining the user’s awareness toward hidden elements in the collapsed domain cluster. In addition, according to usability heuristic by Norman Nielsen [19], a system should always keep users informed about what is going on; the proposed design keep users informed about the selected element and the hidden element(s).

The use of forced-directed network graph supports a flexibility for layout positioning (Table 1 point 3.2). With the interactive web technology, the graph is generated automatically. It is easier and more intuitive for the user rather than clicking on the zoom icon first. This offers user control and freedom, as one of important aspects in usability [19].

4.2.2 Interaction dynamics

The high-fidelity prototype in the form of a web application made it possible to support interaction dynamics to improve the UX of the visualization. In the existing modelling tool, in order to zoom out the screen interface, the users need to click the zoom icon first and then click on the screen to change the perspective. The interaction is limited and makes it difficult to see the macro perspective of the diagram. As required in the findings (Table 1 point 3.3), with the web prototype the user is able to change the viewing perspective by using scroll-up and scroll-down with the mouse interaction. It is easier and more intuitive for the user rather than clicking on the zoom icon first. This offers user control and freedom, as one of important aspects in usability [19].

Another thing added to the interaction is the ability to highlight the nodes when the cursor is over a node area. The visualization highlighted the connected nodes when the user moves the cursor over a node. This answer one of the need for more useful interaction features (Table 1 point 3.4). The rationale of providing a feature to highlight connected nodes, is that it tends to be easier for the user to find information and focus on the selected resource shapes or properties. The existing visualization does not provide this interaction feature and from the interviews it was discovered that it is hard to find information in a large-scale diagram. Thus, complexity in a large diagram interface (Table 1 point 2.6) can be alleviated with the implementation of dynamic interaction. The preview of the interactive prototype is shown in Figure 4.

The high-fidelity prototype also contained a feature to show and hide elements which is a feature that is problematic in the existing tool (Table 1 point 3.1). In the existing tool, when the user hides the elements, there is no information about the hidden elements. It is important to keep the user notified and aware of the hidden elements. In the prototype, the user can collapse and expand a domain cluster so it will reduce cluttered information in a large-scale diagram. In addition, there is a clear notation of how many elements were hidden in the selected domain, providing information to the user about the hidden elements. The ability to collapse and expand the domain cluster also provides an opportunity to the user to focus only in one specific domain while maintaining the user’s awareness toward hidden elements in the collapsed domain cluster. In addition, according to usability heuristic by Norman Nielsen [19], a system should always keep users informed about what is going on; the proposed design keep users informed about the selected element and the hidden element(s).

The UX findings also translated into a list of tasks that was used for the usability testing (Table 2).

4.2.3 Color representation

To answer UX issues about color representations (Table 1, point 4), a rich color representation was also introduced in the prototype. The use of rich color notations will probably contribute to a more uniform look for the resource shapes and properties. Resource shapes and properties in the same domain have the same colors, and this might contribute to the easier identification of domain grouping. As mentioned in Table 1 (point 4.1), the existing visualization only has black and white color, and from interviews in phase 1 it emerged that a visualization with only black and white colors is less motivating. In VOWL [18], it is mentioned that colors help to identify elements and find information in the graphical representation. Although, in one of the interviews it was also mentioned a colorful visualization might not be the optimal solution especially for people with some color impairment. Also, sometimes users want to preserve black and white color in the printed materials. Consequently, feature for the ability to choose to preserve black and white colors is not added to the prototype (Table 1, point 4.2), also because in the discussion with the project supervisors this feature was ranked low as a priority.

The UX findings also translated into a list of tasks that was used for the usability testing (Table 2).

4.3 Usability evaluation

The findings from the usability testing and the notes from the think-aloud protocol show no significant problems related to the tasks or the understanding of the Linked Data information.
This is also supported by the fact that all the participants were familiar with Linked Data. The introduction of new notation was easily comprehended by all the participants. The sections below present the quantitative and qualitative results.

### General information
- **Task 1**: What are the domain and range of property xxx?
- **Task 2**: What class describes resource xxx?

### Visual notation
- **Task 3**: Name all literal property for resource xxx

### Interaction dynamics
- **Task 4**: Shows only domain xxx
- **Task 5**: How many resources are actually in the domain xxx?

### Color visualization
- **Task 6**: Name all the resources belongs to domain xxx

#### Table 2. List of tasks for the usability test

#### 4.3.1 Quantitative results
The effectiveness aspect is operationalized with the time to complete tasks. Overall, a big difference in task completion time between the original Lyo Designer version and the proposed prototype was found in Task 4 and Task 5 (see Table 2); and a slightly faster performance for the prototype was found in Task 6. While in the other tasks, completion times are the same or more than the Lyo Designer.

From looking at the graph in Figure 6, there was no significant difference (less than 1 sec) in task completion time for Task 1 and Task 3. It shows that in terms of finding information and recognizing visual notation, the user easily understands the node-link network graph. The changes from UML diagram to node-link networks show no significant problems. Although in Task 3, users were faster in finding literal properties in Lyo Designer because the literal properties in Lyo Designer are presented directly inside the box of the selected resource shapes (similar to a class-attributes relation in OOP). While in the prototype, although the users easily understand the meaning of dotted-bordered-box as literal properties, they need a little more time to find all literal properties connected with the selected node.

The same pattern was also found in Task 6, where there was only 0.33 sec difference between the Lyo Designer tool and the prototype. It was faster for the user to name all the resources after color notation was given to the model. But it shows no significant difference and since the users were more familiar with Lyo Designer, they can also easily interpret each domain in it. This good result for the prototype is also supported with the addition of highlighting selected resources by giving a different color in the selected node and the links related to the node. The highlight makes it easier to focus on the selected node and it helps finding information faster.

In Task 2, there was a larger difference between the two visualizations. In the prototype, to see the class that describes a resource shape, the user needs to click in a node and the information will be displayed in the floating box on the right side of the screen. Most of the users were not aware that they were able to click the nodes to see more information about the selected node. This might be because they were not familiar with the new prototype. In Lyo, it has the same interaction. When the user clicks a resource, all information will be displayed at the bottom part of the screen.

In terms of interaction dynamics, the prototype performed better than Lyo. It can be seen in Figure 6, Task 4 and Task 5 were done faster in the prototype. This was supported by the ability to collapse and expand the domain easily by clicking on the domain area. Compared to Lyo Designer, the user needs to click several times before they can show or hide domains. In addition, in the prototype, there is a clear information about what is hidden in the collapsed domain. While in the Lyo Designer, the users were not able to see any information about hidden elements. In the number of errors graph in Figure 7, one participant was not able to complete these tasks using the Lyo Designer. This might mean that the interaction of collapsing and expanding domains and showing clear information about hidden elements can increase both the effectiveness and the efficiency. In terms of efficiency, Figure 7 shows that participants can successfully completed all other tasks.

The satisfaction aspect is evaluated in this research work with the SUS scores. The prototype received a score of 79.73. According to [2], a score of 79.37 can be considered as good.
The score for each of the SUS questions is presented in Figure 8. The odd number shows positive statements and the even number shows negative statements. In Figure 8, the green cells represent a high satisfaction score in the corresponding question, and the red cell means the number shows a low satisfaction score. By looking at Figure 8, for the positive statements, question number 5 and 9 are around the middle score between strongly agree and strongly disagree. In number 5, some participants felt confused about how they defined a well-integrated system. A participant gave a low score of two on this point because he thought the description of each resource and properties in the link relation is missing. In the post-test interview, the participants also mentioned that it will be more helpful if they could modify the visualization. With respect to the statement in number 9 in the SUS questionnaire, several participants mentioned that they can easily understand each of the notation but they need time to get used to it. While for the even number which demonstrated negative statements, there was no significance score which was above 3. Overall, it can be concluded that the prototype was well received by the participants who responded positively in a satisfactory level.

4.3.2 Qualitative results
From the think-aloud notes and the interview transcripts, the most important comments were the absence of the reference names that explain the relations between resources. Another important note was the need to put the namespace field and the URI field in the domain title. Right now the prototype only shows the domain name, several participants mentioned that the importance of the namespace information as it is a crucial concept of Linked Data.

Users also found the prototype is intuitive and easy to use. They mentioned the interaction which has more capabilities to make it more comfortable to use. For example, the interaction of showing and hiding the element was repetitively mentioned as their favorite implementation. They found it difficult to do the same activity in the existing tool. Another interaction dynamic that participant found very useful is the ability to zoom in and zoom out only using the mouse scroll. In the existing visualization, users need to do several clicks to be able to zoom the view. Of course, this change improves the UX and the efficiency of interacting with the visualization.

The users also found the visual representation of the prototype as being more beautiful and felt more motivated than the existing diagram. The use of circle and flexible positioning make it more intuitive. The way the diagram presented with automatic positioning also make the users feel positive about the visualization. Several positive keywords mentioned during the test, such as "beautiful", "attractive", "interactive", "intuitive", shows that the prototype received a positive response and raised user’s expectation towards the visualization of Linked Data.

Overall, in the results presented above, the most significance improvements were the addition of interaction dynamics and the use of color representation. It is interesting to note that the introduction of node-link graph to represent Linked Data received positive response but not showing a significance difference when it comes to effectiveness and efficiency.

5. DISCUSSION

5.1 Research method
The interview sessions and the literature review pertained to the existing system provided knowledge about the user experience issues and adequate understanding in order to propose design solutions for the visualization. User research helps to understand how users make use of and interact with the tool in order to model information and what are the important visual characteristics of the existing tool. It is also worth to mention that there was a difficulty to find a proper interview protocol for the first phase on the topic of visualizing Linked Data. As mentioned before, there are various technique for visualizing semantic web but only a few are focusing on the UX aspect. Therefore, interview protocol which is used to evaluate a set of features from a visualization [15] and interview protocol to understand user’s goal from [8] were used as inspiration. Although the sources from [15] is not exactly evaluating visualization for Linked Data information, the result from the interview sessions discovered the design issues and was used as an important input for the next phase of the research.

The literature review as a part of user research was also proven to be very useful to understand the development of the existing tool and the prior usability research conducted. Following the method in [8], literature review was done in parallel or before the interviews. The literature review also helped to set a context for developing questions to the participants. The output from this literature review contributed to the compilation of UX issues. Previous usability research in [25] provides initial UX problems that were matched with the findings from the user interviews. This method contributed to an efficient interviews strategy as well because it is easier to understand the interviews when you know the context beforehand.

5.2 Visualization prototype
In order to answer the research question, the design requirements then translated into prototype so it can be validated with an evaluation test. The use of high fidelity prototype makes it possible to present important UX improvements which were difficult to present in the low fidelity form. For example, the interaction dynamics is easily delivered to the user using the web application and D3js technology.

The use of forced-directed layout allows the ability to represent RDF graphical notation which is commonly used to describe Linked Data. The notation also similar with the Linked Data visualization in GViz [22], ISWIVE [7], and VOWL [18] where a node-links graph was used to represent Linked Data information. The change from using a UML-like diagram to new visual notations give a new perspective on the information presented in the visualization. It stimulates the users to not think about object-oriented modelling but instead building a correct perspective about Linked Data. Although the result shows the effectiveness and efficiency score for both diagrams are similar, this might be caused by the fact that all of the participants were familiar with UML diagram and use Lyo Designer frequently.

The prototypes produced in the design process changed continuously after several informal feedback sessions. But the
main idea to bring a forced-directed network as the final visualization did not change through the iteration. Although the forced-directed network concept received good feedback from the users, there is also a critique towards the use of new visual notation for representing information model in Lyo Designer. The critique is that the current use of UML diagram might be better to use among developers although it might be misinterpreted as an object-oriented model. It is because the form of UML is commonly used among developers and in a software development environment.

As mentioned in section 4.2.2, the interaction dynamics might provide a lot of helps in a large diagram. With the feature of collapsing and expanding, the users were able to focus only on the selected domain and reduce cluttered space on the screen. In terms of general information, the node-links diagram received positive feedback although from the quantitative data emerged that both UML and node-links representations received similar scores for effectiveness and efficiency. Furthermore, the use of new visual notation did not reduce the user’s ability to find information, and the addition of colour grouping makes it easier for the users to comprehend the whole picture of the Linked Data information.

The use of web technology as the media to develop the high-fidelity prototype also received positive responses. As mentioned in section 4.1.1 and 4.1.2, the schema generated from the diagram is used by different organisation units and the sharing issues also appeared in the interviews in phase 1. People use the schema to discuss and collaborate, and the most common way is by having the printed version. By having one more layer above Lyo Designer in a form of a web application, it opens the possibility to have a communal platform for everyone so it can be accessed easily and can be easily integrated with other existing interoperability tools which is also in a form of a web application.

5.3 Usability evaluation

To discover the important design decisions that could improve the UX of Linked Data visualization, the use of usability testing towards the prototype was an effective method. More useful data can be generated from the evaluation phase by the use of quantitative data and qualitative data together in the testing. It was also found that it was very useful to ask the users to do think aloud while they were doing the tasks. In addition to the interview transcripts, a lot of insightful comments were gathered from the think-aloud notes.

The test plan and test procedure were showing an effective result. One improvement was made during the first usability test. It was quite difficult for the first participant to operate the laptop without a mouse. Thus, started from the second test, a mouse was prepared for the test. The use of screen recorder was really helpful for the interpretation phase. Unfortunately several screen recordings were not successfully generated because of the software problem. To overcome this issue, a note of observation was directly made after the interview.

5.4 Future work and research

As mentioned before in section 4.1.1, in the setting of embedded system environment in Scania, people also use Lyo Designer to model information, not only viewing and finding information. This research highlight only the viewing and finding information part. There should be further research in the creation part, which is to model Linked Data information. During the interview sessions in phase 1, several participants mentioned the difficulty they encountered in the creation part. Such as the drag and drop interaction which is not intuitive and the need to make it more efficient. Another finding also the absence of help and tutorial for the first-time users. The first time users find it difficult to know what should they do first on their first contact with the interface. In order to improve a more comprehensive user experience, there should be more analysis according to this matter.

There is an opportunity to do a deeper research in the area of Linked Data visualization. This research was done with a rationale of improving the existing visualization in Lyo Designer. But Linked Data is used anywhere and it is a potential technology helping a better data integration in the industry. A wider audience for Linked Data in industrial settings could be explored for the user research to find out different needs and different user persona. There should be more research regarding the absence of standard visual notation for Linked Data; many people have developed different interpretations of Linked Data. It could be interesting to see one standard visual notation for Linked Data visualization so people can adapt and implement the standard in their organisation.
6. CONCLUSION

The research question this study set out to answer was how to optimize the visualization by improving the UX of a product information tool which is developed with Linked Data approach. In order to answer this question, this study was conducted in four phases: user research, define design requirements, design prototype, and usability testing. The methods used in this study shows an effective way to answer the research question. The early phase of conducting user research was effective to understand the goal and user’s needs. The information then used together with a list of UX issues which also compiled from the interviews and literature review in the first phase. These activities provide valuable resources to define design requirements which were targeted to improve the UX of Linked Data visualization. Finally, the prototyping and the usability test validated the proposed design requirements as the aspects that can answer the research questions. On the other hand, the low number of participants of the usability testing affects the quantitative and qualitative results, but especially the quantitative results. The generalizability and transferability can be unreliable because of the low number. The results presented in this paper provide indications that the proposed solution might actually work.

In conclusion, several design decisions introduced in the new prototype could improve the UX of Linked Data visualization. Four categories were introduced in the UX (see Table 1). The most significant design improvement is the change in the interaction dynamics and the use of rich colour representation. The addition of highlight feature, intuitive zoom interaction, and collapse-expand feature might help in improving the usability of the visualization. Furthermore, the use of rich color notation provide uniformity and consistency of the visualization.

7. ACKNOWLEDGMENTS

I would like to thank all participants in this study as well as Scania and the Systems engineering department for the help and opportunity to work on this project. I would also like to thank my supervisors in Scania, Sofia Cassel and Jad El-Khoury for the valuable knowledge and feedback for this project. I also express my gratitude to my supervisor at KTH Anna Mavroudi for the critical input and my examiner Olle Bälter at KTH for the valuable feedback. I gratefully acknowledge the Swedish Institute Study Scholarships from the Swedish Institute that made it possible for me to pursue this master degree. Lastly, thank you so much for my family back in Indonesia for a never-ending support.

8. REFERENCES

5. Mike Bostock. Data-Driven Documents. (????). https://d3js.org/


