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

Muhammad Rohan Ali Asmat,
Vitalis Wiens,
and Steffen Lohmann

October 8th, 2018
Introduction

- Ontologies are used in scientific and industrial contexts
- Various ontology engineering tools
- Different modeling paradigms
Introduction

- Modeling of ontologies
  - Not limited to ontology engineers
  - Different communities pursuing formal representation of domain knowledge
  - Modeling tools are designed for experts (with profound modeling knowledge)
  - Different modeling paradigms (text input, UML-based graphs, widget and hierarchical based GUI, node-link diagrams and hybrid solutions)
We present an evaluation: comparing node-link diagram and widget-based modeling paradigms (WebVOWL Editor and Protégé)
Introduction

Protégé

Asmat et al.: Comparative User Evaluation
Introduction

WebVOWL Editor
Overview

- Pre-test
- User study
  - Experimental design
  - Results
- Conclusion
Pre-Test

- Definition of five small concept spaces
- Comprising of common, every-day knowledge (university, zoo, city traffic, media, and family tree)
- Concept spaces defined in a tabular form.

Example: **University concept space**

<table>
<thead>
<tr>
<th>Staff Member</th>
<th>Person</th>
<th>Professor</th>
</tr>
</thead>
<tbody>
<tr>
<td>University</td>
<td>Student</td>
<td>Graduate Student</td>
</tr>
<tr>
<td>has name</td>
<td>teaches</td>
<td>Course</td>
</tr>
<tr>
<td>Undergraduate Student</td>
<td>course name</td>
<td></td>
</tr>
<tr>
<td>has*</td>
<td>is a*</td>
<td></td>
</tr>
</tbody>
</table>

* concepts can be used multiple times
Pre-Test

Participants

- Four male participants (without any visual, physical or color blind impairment)
- Profound experience (>2 years) with ontology modeling
- Age range 27–39
Pre-Test

- Participants
  - Four male participants
    - (without any visual, physical or color blind impairment)
  - Profound experience (>2 years) with ontology modeling
  - Age range 27–39

- Measuring cognitive complexity
  - Participants were asked to model all concept spaces in Protégé
  - Training session with university concept space
  - Alternating order for other concept spaces
  - Recorded modeling completion times
Completion times measured in seconds.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Modeling Completion Times</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Family Tree</td>
</tr>
<tr>
<td>A</td>
<td>237</td>
</tr>
<tr>
<td>B</td>
<td>330</td>
</tr>
<tr>
<td>C</td>
<td>389</td>
</tr>
<tr>
<td>D</td>
<td>343</td>
</tr>
<tr>
<td>Sum</td>
<td>1299</td>
</tr>
<tr>
<td>Mean</td>
<td>324.75</td>
</tr>
</tbody>
</table>

Asmat et al.: Comparative User Evaluation
Pre-Test

- Completion times indicate cognitive complexity
- Family tree and city traffic have the lowest mean difference

**Family tree concept space**

<table>
<thead>
<tr>
<th>Child</th>
<th>child's birthplace</th>
<th>Family Tree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>Mother</td>
<td>Grandmother</td>
</tr>
<tr>
<td>Male</td>
<td>gives birth</td>
<td>Father</td>
</tr>
<tr>
<td>Person</td>
<td>person name</td>
<td></td>
</tr>
</tbody>
</table>

**City traffic concept space**

<table>
<thead>
<tr>
<th>Vehicle</th>
<th>Bus</th>
<th>model name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manufacturer</td>
<td>City Traffic</td>
<td>Public Vehicle</td>
</tr>
<tr>
<td>Car</td>
<td>manufactured by</td>
<td>Private Vehicle</td>
</tr>
<tr>
<td>Train</td>
<td>manufacturing date</td>
<td></td>
</tr>
</tbody>
</table>

Asmat et al.: Comparative User Evaluation
Experimental Design

- User Study:
  - Involved 12 voluntary participants
  - Comprised of 9 tasks
  - Duration per participant 45–60 minutes

- Tasks:

<table>
<thead>
<tr>
<th>Task</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>Demographic questionnaire</td>
</tr>
<tr>
<td>T2</td>
<td>Model with Protégé</td>
</tr>
<tr>
<td>T3</td>
<td>After-Scenario-Questionnaire (ASQ) for T2</td>
</tr>
<tr>
<td>T4</td>
<td>Cued Recall Process (highlight concepts in a table)</td>
</tr>
<tr>
<td>T5</td>
<td>Computer-System-Usability-Questionnaire (CSUQ)</td>
</tr>
<tr>
<td>T6</td>
<td>Model with WebVOWL Editor</td>
</tr>
<tr>
<td>T7</td>
<td>After-Scenario-Questionnaire (ASQ) for T6</td>
</tr>
<tr>
<td>T8</td>
<td>Cued Recall Process (highlight concepts in a table)</td>
</tr>
<tr>
<td>T9</td>
<td>Computer-System-Usability-Questionnaire (CSUQ)</td>
</tr>
</tbody>
</table>
Experimental Design

- Demographic questionnaire (T1):
  - Name
  - Age
  - Profession
  - Experience in ontology modeling
  - Experience with Protégé
  - Experience with WebVOWL
  - Any sort of physical impairment
Experimental Design

- Demographic questionnaire (T1):
  - Name
  - Age
  - Profession
  - Experience in ontology modeling
  - Experience with Protégé
  - Experience with WebVOWL
  - Any sort of physical impairment

- User statistics
  - Age range: 25–36
  - No physical or visual impairment
  - One participant was color blind
  - Participants were employees of Fraunhofer IAIS and students of the universities of Bonn and Aachen
Experimental Design

- Based on answers of task T1
  - Users divided into 2 groups
  - $G_1$ contained 6 participants with prior experience
  - $G_2$ contained 6 participants without prior experience

- Training sessions
  - Training with Protégé and WebVOWL Editor
  - Media and zoo concept spaces
  - Approximately 10 min. training for each tool
Experimental Design

- Evaluation setup
  - Dell Precision 3520 laptop
  - Screen size 16”9
  - Resolution 1920 × 1080
  - Protégé version 5.2.0
  - WebVOWL Editor version 0.0.2
Experimental Design

- **Task groups**
  - $T_{G1}$: T2–T5 Protégé related tasks
  - $T_{G2}$: T6–T9 WebVOWL Editor related tasks

- **Counter balancing**
  - Alternating order of task groups
    (starting either with Protégé or WebVOWL Editor)
  - Alternating order of concept space
    (starting either with family tree or city traffic)

<table>
<thead>
<tr>
<th>Concept space \ Tool</th>
<th>Protégé</th>
<th>WebVOWL Editor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family tree</td>
<td>Session 1</td>
<td></td>
</tr>
<tr>
<td>City traffic</td>
<td></td>
<td>Session 2</td>
</tr>
</tbody>
</table>
Experimental Design

Post modeling tasks

- After-Scenario-Questionnaire (ASQ)
  - ease of task completion
  - satisfaction with completion time
  - support of information
  - e.g. :
    “Overall, I am satisfied with the ease of completing the tasks in this scenario.”

- Measured using a Likert scale (1–7)
  1 refers to strong disagreement
  7 refers to strong agreement
Experimental Design

Post modeling tasks

- Cued recall process

Family tree concept space

<table>
<thead>
<tr>
<th>Child</th>
<th>Grandfather</th>
<th>Father</th>
<th>family members</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grandson</td>
<td>child’s birthplace</td>
<td>has kids</td>
<td>gives birth</td>
</tr>
<tr>
<td>Person</td>
<td>Sister</td>
<td>has*</td>
<td>time of birth</td>
</tr>
<tr>
<td>is a*</td>
<td>Male</td>
<td>Aunt</td>
<td>Family Tree</td>
</tr>
<tr>
<td>Grandmother</td>
<td>Daughter</td>
<td>Female</td>
<td>Cousin</td>
</tr>
<tr>
<td>Son</td>
<td>Mother</td>
<td>Uncle</td>
<td>person name</td>
</tr>
</tbody>
</table>
Experimental Design

Post modeling tasks

- Cued recall process

### Family tree concept space

<table>
<thead>
<tr>
<th>Child</th>
<th>Grandfather</th>
<th>Father</th>
<th>family members</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grandson</td>
<td>child’s birthplace</td>
<td>has kids</td>
<td>gives birth</td>
</tr>
<tr>
<td>Person</td>
<td>Sister</td>
<td>has*</td>
<td>time of birth</td>
</tr>
<tr>
<td>is a*</td>
<td>Male</td>
<td>Aunt</td>
<td>Family Tree</td>
</tr>
<tr>
<td>Grandmother</td>
<td>Daughter</td>
<td>Female</td>
<td>Cousin</td>
</tr>
<tr>
<td>Son</td>
<td>Mother</td>
<td>Uncle</td>
<td>person name</td>
</tr>
</tbody>
</table>

### City traffic concept space

<table>
<thead>
<tr>
<th>Vehicle</th>
<th>Motor Bike</th>
<th>Private Vehicle</th>
<th>Transport</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jeep</td>
<td>Bus</td>
<td>car name</td>
<td>manufactured by</td>
</tr>
<tr>
<td>is a*</td>
<td>Brand name</td>
<td>model name</td>
<td>BMW</td>
</tr>
<tr>
<td>Wagon</td>
<td>Car</td>
<td>Suzuki</td>
<td>City Traffic</td>
</tr>
<tr>
<td>Public Vehicle</td>
<td>Horsepower</td>
<td>Manufacturer</td>
<td>Intercity Express</td>
</tr>
<tr>
<td>manufacturing date</td>
<td>has*</td>
<td>Traffic Signals</td>
<td>Train</td>
</tr>
</tbody>
</table>
Experimental Design

Post modeling tasks

- Computer-System-Usability-Questionnaire (CSUQ)

<table>
<thead>
<tr>
<th>effectiveness</th>
<th>efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>satisfaction</td>
<td>discriminability</td>
</tr>
<tr>
<td>guidance</td>
<td>workload</td>
</tr>
<tr>
<td>error management</td>
<td></td>
</tr>
</tbody>
</table>

- CSUQ contains 19 questions
  - “It was easy to learn to use this system.”
  - “I believe I became productive quickly using this system.”
  - “It was simple to use this system.”

- Measured using a Likert scale (1–7)
## Summary

<table>
<thead>
<tr>
<th>Pre-test</th>
<th>User study</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 Participants</td>
<td>12 Participants</td>
</tr>
<tr>
<td>Measuring cognitive complexity</td>
<td>Modeling task</td>
</tr>
<tr>
<td>Only Protége</td>
<td>Protégé and WebVOWL Editor</td>
</tr>
<tr>
<td>All concept spaces</td>
<td>Only family tree and city traffic</td>
</tr>
</tbody>
</table>
Results

- Performance scores (modeling completion time)
- Scores for recall of concepts (highlighting errors)
- Questionnaire scores (ASQ and CSUQ)
Results

Performance scores

- **Protégé**
- **WebVOWL**

- Protégé Experienced Participants
- WebVOWL Experienced Participants
- Protégé Non Experienced Participants
- WebVOWL Non Experienced Participants

Time required to model an ontology (seconds)

Asmat et al.: Comparative User Evaluation
Cued recall scores

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

- Total number of errors for individual tool is 8
- Seven participants were incorrect w.r.t Protégé
- Five participants were incorrect w.r.t WebVOWL Editor
Results

ASQ scores

- Protégé
- WebVOWL

A: Ease of Completing Task
B: Time It Takes to Complete Task
C: Support Information Provided

- Protégé Experienced Participants
- WebVOWL Experienced Participants
- Protégé Non-Experienced Participants
- WebVOWL Non-Experienced Participants

Graph showing ASQ scores for Protégé and WebVOWL in different categories.
Results

CSUQ scores

Asmat et al.: Comparative User Evaluation
Conclusion

- Compared node-link diagrams and widget-based modeling
- Predefined concept spaces analyzed in pre-test
- Experiments included
  - Modeling of ontologies
  - Answering reflective questions
Conclusion

- Results indicate minor difference in mean performance, WebVOWL Editor having slightly better scores.
- Results of the questionnaires (ASQ and CSUQ) indicate the potential of visual modeling being:
  - more efficient
  - supporting mental map creation
  - satisfying more users.
Conclusion

Limitations

- Small sample size
- Small ontology (controlled setting)
- Only small increase in performance

Future Work

- Increased number of participants
- Field study
- Larger ontology
- Controlling prior experience
Thank You

WebVOWL Editor will be presented
in the demo and poster session of ISWC

Give it a try at
http://w3id.org/webvowl/editor

Contact:
Muhammad.Rohan.Ali.Asmat@iais.fraunhofer.de
Vitalis.Wiens@gmail.com
Steffen.Lohmann@iais.fraunhofer.de