

Universidad de Oviedo
Universidá d'Uviéu
University of Oviedo


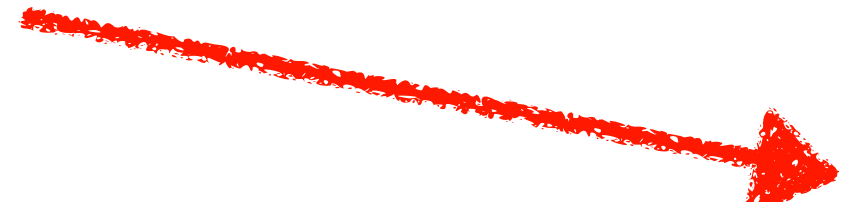

Everything for the Users, Nothing by the Users

Lessons Learnt From a Heterogeneous Data Mapping Languages User Study

Herminio García González - garciaherminio@uniovi.es - @herminio_gg

Introduction

Why usability?

- Declarative mapping rules
 - Faster and more flexible process 
 - Less time and resources
 - Some exiting languages claim to be user friendly
 - YARRRML and ShExML 
 - Others to be easy to learn by semantic web experts 
 - SPARQL-Generate
- Final goal: to ease users' workflow
- Not quantified

Introduction

More on Why usability? The recent change of perspective

- Semantic Web community used to center on new features and technical improvements
 - Historical analogy -> Enlightened Despotism
 - *"Everything for the users, nothing by the users"*
- Recent trends claim to develop more user-centric approaches
 - Understand users
 - Improve their productivity
- A huge analysis tool to decide future actions on a topic

Introduction

Our usability experiment

- **TTBOMK, only our recent study [1] has tackled this aspect for heterogeneous data mapping languages**
- **We briefly summarise and explain it**
- **From its outcomes we envisage next actions in the community**
 - **To better understand and address users' problems**

[1] García-González, H., Boneva, I., Staworko, S., Labra-Gayo, J. E., & Lovelle, J. M. C. (2020). ShExML: improving the usability of heterogeneous data mapping languages for first-time users. *PeerJ Computer Science*, 6, e318.

Brief Experiment Description

Brief experiment description

Language selection criteria

- Languages which goal is to be user friendly
 - SPARQL-Generate, YARRRML and ShExML
- Why not to include RDF-based syntax approaches?
 - Verbosity (solutions are much longer)
- Therefore, similar syntax in terms of verbosity




Not fair and a bias from the beginning!!!

Brief experiment description

Methodology

- Mixed-method approach
 - Quantitative (objective variables measure: behavioural and performance metrics)
 - Qualitative (subjectives variables measure: users' perceptions)
 - Qualitative results can give a better understanding of quantitative results
- 20 students (randomly assigned to languages) of MSc in Web Engineering
- Semantic web course (RDF, SPARQL, ShEx, etc.)
- Task1: Generate mapping rules given inputs and a desired output
- Task 2: Modify the previously generated mapping rules to match a new output




First-time users
with some
background
knowledge

Results & Highlights

Results & Highlights

Statistical results

- Statistical analysis (hypothesis testing) with pair-wise comparisons 
 - Task 1 - Quantitative analysis
 - Significant differences on:
 - Elapsed seconds (ShExML and YARRRML)
 - Completeness percentage and precision (ShExML and SPARQL-Generate)
 - No significant differences on:
 - Keystrokes (no difference in language verbosity)
 - Left & Right button clicks, mouse wheel scroll and meters traveled by the mouse (Similar web playground)

 We'll come back later to this

Results & Highlights

Statistical results

- **Task 1 - Qualitative analysis**
 - **Significant differences on:**
 - **General satisfaction & Easiness of use (ShExML and YARRRML)**
 - **Learnability & Mapping definition easiness (ShExML and both other languages)**
 - **Differences align with quantitative ones:**
 - **Difficulties in SPARQL-Generate -> Worse learnability and mapping definitions easiness**
 - **More time consumed with YARRRML -> Lower levels on general satisfaction and easiness of use**

Results & Highlights

Statistical results

- **Task 2:**
 - **No significant differences due to low sample sizes (6 for ShExML & 1 for YARRRML)**
 - **Modifiability: 5 by 83% of the ShExML users, 3 by the only YARRRML user**
 - **SPARQL-Generate users didn't reach this task due to difficulties to finish the first one**

Results & Highlights

Discussion

- SPARQL-Generate -> Its design is having a bad effect on first-time users -> Difficult to use and learn
- ShExML & YARRRML -> Where's the difference?
 - Hypothesis: Difference in syntax
- Bad scores in the three languages -> Call to action!!!!
 - Language design lead to commit errors
 - Bad error reporting systems
 - No applicability

Actions To Take

Actions to take

Outputs from the experiment

- Take care of how new features are added and designed in languages
 - Avoid bad impact on usability and learnability
- Take care of badly scored variables in the three languages
- Applicability & Learnability on first-time users -> Adoption!!!
- Semantic Web community
 - Focused in new features and technical improvements
 - Need to develop more user-centric approaches (new and recent shift of paradigm)

Actions to take

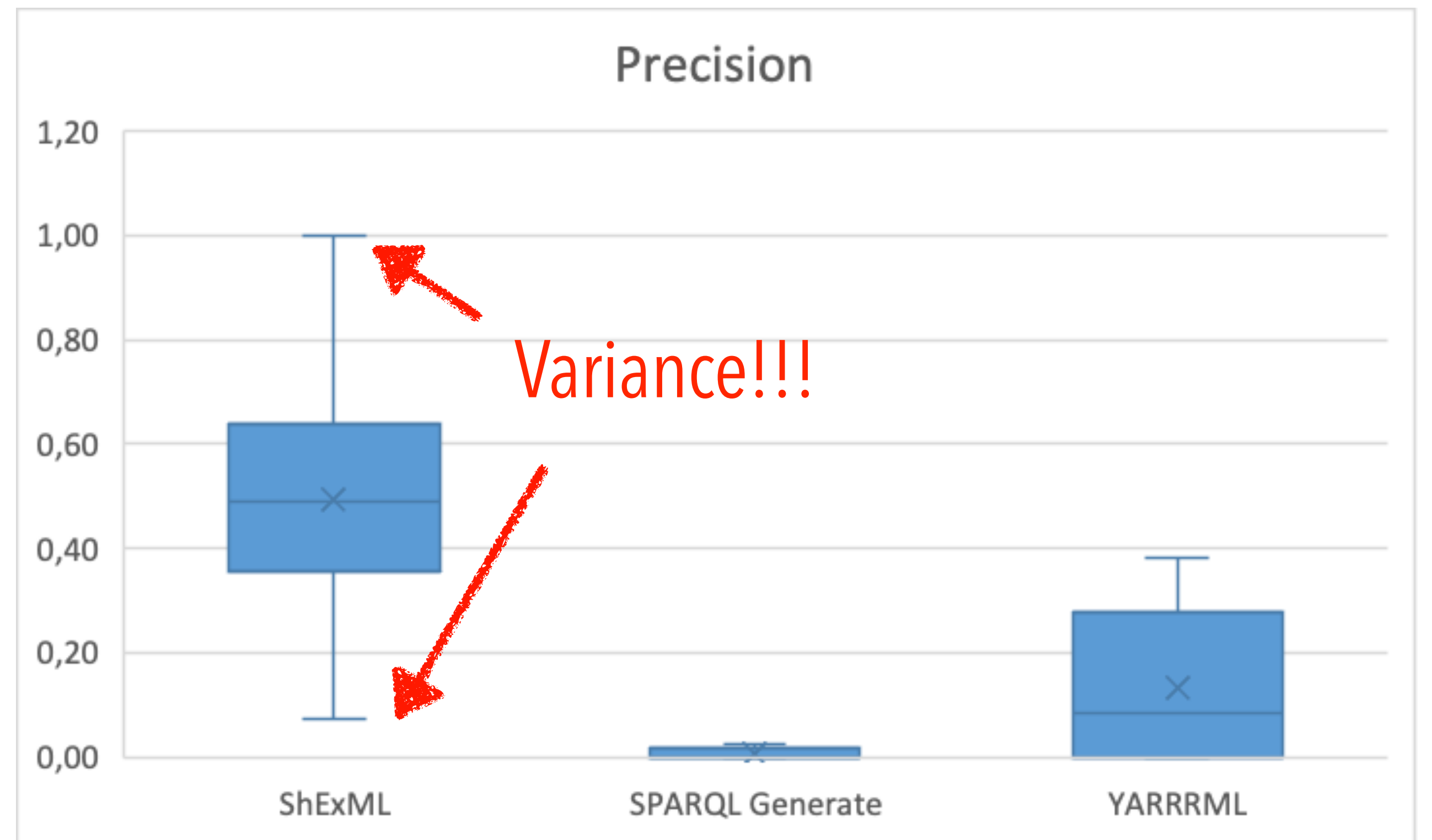
Methodological tools

- Stronger methods to support our hypothesis and conclusions
- In our study -> Statistical hypothesis testing, why?
 - Avoid erroneous conclusions
 - Corroborate that our findings were not obtained just by chance
 - Take into account the variance
 - Measure the evidence strength (effect size)

Actions to take

Methodological tools

- Example from our data
- Precision variable
 - ShExML mean: 0.495
 - YARRRML mean: 0.131
- Intuitively ShExML users are more precise
- Statistically they are not!!!
 - Why?



Actions to take

More studies

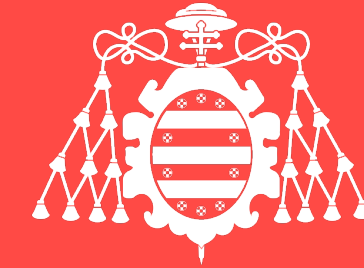
- We only covered first-time users with some background knowledge
 - More profiles -> Whole perspective
- We also have to compare visual and non-visual approaches
 - Discern preferences by profiles
- Differences in syntax
- Experiments that come closer to users' mental processes
 - One possibility: Cognitive models and frameworks
 - Deliver explanations to empirical studies

Conclusions

Conclusions

And take-home lessons

- Focus on users, understand them and take care of their needs (put them in the center)
 - Recent trend in the Semantic Web community
- Example of an heterogeneous data mapping languages study and its outcomes
 - Take care of features design!
 - Do more experiments!
 - Involve users!
 - Use strong methodological and analysis systems (statistics are an ally not an enemy)
 - Learn from other scientific fields



Universidad de Oviedo
Universidá d'Uviéu
University of Oviedo

Everything for the Users, Nothing by the Users

Lessons Learnt From a Heterogeneous Data Mapping Languages User Study

Herminio García González - garciaherminio@uniovi.es - @herminio_gg