

The Efficacy of OWL and DL on User Understanding of Axioms and Their Entailments

Eisa Alharbi, John Howse, Gem Stapleton, Ali Hamie and Anestis Touloumis

University of Brighton

Motivation

- Ontology engineering has become a major activity with many stakeholders involved in producing ontologies.
- The W3C OWL working group devised several different syntaxes.
- The Manchester syntax was created with a view that it “would be easier to write and understand, particularly for non-logicians”. (Horridge et al.)
- “The Manchester syntax is a user-friendly compact syntax for OWL 2 ontologies”. (W3C)

Motivation

- Warren et al. provide insight into the relative efficacy of different Manchester OWL constructs.
- Warren study revealed that users were “prone to certain misconceptions”.
- Rector et al. considered the understandability of OWL.
- Horridge et al. provided insight into the relative cognitive complexity of OWL justifications.

Aim

To provide empirical evidence to support the untested belief that the Manchester syntax is more effective for users than Description Logic.

Tasks: Understanding Axioms and Inference

1. Demon **SubClassOf** Elf
 2. Korrigan **SubClassOf** Demon
 3. Mermaid **SubClassOf** Spirit
 4. Elf **DisjointWith** Nisse
 5. Demon **SubClassOf** hates **only** Goblin
 6. Elf **SubClassOf** chases **some** Spirit
 7. Halfling **SubClassOf** watches **some** Fairy
 8. guides **Domain** Mermaid
- 'All Korrigans are Demons': follows from axiom 2.
 - 'No Demon is a Nisse': follows from axiom 1 and from axiom 4.
 - 'No Halfling is a Spirit': does not follow.

Main Hypotheses

We identify the following hypotheses:

- People more accurately *understand* axioms using MOS than DL.
- People are able to identify *sound inferences* more accurately using MOS than DL.
- People make fewer *unsound inferences* using MOS than DL.
- People perform tasks more quickly *overall* using MOS than DL.

Empirical Study Design

We focused on six axiom types:

Axiom Type	MOS (Protégé 4.3)	DL	Statement Style
Subsumption	C_1 SubClassOf C_2	$C_1 \sqsubseteq C_2$	All C_1 are C_2
Disjointness	C_1 DisjointWith C_2	$C_1 \sqcap C_2 \sqsubseteq \perp$	No C_1 is a C_2
All VF	C_1 SubClassOf p only C_2	$C_1 \sqsubseteq \forall p.C_2$	C_1 p only C_2
Some VF	C_1 SubClassOf p some C_2	$C_1 \sqsubseteq \exists p.C_2$	C_1 p at least one C_2
Domain	p Domain C	$\exists p.\top \sqsubseteq C$	Only C p Things
Range	p Range C	$\top \sqsubseteq \forall p.C$	Things p only C

Designing Questions for the Study

Banshee \sqsubseteq Spirit

Gnome \sqsubseteq Dwarf

Imp \sqsubseteq Troll

Pixie \sqsubseteq Spirit

Gnome \cap Fiend $\sqsubseteq \perp$

Imp \cap Dwarf $\sqsubseteq \perp$

Dwarf $\sqsubseteq \forall$ frightens.Pixie

Elf $\sqsubseteq \forall$ watches.Halfling

Fiend $\sqsubseteq \exists$ scares.Elf

Spirit $\sqsubseteq \exists$ annoys.Fiend

\exists guides.T \sqsubseteq Elf

\exists helps.T \sqsubseteq Elf

T $\sqsubseteq \forall$ chases.Halfling

T $\sqsubseteq \forall$ likes.Banshee

Which of the following statements hold?	
<input type="checkbox"/>	Elves watch only Halflings
<input type="checkbox"/>	Things like only Pixies
<input type="checkbox"/>	Banshees annoy at least one Fiend
<input type="checkbox"/>	No Gnome is a Fiend
<input type="checkbox"/>	Fiends scare at least one Troll
<input type="checkbox"/>	Things like only Spirits
<input type="checkbox"/>	All Gnomes are Dwarfs
<input type="checkbox"/>	No Imp is a Gnome
<input type="checkbox"/>	Things chase only Imps

Understanding Axioms

Banshee \sqsubseteq Spirit

Gnome \sqsubseteq Dwarf

Imp \sqsubseteq Troll

Pixie \sqsubseteq Spirit

Gnome \sqcap Fiend $\sqsubseteq \perp$

Imp \sqcap Dwarf $\sqsubseteq \perp$

Dwarf $\sqsubseteq \forall$ frightens.Pixie

Elf $\sqsubseteq \forall$ watches.Halfling

Fiend $\sqsubseteq \exists$ scares.Elf

Spirit $\sqsubseteq \exists$ annoys.Fiend

\exists guides.T \sqsubseteq Elf

\exists helps.T \sqsubseteq Elf

T $\sqsubseteq \forall$ chases.Halfling

T $\sqsubseteq \forall$ likes.Banshee

Which of the following statements hold?	
✓	Elves watch only Halflings
	Things like only Pixies
	Banshees annoy at least one Fiend
✓	No Gnome is a Fiend
	Fiends scare at least one Troll
	Things like only Spirits
✓	All Gnomes are Dwarfs
	No Imp is a Gnome
	Things chase only Imps

Making Sound Inferences from Axioms

Banshee \sqsubseteq Spirit
Gnome \sqsubseteq Dwarf
Imp \sqsubseteq Troll
Pixie \sqsubseteq Spirit
Gnome \sqcap Fiend $\sqsubseteq \perp$
Imp \sqcap Dwarf $\sqsubseteq \perp$
Dwarf $\sqsubseteq \forall$ frightens.Pixie
Elf $\sqsubseteq \forall$ watches.Halfling
Fiend $\sqsubseteq \exists$ scares.Elf
Spirit $\sqsubseteq \exists$ annoys.Fiend
 \exists guides.T \sqsubseteq Elf
 \exists helps.T \sqsubseteq Elf
T $\sqsubseteq \forall$ chases.Halfling
T $\sqsubseteq \forall$ likes.Banshee

Which of the following statements hold?	
✓	Elves watch only Halflings
	Things like only Pixies
✓	Banshees annoy at least one Fiend
✓	No Gnome is a Fiend
	Fiends scare at least one Troll
✓	Things like only Spirits
✓	All Gnomes are Dwarfs
✓	No Imp is a Gnome
	Things chase only Imps

Making Unsound Inferences from Axioms

Banshee \sqsubseteq Spirit
Gnome \sqsubseteq Dwarf
Imp \sqsubseteq Troll
Pixie \sqsubseteq Spirit
Gnome \sqcap Fiend $\sqsubseteq \perp$
Imp \sqcap Dwarf $\sqsubseteq \perp$
Dwarf $\sqsubseteq \forall$ frightens.Pixie
Elf $\sqsubseteq \forall$ watches.Halfling
Fiend $\sqsubseteq \exists$ scares.Elf
Spirit $\sqsubseteq \exists$ annoys.Fiend
 \exists guides.T \sqsubseteq Elf
 \exists helps.T \sqsubseteq Elf
T $\sqsubseteq \forall$ chases.Halfling
T $\sqsubseteq \forall$ likes.Banshee

Which of the following statements hold?	
✓	Elves watch only Halflings
	Things like only Pixies
✓	Banshees annoy at least one Fiend
✓	No Gnome is a Fiend
	Fiends scare at least one Troll
✓	Things like only Spirits
✓	All Gnomes are Dwarfs
✓	No Imp is a Gnome
	Things chase only Imps

MOS Version

Banshee **SubClassOf** Spirit
Gnome **SubClassOf** Dwarf
Imp **SubClassOf** Troll
Pixie **SubClassOf** Spirit
Gnome **DisjointWith** Fiend
Imp **DisjointWith** Dwarf
Dwarf **SubClassOf** frightens **only** Pixie
Elf **SubClassOf** watches **only** Halfling
Fiend **SubClassOf** scares **some** Elf
Spirit **SubClassOf** annoys **some** Fiend
guides **Domain** Elf
helps **Domain** Elf
chases **Range** Halfling
likes **Range** Banshee

Which of the following statements hold?	
✓	Elves watch only Halflings
	Things like only Pixies
✓	Banshees annoy at least one Fiend
✓	No Gnome is a Fiend
	Fiends scare at least one Troll
✓	Things like only Spirits
✓	All Gnomes are Dwarfs
✓	No Imp is a Gnome
	Things chase only Imps

- We employed a Generalized Estimating Equations (GEE) methodology.
- 60 participants took part in the study (30 per group).
- Total of 3240 accuracy observations.
- Total of 360 time observations.

Understanding Tasks

Statement Style	MOS	DL	p -value	Significant
All C_1 are C_2	90.00%	90.00%	1.00	×
No C_1 are C_2	82.22%	83.33%	0.88	×
C_1 p only C_2	87.78%	92.22%	0.36	×
C_1 p at least one C_2	85.56%	93.33%	0.14	×
Only C p things	84.44%	80.00%	0.48	×
Things p only C	81.11%	83.33%	0.76	×

Sound Inference Tasks

Statement Style	MOS	DL	p -value	Significant
All C_1 are C_2	78.89%	63.33%	0.11	×
No C_1 are C_2	70.00%	56.67%	0.10	×
C_1 p only C_2	67.78%	82.22%	0.06	×
C_1 p at least one C_2	71.11%	80.00%	0.16	×
Only C p things	64.44%	67.78%	0.71	×
Things p only C	58.89%	83.33%	0.00	✓

Unsound Inference Tasks

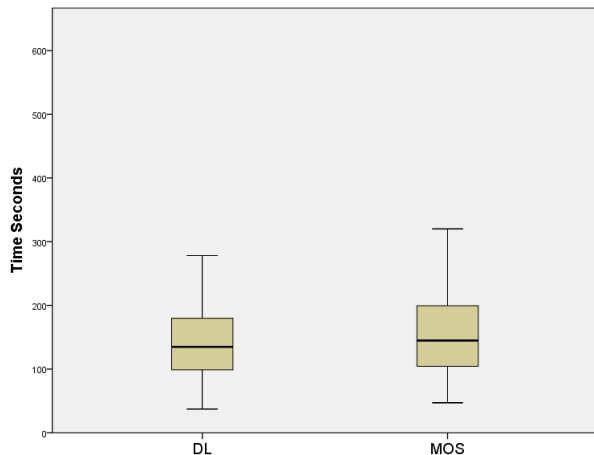
Statement Style	MOS	DL	p -value	Significant
All C_1 are C_2	93.33%	100.00%	0.00	✓
No C_1 are C_2	77.78%	77.78%	1.00	×
C_1 p only C_2	90.00%	90.00%	1.00	×
C_1 p at least one C_2	91.11%	91.11%	1.00	×
Only C p things	90.00%	92.22%	0.62	×
Things p only C	88.89%	81.11%	0.09	×

Time Analysis

DL: 2 minutes 22.46 seconds.

MOS: 2 minutes 37.88 seconds.

p -value = 0.075 (No significant difference)



Discussion

- We hypothesized that participants using MOS would perform significantly better than those using DL due to its textual nature.
- The results of this empirical study are surprising: there were few significant differences between MOS and DL.
- Where there were significant differences, it was DL that performed better.
- This result chimes with Keet who reported that non-English language modellers preferred Protégé v3 with a symbolic DL interface over Protégé v4 using MOS.
- There could be a problem with using natural language in notations: some people may interpret natural language in a reasonable but incorrect way.
- A case of a little knowledge being a dangerous thing.

Conclusion

- Aim: To provide empirical evidence to support the untested belief that the MOS is more effective for users than DL.
- Surprising result: there were few significant differences between MOS and DL .
- Where there were, DL performed better.
- Future work:
 - More complex axioms and tasks.
 - Expert users.