Computing Authoring Tests from Competency Questions: Experimental Validation

Matt Dennis, Kees van Deemter, Daniele Dell'Aglio, Jeff Z. Pan

matt.dennis@port.ac.uk

Competency Questions

- Constructing ontologies is difficult for beginners
- Users and developers frequently fail to comprehend important implications of the information contained in the ontology
- These challenges have led to the notion of a Competency Question (CQ): a question that the finished ontology should be able to answer.
- For example, a CQ for a restaurant domain might ask: "What is the price of asparagus soup?".

CQ-Driven Ontology Authoring

Recently a number of authors have proposed that CQs should become part of the authoring interface.

One approach, which comes to terms with a particularly wide range of CQs, was Ren et al.'s , who proposed CQ-driven Ontology Authoring (CQOA).

An authoring interface checks continually, during authoring, which of the CQs are answered by the ontology.

"Correctly" handling a CQ

We draw on a key concept in linguistics: presupposition

A presupposition is a proposition whose truth is a precondition to assessing truth: if the presupposition does not hold, the sentence is neither true nor false.

For example, the question "What is the price of the cutlery?", when asked in a restaurant, presupposes that cutlery is on sale in that restaurant: if it is not, then the question cannot be answered.

From CQs to Authoring Tests

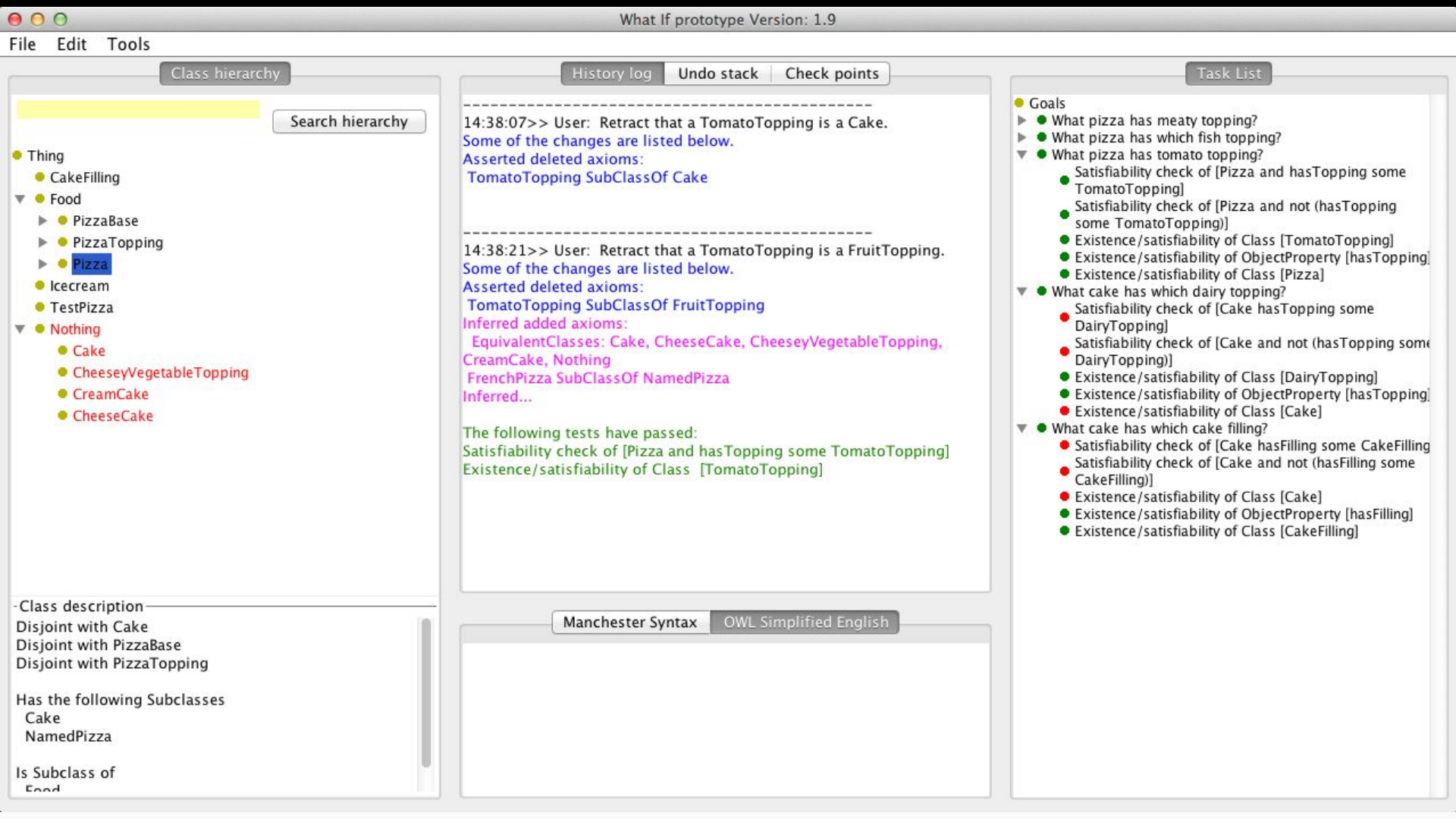
"Which pizzas contain chocolate"

Positive Presupposition: Pizza ⊓ ∃contains.Choc is satisfiable

Complement Presupposition: Pizza ⊓ ∀contains.¬Choc is satisfiable

Ren et al mapped the types of CQ that authors tend to ask and defined archetypes.

Authoring tests can be generated for each archetype


What If prototype Version: 1.9
File
Edit
Tools
Class hierarchy
Search hierarchy
Thing
CakeFilling
Food
PizzaBase
PizzaTopping
Pizza
Icecream
TestPizza
Nothing
Cake
CheeseyVegetableTopping
CreamCake
CheeseCake
Class description
Disjoint with Cake
Disjoint with PizzaBase
Disjoint with PizzaTopping
Has the following Subclasses
Cake
NamedPizza
Is Subclass of
History log
Undo stack
Check points
14:38:07>> User: Retract that a TomatoTopping is a Cake.
Some of the changes are listed below.
Asserted deleted axioms:
TomatoTopping SubClassOf Cake
14:38:21>> User: Retract that a TomatoTopping is a FruitTopping.
Some of the changes are listed below.
Asserted deleted axioms:
TomatoTopping SubClassOf FruitTopping
Inferred added axioms:
EquivalentClasses: Cake, CheeseCake, CheeseyVegetableTopping, CreamCake, Nothing
FrenchPizza SubClassOf NamedPizza
Inferred...
The following tests have passed:
Satisfiability check of [Pizza and hasTopping some TomatoTopping]
Existence/satisfiability of Class [TomatoTopping]
Manchester Syntax
OWL Simplified English
Task List
Goals
What pizza has meaty topping?
What pizza has which fish topping?
What pizza has tomato topping?
Satisfiability check of [Pizza and hasTopping some TomatoTopping]
Satisfiability check of [Pizza and not (hasTopping some TomatoTopping)]
Existence/satisfiability of Class [TomatoTopping]
Existence/satisfiability of ObjectProperty [hasTopping]
Existence/satisfiability of Class [Pizza]
What cake has which dairy topping?
Satisfiability check of [Cake hasTopping some DairyTopping]
Satisfiability check of [Cake and not (hasTopping some DairyTopping)]
Existence/satisfiability of Class [DairyTopping]
Existence/satisfiability of ObjectProperty [hasTopping]
Existence/satisfiability of Class [Cake]
What cake has which cake filling?
Satisfiability check of [Cake hasFilling some CakeFilling]
Satisfiability check of [Cake and not (hasFilling some CakeFilling)]
Existence/satisfiability of Class [Cake]
Existence/satisfiability of ObjectProperty [hasFilling]
Existence/satisfiability of Class [CakeFilling]

Research Question

Is the interpretation of CQs embodied in the mapping from presuppositions to ATs is in accordance with the user's understanding?

Experiments

2 Experimental designs

One used an English Representation of Ontology, CQs and ATs

The other used DL syntax from the same ontology

Costabucks is a hot drinks company. They are creating a robot that can answer questions from customers about the hot drinks that they sell. The robot's programmers have to tell the robot some facts about hot drinks so that it understands enough to answer the questions. To do this, the robot's designers give the robot "rules" about the world.

Once all of the Customer Questions (CQs) can be answered by the robot, its knowledge of the coffee menu is considered complete, and it can be used in the shop.

The programmers are using a special programming tool which allows them to add possible customer questions to its interface, and the tool can inform them when the questions are able to be answered by the current set of rules. To do this, the tool breaks down the questions into several smaller authoring tests, which all must be passed in order for the question to be judged as answerable. The authoring tests are automatically generated by the tool, based on what the customer question is.

Our “Hot Drinks” Ontology

	DL	Non DL
1	$hasContent \circ hasContent \sqsubseteq hasContent$	The robot understands that things can ‘contain’ other things, and that this is transitive. Transitive means that, for example, if flour contains gluten, and a loaf of bread contains flour, then the loaf of bread therefore contains gluten.
2	$Drink \equiv CoffeeDrink \sqcup TeaDrink$	All drinks are coffee drinks or tea drinks
3	$Coffee \sqsubseteq \exists hasContent.Caffeine$	Coffee beans contain caffeine
4	$CoffeeDrink \equiv \exists hasContent.Coffee$	Coffee drinks contain coffee beans
5	$TeaDrink \equiv \exists hasContent.Tea$	Tea Drinks contain tea leaves
6	$CoffeeDrink \sqcap TeaDrink \sqsubseteq \bot$	Nothing can be both a coffee drink and a tea drink at the same time
7	$Cappuccino \sqsubseteq Drink \sqcap \exists hasContent.SteamedMilk \sqcap \exists hasContent.Coffee$	A cappuccino is a drink that contains steamed milk and coffee beans
8	$Americano \sqsubseteq CoffeeDrink$	An Americano is a coffee drink

Participants

Study 1 (English): 54 participants, 54% male. Recruited through crowd-sourcing on Mechanical Turk. Lay audience.

Study 2 (DL). 15 participants, 86% male. Recruited through 12th Reasoning Web Summer School, Aberdeen. Beginner and intermediate DL practitioners (40% beginner, 14% intermediate).

Study 3 (DL). 67 participants, 42% female. Recruited at CCKS2016 conference, China. Beginner and intermediate DL practitioners.

Variables

Independent variable:
Type of Authoring Test

Dependent variable:
***Relevance*: whether a participant judged an AT to be relevant to a given CQ or not.**

Type	DL	English
Occurrence (concept)	CoffeeDrink should occur in the ontology	A coffee drink should be defined
Occurrence (property)	hasContent should occur in the ontology	It must be possible for something to contain something
Relation Satisfiability	CoffeeDrink ⊓ ∃ hasContent.Caffeine should be satisfiable in the ontology	It must be possible for a coffee drink to contain caffeine
Relation Satisfiability (complement)	CoffeeDrink ⊓ ∀ hasContent.(¬TeaLeaf) should be satisfiable in the ontology	It must be possible for a coffee drink to not contain tea leaves

Hypotheses

H1: Occurrence ATs are agreed with more often than disagreed with.

H2: satisfiability ATs that focus on a concept mentioned in a CQ are agreed with more often than disagreed with.

H3: satisfiability ATs that focus on the complement of a concept mentioned in a CQ are agreed with more often than disagreed with.

H4: satisfiability ATs that focus on a concept mentioned in a CQ are agreed with more often than satisfiability ATs that focus on the complement of a concept mentioned in a CQ.

Procedure

- Demographics captured
- Participants read scenario
- Participants were shown the ontology
- Participants were shown ATs, with symbols to show whether the test passed or not
- Shown 7 CQs, one by one.
- Participants were asked to mark which ATs were relevant for each CQ
- Participants could give comments if they wished
- Participants were asked whether they agreed with the answerability judgement of the CQ

Question 2 of 7

CQ2: Which coffee drinks contain caffeine?

The programming tool has generated the following Authoring Tests:

- 2.1: There must be more than one type of drink that contains coffee and tea. ✖
- 2.2: *hasContent* should occur in the ontology ✔
- 2.3: *TeaDrink* ⊓ ¬(∃*hasContent.CoffeeBean*) should be satisfiable in the ontology ✔
- 2.4: *TeaDrink* ⊓ ∃*hasContent.Caffeine* should be satisfiable in the ontology ✔
- 2.5: *CoffeeDrink* ⊓ ¬(∃*hasContent.Caffeine*) should be satisfiable in the ontology ✖
- 2.6: *CoffeeDrink* ⊓ ∃*hasContent.Caffeine* should be satisfiable in the ontology ✔
- 2.7: *Caffeine* should occur in the ontology ✔
- 2.8: *CoffeeDrink* should occur in the ontology ✔

Task 1: For each authoring test, please state whether you think it is relevant to this question or not:

Authoring Test	Relevant	Reason
2.1 There must be more than one type of drink that contains coffee and tea.	○ Yes ◉ No	The CQ does not talk about tea
2.2 *hasContent* should occur in the ontology	◉ Yes ○ No	
2.3 *TeaDrink* ⊓ ¬(∃*hasContent.CoffeeBean*) should be satisfiable in the ontology	○ Yes ◉ No	The CQ does not aks about tea drinks
2.4 *TeaDrink* ⊓ ∃*hasContent.Caffeine* should be satisfiable in the ontology	○ Yes ○ No	
2.5 *CoffeeDrink* ⊓ ¬(∃*hasContent.Caffeine*) should be satisfiable in the ontology	○ Yes ○ No	
2.6 *CoffeeDrink* ⊓ ∃*hasContent.Caffeine* should be satisfiable in the ontology	○ Yes ○ No	
2.7 *Caffeine* should occur in the ontology	○ Yes ○ No	
2.8 *CoffeeDrink* should occur in the ontology	○ Yes ○ No	

Task 2: Do you think that the CQ can be meaningfully answered?

Yes | No

Results - Study 1

H1, H2 and H3 are confirmed with a significant majority of participants agreeing with the generated authoring tests. H4 is not supported

Filler threshold	0%		50%		66%	
AT type	relevant	not relevant	relevant	not relevant	relevant	not relevant
Occurrence (conc)	96% (830)	4% (34) *	97% (482)	3% (14) *	98% (298)	2% (6) *
Occurrence (prop)	91% (296)	9% (28) *	90% (168)	10% (18) *	84% (96)	16% (18) *
Satisfiability (conc)	76% (245)	24% (79) *	82% (152)	18% (34) *	83% (95)	17% (19) *
Satisfiability (comp)	72% (233)	28% (91) *	71% (132)	29% (54) *	72% (82)	28% (32) *

Results - Study 2

H1, H2 and H3 confirmed. However, at a 66% filler threshold, H3 is not supported, with only a small majority of ‘satisfiability of complement of concept’ ATs being marked as relevant. Once again, H4 is not supported.

Filler threshold	0%		50%		66%	
AT type	relevant	not relevant	relevant	not relevant	relevant	not relevant
Occurrence (conc)	95% (227)	5% (13) *	94% (181)	6% (11) *	96% (108)	4% (4) *
Occurrence (prop)	100% (90)	0% (0) *	100% (72)	0% (0) *	100% (42)	0% (0) *
Satisfiability (conc)	78% (70)	22% (20) *	82% (59)	18% (13) *	93% (39)	7% (3) *
Satisfiability (comp)	64% (58)	36% (32) *	62.5% (45)	37.5% (27) *	**55%** (23)	**45%** (19)

Results - Study 3

H1 and H2 confirmed. There is no support for H3: For the filler thresholds of 50% and 66% (representing the DL-logically more capable participants), significant majorities marked these ATs as **non-relevant.** H4 is **supported.**

Filler threshold	0%		50%		66%	
AT type	relevant	not relevant	relevant	not relevant	relevant	not relevant
Occurrence (conc)	83% (893)	17% (179) *	82% (499)	18% (109) *	78% (250)	22% (70) *
Occurrence (prop)	86% (347)	14% (55) *	88% (201)	12% (27) *	79% (95)	21% (25) *
Satisfiability (conc)	70% (280)	30% (122) *	74% (168)	26% (60) *	73% (88)	27% (32) *
Satisfiability (comp)	**50%** (202)	**50%** (200)	**43%** (98)	**57%** (130)	36% (43)	64% (77) *

Summary

We have found that occurrence ATs are almost universally agreed with

We also found broad agreement that the key concept involved in a Selection Question must be satisfiable.

However, when dealing with the complement of such a concept, participants in study 3 did not agree that this had to satisfiable.

Discussion

Post hoc testing shows that the more proficient DL practitioners are, the less they mark the complement of relation satisfiability as relevant (see paper)

Nevertheless, we have evidence that our mapping from CQs to ATs is on the right track

Further experimentation needed to investigate the findings of study 3, and measure the effectiveness of CQOA vs standard ontology authoring.