Fact-checking with explanations

Adrian Groza and Aron Katona

Department of Computer Science
Technical University of Cluj-Napoca, Romania
Adrian.Groza@cs.utcluj.ro

Myth #1: The day in which you were born *guides you to choose a profession*

best suited for your abilities. It also *increases your chances of success* in the career.

Example (Those born on 1st, 10th, 19th, 28th)

The ruling planet for *number 1* is the Sun. They are born *leaders*. They are risk takers with a 'never surrender' attitude, which makes them excel in *business*. Elon Musk and Bill Gates belong to these dates. Business is the best career for number 1 individuals.

Example (Those born on 2nd, 11th, 20th, 29th)

They are *creative* people. The best suitable career for them is diplomacy, arts, acting, fashion designing. Jennifer Aniston and Leornado Di Caprio were born on these dates.

1. Pick classes from YAGO/DBPEDIA, representing succesfull *careers* according to the myth: *leader*, *political leader*, *executive director*, *managing director*.
2. Calculate the proportion p_1 of people born in days 1, 10, 19 and 28 that are part of the chosen classes. Find the sample size n_1.
3. Calculate the proportion p_2 of people born in the other days that are part of the chosen classes. Find the sample size n_2.
4. Compare the proportions. Check the statistical relevance.

How many leaders were born on day 10?

```
1  SELECT COUNT(?person) as ?nr_persons
2   WHERE { ?person yago:wasBornOnDate ?date ;
3           rdf:type ?person_type ;
4           rdf:type ?class .
5           ?class rdfs:label "leader"@eng
6           ?person_type rdfs:label "person"@eng .
7   FILTER (datatype(?date) = xsd:date
8  && fn:month-from-date(xsd:date(?date)) = 7
9  && fn:day-from-date(xsd:date(?date)) = 10)}
```

Career	n_1	n_2	p_1	p_2	
leader	1,660,382	10,441,109	0.00848	0.00832	1.89×10^{-46}
political leader			0.00292	0.00278	6.27×10^{-10}
executive director			0.00122	0.00116	5.25×10^{-5}
managing director			0.00148	0.00145	4.95×10^{-10}
artist	1,546,551	10,554,940	0.004274	0.00437	2.84×10^{-41}
actor			0.003862	0.003834	5.6×10^{-25}
diplomat			0.000361	0.000348	0.00778

Myth #2: The 'life path number' tells us if two persons are compatible

Repeatedly summing all the digits in one's birth date (e.g. 01.02.1998 $\rightarrow$ 2)

Example (LPN=1 compatible 3, 5, and 6)

Number 1's can sometimes be a little bossy and opinionated. 3 and 5 can put up with this sort of character as they have traits that can help them deal with this easier. [. . .]

Example (LPT=2 compatible 6, 8, and 9)

LPT=2, you are a natural peacemaker and will often see all sides of a story. You find it easy to deal with difficult situations too. This makes you very compatible with the number 8, who can sometimes become quite indecisive at times.

1. Get the dates of birth for each married couple in DBpedia
2. Calculate the proportion p_1 of people with $lpn = 1$ which are married to someone with $lpn \in \{3, 5, 6\}$
3. Calculate the proportion p_2 of people with $lpn \notin \{1, 3, 5, 6\}$ which are married to someone with $lpn \in \{3, 5, 6\}$
4. Compare the proportions. Check the statistical relevance

Retrieving the birth dates of each married couple

```
1   SELECT ?date1, ?date2
2   WHERE { ?person1 rdf:type ?person_type;
3             dbo:birthDate ?date1;
4             dbo:spouse ?person2.
5           ?person2 rdf:type ?person_type;
6             dbo:birthDate ?date2.
7           ?person_type rdfs:label "person"@en.}
```

lpn	'*compatible*' *lpns*	n_1	n_2	p_1	p_2	
1	3, 5 ,6	3,316	25,837	0.3543	0.3638	2.06×10^{-13}
2	6, 8 ,9	3,228	25,925	0.3782	0.3767	6.01×10^{-10}
3	1, 5, 7	3,011	26,142	0.3251	0.2996	0.0155
4	1, 8, 7	3,400	25,753	0.3321	0.3507	6.41×10^{-17}
5	1, 3, 7	3,283	25,870	0.3290	0.3340	1.26×10^{-10}
6	1, 2, 8, 9	3,199	25,954	0.4495	0.4432	5.27×10^{-11}
7	3, 5	3,223	25,930	0.2144	0.2132	2.97×10^{-5}
8	2, 4, 6	3,332	25,821	0.3190	0.3293	1.81×10^{-12}
9	2, 6	3,161	25,992	0.2186	0.2127	0.000 418

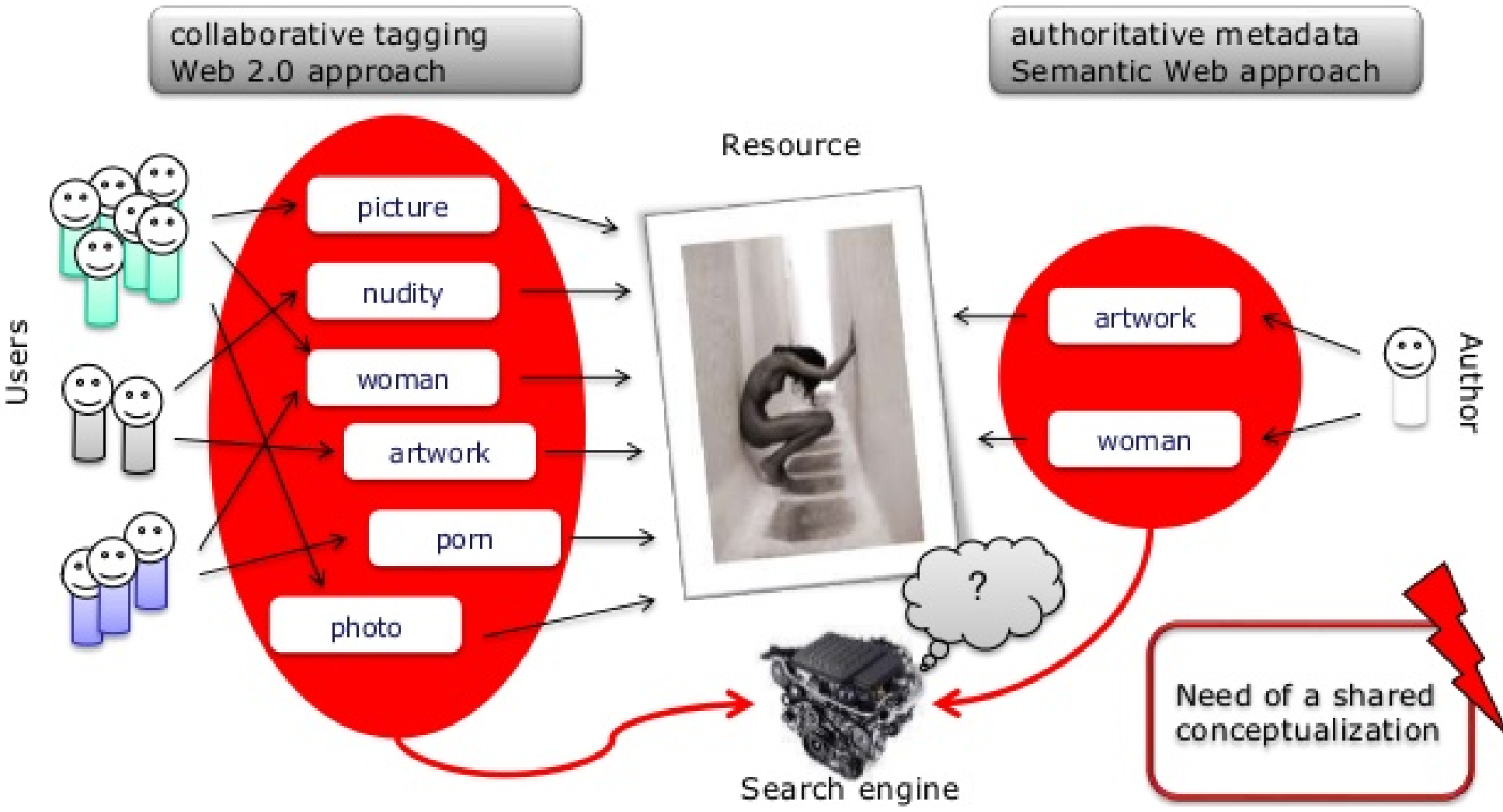
Semantic Web Vision

”The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation.”

Tim Barners Lee

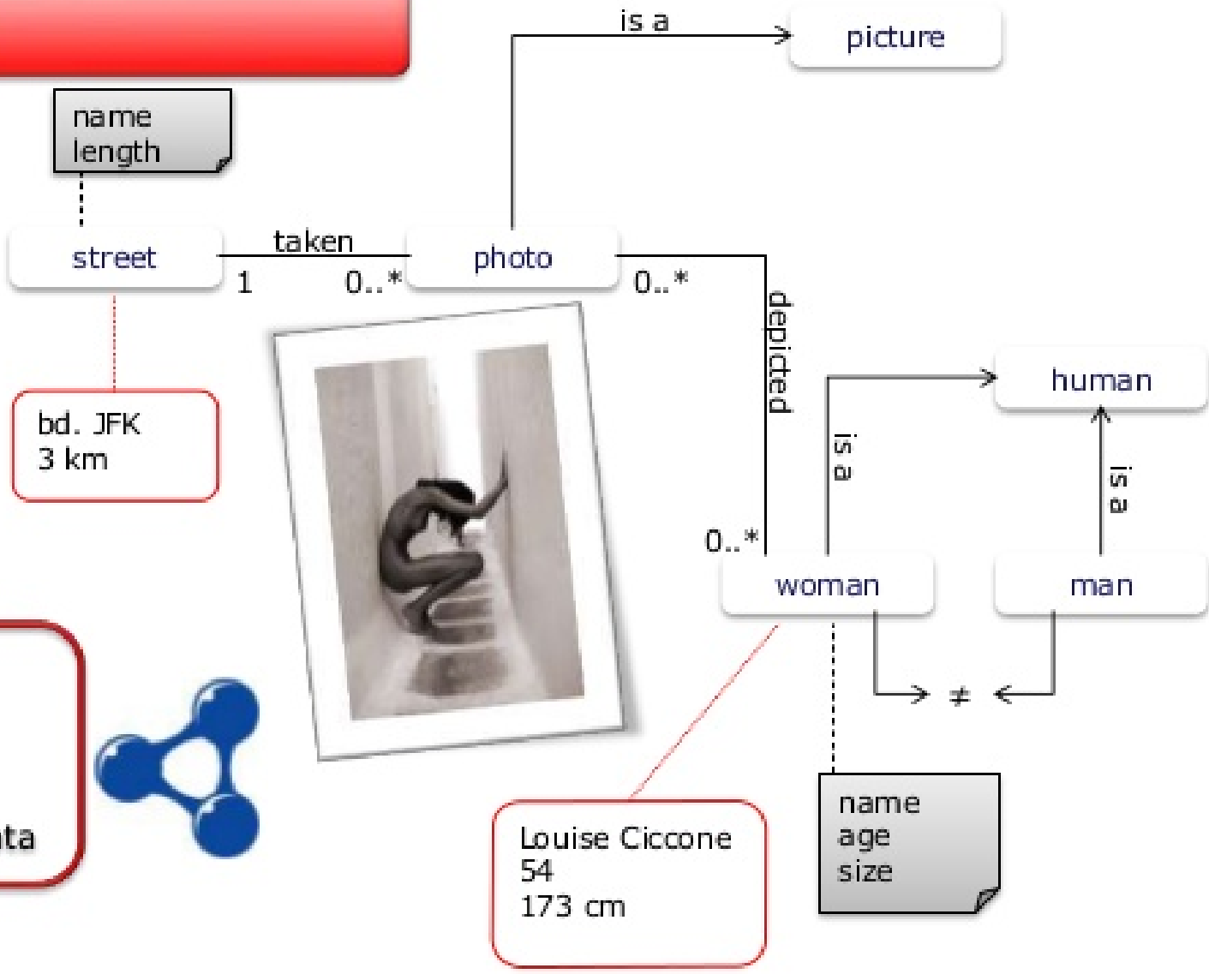
Social Web

Semantic Web


collaborative tagging
Web 2.0 approach
authoritative metadata
Semantic Web approach
Resource
Users
picture
nudity
woman
artwork
porn
photo
artwork
woman
Author
?
Search engine
Need of a shared
conceptualization

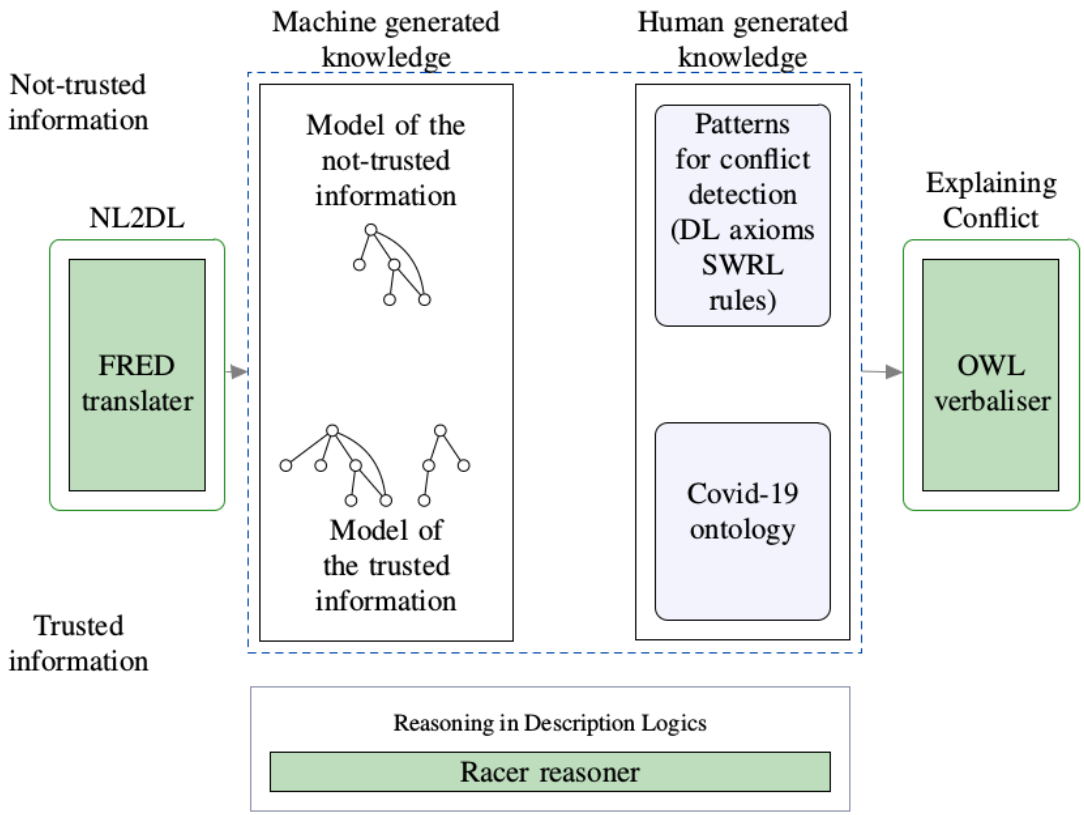
Conceptualization

- concepts
- relations between concepts
- attributes
- instances



Using the same ontology allows two different systems to communicate and to reason over (meta)data

Myth
5G mobile networks spread Covid-19
Exposing yourself to temperatures higher than 25C degrees prevents COVID
Drinking excessive amounts of water can flush out the virus
Regularly rinsing your nose with saline help prevent infection with Covid-19
Eating raw ginger counters the coronavirus
Antibiotics are effective in preventing and treating the new coronavirus
High dose of Vitamin C heals Covid-19
The new coronavirus can be transmitted through mosquito
Covid-19 can affect elderly only



Example (Terminological box)

"Coronavirus disease (COVID) is an infectious caused by a newly discovered coronavirus"

$$COVID \equiv CoronavirusDisease \tag{1}$$

$$InfectiousDisease \sqsubseteq Disease \tag{2}$$

$$CoronavirusDisease \sqsubseteq InfectiosDisease \sqcap \forall causedBy.NewCoronavirus \tag{3}$$

Example (Assertional Box)

$$\text{SARS-CoV-2} : Virus \tag{4}$$

$$hasSource(\text{SARS-CoV-2}, bat) \tag{5}$$

Example (Incoherent ontology)

$$\text{Covid-19} \sqsubseteq InfectionDisease \tag{6}$$

$$\text{Covid-19} \sqsubseteq \neg InfectionDisease \tag{7}$$

Example (Reasoning to detect incoherence)

$$\text{Covid-19} \sqsubseteq InfectionDisease \tag{8}$$

$$InfectiousDisease \sqsubseteq Disease \sqcap \exists causedBy.(Bacteria \sqcup Virus \sqcup Fungi \sqcup Parasites) \tag{9}$$

$$\text{Covid-19} \sqsubseteq \neg Disease \tag{10}$$

Example (OIL antipattern)

$$OIL_1 : Antibiotics \sqsubseteq \forall kills.Virus \tag{11}$$

$$OIL_2 : Antibiotics \sqsubseteq \forall kills.Bacteria \tag{12}$$

$$OIL_3 : Virus \sqsubseteq \neg Bacteria \tag{13}$$

Myth: 5G mobile networks spread Covid-19

$$5G : MobileNetwork \tag{14}$$
$$covid19 : Virus \tag{15}$$
$$spread(5G, covid19) \tag{16}$$

Fact: Viruses can not travel on radio waves/mobile networks

$$Virus \sqsubseteq \neg\exists travel.(RadioWaves \sqcup MobileNetworks) \tag{17}$$

Detecting inconsistencies

$$Virus \sqsubseteq \neg\exists travel.MobileNetworks \tag{18}$$
$$Virus \sqsubseteq \forall travel.\neg MobileNetworks \tag{19}$$
$$Virus \sqsubseteq \forall spread.\neg MobileNetworks \tag{20}$$

Here we need the subsumption relation between roles ($travel \sqsubseteq spread$). The reasoner finds that the individual $5G$ (which is a mobile network by axiom (14)) that spreads $COVID19$ (which is a virus by axiom (15)) is in conflict with the axiom (20).

Myth: Covid-19 can affect elderly only

$$\text{Covid-19} \sqsubseteq \forall affects.Elderly \tag{21}$$

Fact: Covid-19 can affect anyone

$$\text{Covid-19} \sqsubseteq \forall affects.Person \tag{22}$$

Detecting inconsistency: Abox + background knowledge

$$affectedBy(jon, \text{Covid-19}) \tag{23}$$

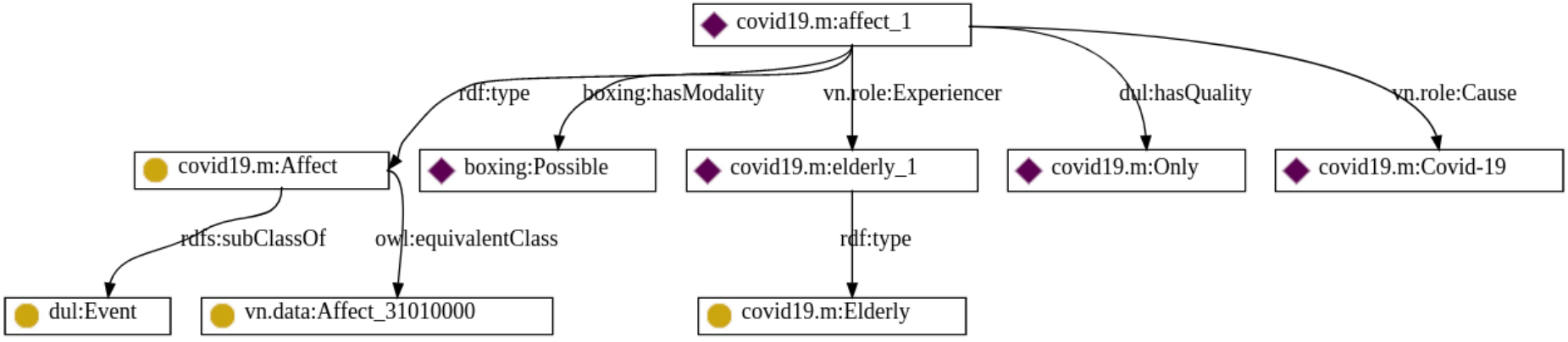
$$hasAge(jon, 40) \tag{24}$$

$$Elderly \sqsubseteq Person \sqcap (> hasAge\ 65) \tag{25}$$

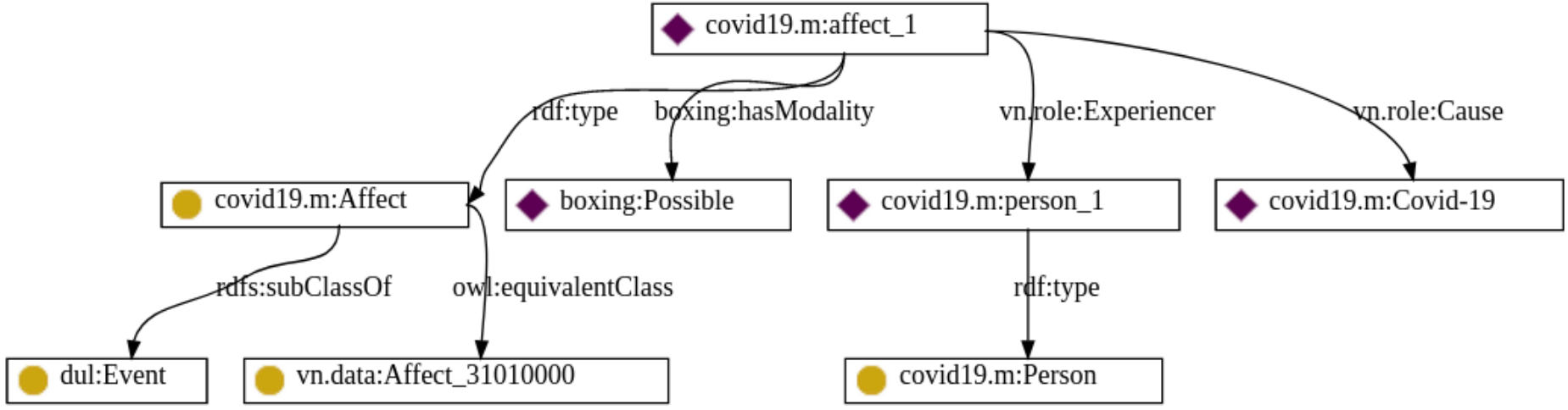
$$affects^{-} \equiv affectedBy \tag{26}$$

$$\text{Covid-19} \equiv one - of(\text{Covid-19}) \tag{27}$$

Step 1: “Covid-19 can affect elderly only“



Step 2: “Covid-19 can affect anyone”



Step 3: Merge ontologies
Step 4: Add background knowledge
Step 5: Detect inconsistency/incoherence
Step 6: Verbalise the inconsistency
Covid-19 $\sqsubseteq \forall affects.Person$ can be learned from data

Test sentences

#	Sentence	Verdict	Comment
1	Viruses are not bacteria.	true	Type Negation Pattern - plural
2	A virus is not a bacterium.	true	Type Negation Pattern - singular
3	Bacteria are viruses.	false	Positive form of (2), incoherent, DS
4	Antibiotics kill viruses.	false	VerbNet Role Pattern: Agent, Patient; inconsistent
5	Bacitracin kills viruses.	false	(4) + WordNet
6	Amoxicillin kills retrovirus.	false	(4) + WordNet
7	Amoxicillin kills retrovirus. Bacteria are not viruses.	false	Composed: (1) + (6), with entailed axioms
8	Covid-19 is caused by SARS-CoV-2.	true	VerbNet Role Pattern: Theme1, Theme2
9	Covid-19 is caused by a coronavirus.	true	(8) + concept matching
10	Covid-19 is caused by a bacterium.	false	(9) + inconsistency
11	SARS-Cov-2 is a virus and a bacterium.	false	and
12	SARS-CoV-2 is a virus or a bacterium or an infection.	true	Type disjunction pattern
13	Vitamin C heals Covid-19.	false	VerbNet Role Pattern, inconsistent
14	Sars-CoV-2 is a virus.	true	
15	Sars-CoV-2 is not a virus.	false	Type Negation Pattern, instance matching
16	Sars-CoV-2 is not a Bacterium.	true	Type Negation Pattern, instance matching

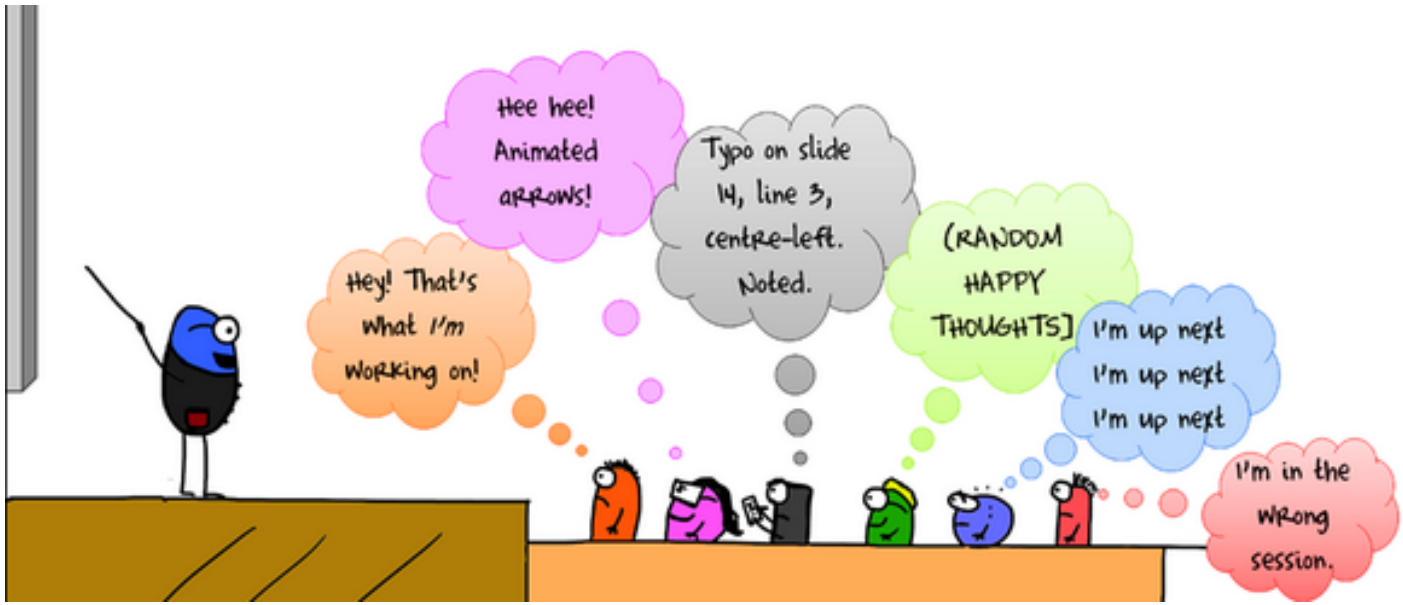
Generating counterspeech

The input text contradicts our trusted sources
The ontology is incoherent: Retrovirus is unsatisfiable
Explanation #1:
    knowledge from input text:
Every Retrovirus is killed by Amoxicillin.
    knowledge from trusted sources (counterspeech):
Amoxicillin is an Antibiotic. (https://wordnet.princeton.edu/)
Every Retrovirus is a Virus. (https://wordnet.princeton.edu/)
Everything that is killed by an Antibiotic is a Bacterium. (https://www.who.
    ↪ int/emergencies/diseases/novel-coronavirus-2019/advice-for-
    ↪ public/myth-busters#antibiotics)
No Virus is a Bacterium. (https://www.healthdirect.gov.au/bacterial-vs-viral
    ↪ -infection)

Some part of the input is entailed by our trusted sources:
Entailed information #1: No Bacterium is a Virus.
Explanation #1: No Virus is a Bacterium. (https://www.healthdirect.gov.au/
    ↪ bacterial-vs-viral-infection)

Conclusion

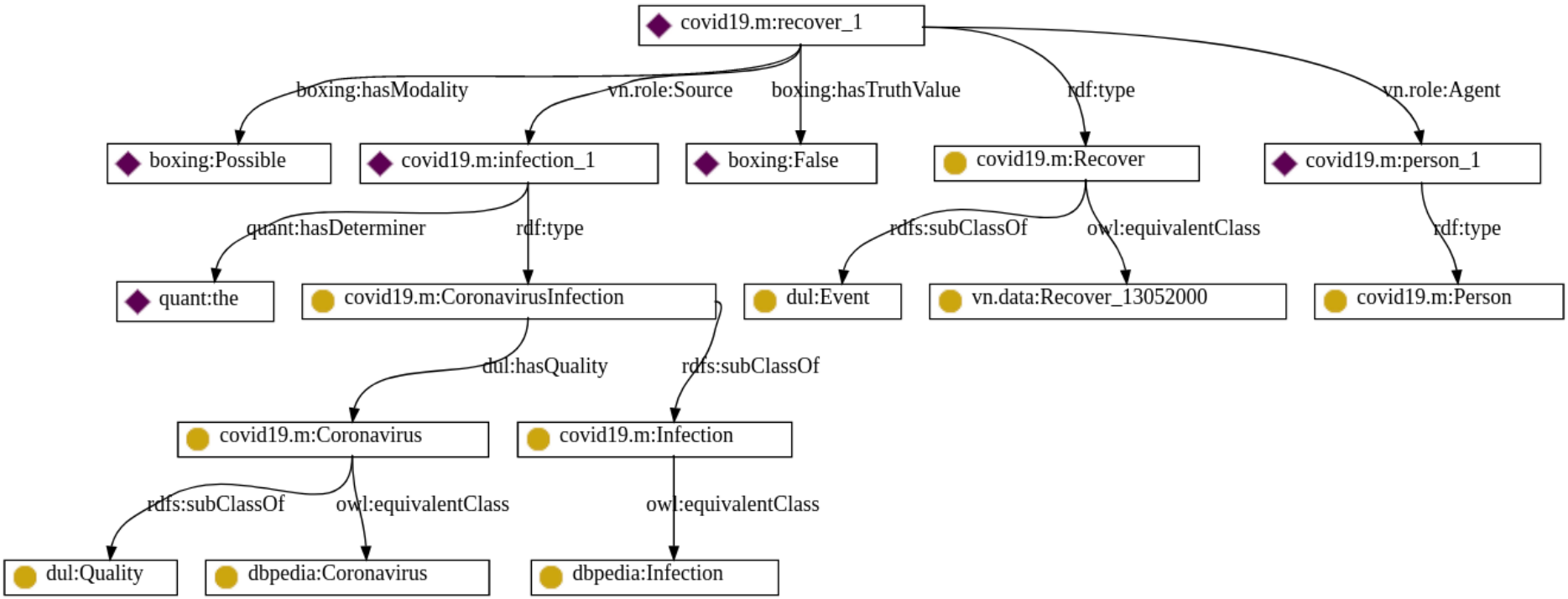
- We built an automated FAct Checker with Explanations (FACE)
- Available at https://github.com/katonaaron/FACE
- Detecting fake news triggers a counterspeech generation

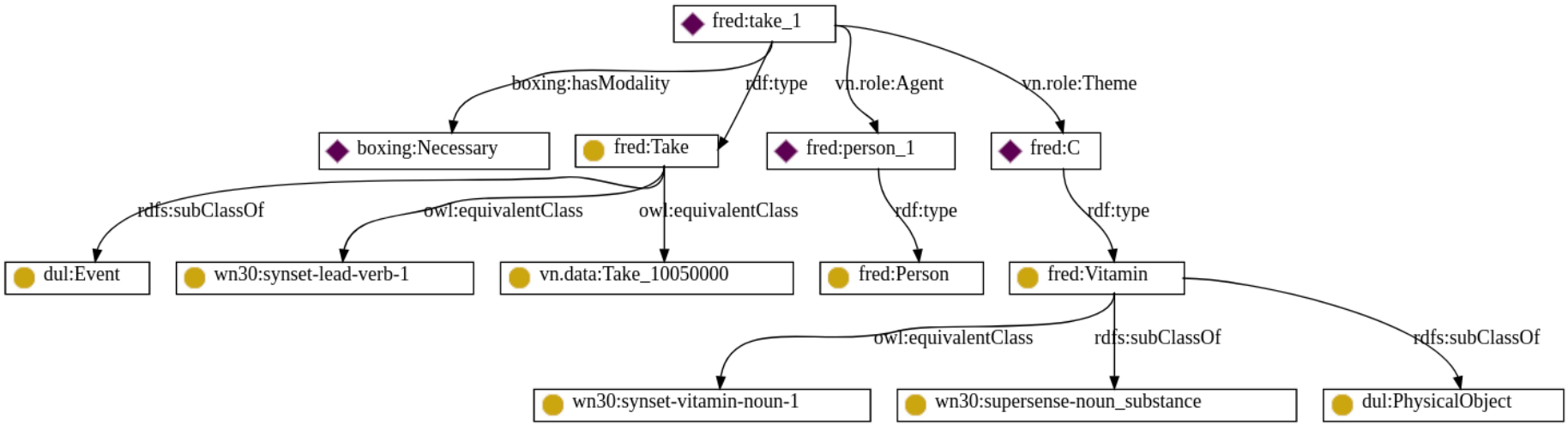


$\exists participates.\{thisSession\} \sqsubseteq \exists thank^{-}.\{I\}$

Formalising negation

You can not recover from the coronavirus infection



Modalities

You should take vitamin C