ISWS INTERNATIONAL SEMANTIC WEB RESEARCH SUMMER SCHOOL

SCHOOL FEE ALL INCLUSIVE **970** EUROS

SCHOOL THEME

KNOWLEDGE GRAPHS AND ARTIFICIAL INTELLIGENCE

July 5th - 11th 2020
Bertinoro, Italy

WHAT IS ISWS?

http://www.semanticwebschool.org

A full immersion, super intensive one-week experience including lectures and keynotes from outstanding speakers, a "learning by doing" teamwork program on open research problems, under the guidance of the best scientists in the field, with whom you will co-author a white paper of high potential impact [1].
ISWS means networking and becoming friends with your peers as well as senior scientists, and have lots of fun with them. At ISWS we shape the future generation of semantic web researchers with creativity, methodology, fun and rigour.

IMPORTANT DATES

REGISTRATION OPEN:
JANUARY 27TH

STUDENTS APPLICATION DEADLINE:
APRIL 3RD

ACCEPTCE NOTIFICATIONS:
APRIL 30TH

REGISTRATION DEADLINE:
MAY 20TH

DIRECTORS

Valentina Presutti
valentina.presutti@unibo.it
Univ. of Bologna & STLab-ISTC CNR (IT)

Harald Sack
harald.sack@fiz-karlsruhe.de
FIZ Karlsruhe & KIT (DE)

TUTORS

Claudia d'Amato - University of Bari (IT)
Irene Celino - Cefriel (IT)
John Domingue - KMi, OPen Univ. & STI International (UK)
Michel Dumontier - Maastricht University (NL)
Aldo Gangemi - Univ. of Bologna and STLab ISTC-CNR (IT)
Heiko Paulheim - University of Mannheim (DE)
Axel Polleres - WU Vienna (AT)

SPEAKERS

Frank van Harmelen - Vrije Univ. Amsterdam (NL)
Elena Simperl - University of Southampton (UK)

COMMUNICATION & ADMINISTRATION

Roberta Partisani - CEUB (IT)
Martina Sangiovanni - STLab ISTC-CNR (IT)

[1] ISWS 2018 report

Visit the website!

Looking for Common Sense in the Semantic Web

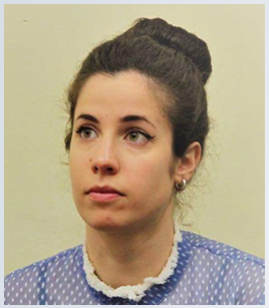
Valentina Presutti

University of Bologna

STLab, ISTC, National Research Council, Italy

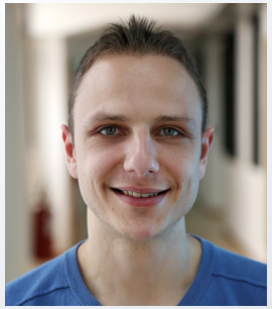
STLAB

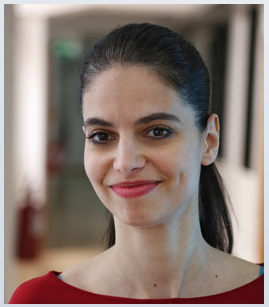
SEMANTIC TECHNOLOGY LABORATORY

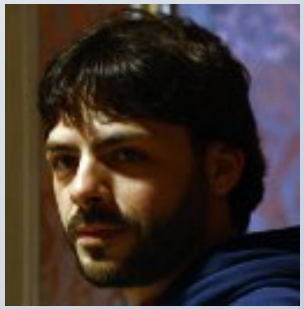







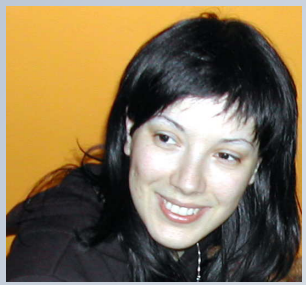





What's in this talk?

- More questions than answers
- More problems than solutions
- A lot of work from others
- Some work from my group
- Some ideas that hopefully will be of inspiration for some of you



undergradStudentAt

phdStudentAt

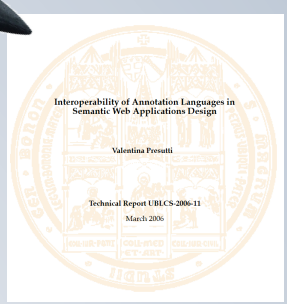
ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

2002

masterIn

phdIn

2006



Computer Science

assistantProfessorAt

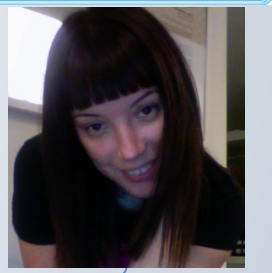
2019 - now

associateResearcherAt

postdocAt

2006
2010

researchLabAt



ISWS

STLAB
SEMANTIC TECHNOLOGY LABORATORY

CNR
ISTC
ISTITUTO DI SCIENZE E TECNOLOGIE DELLA COGNIZIONE

FRED
Machine Reading for the Semantic Web

coordinatorOf

co-founderOf

2015
now

2008

permanentResearcherAt

coDirectorOf

2018 - now

MARIO

2010
2019

IKS


ACTIVE girl

KNOWLEDGE MEDIA INSTITUTE

Thesis #1

The SW today has already reached a level of scale good enough to make it a very useful source of knowledge to support intelligent applications

In other words: the Semantic Web is no longer an aspiration but a reality

The availability of such large scale amounts of formalised knowledge is unprecedented in the history of AI

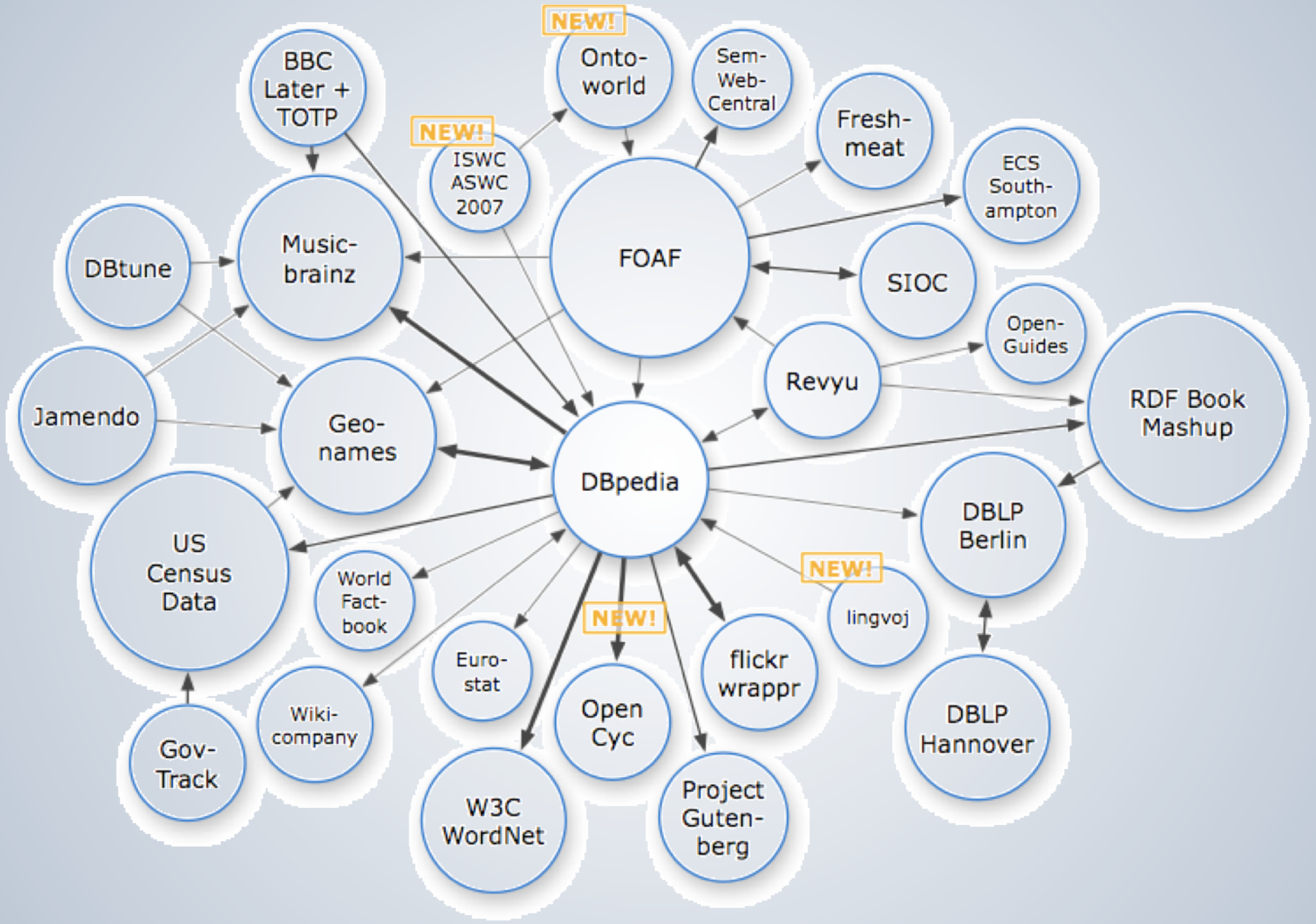
From Enrico Motta's talk at SSSW 2007

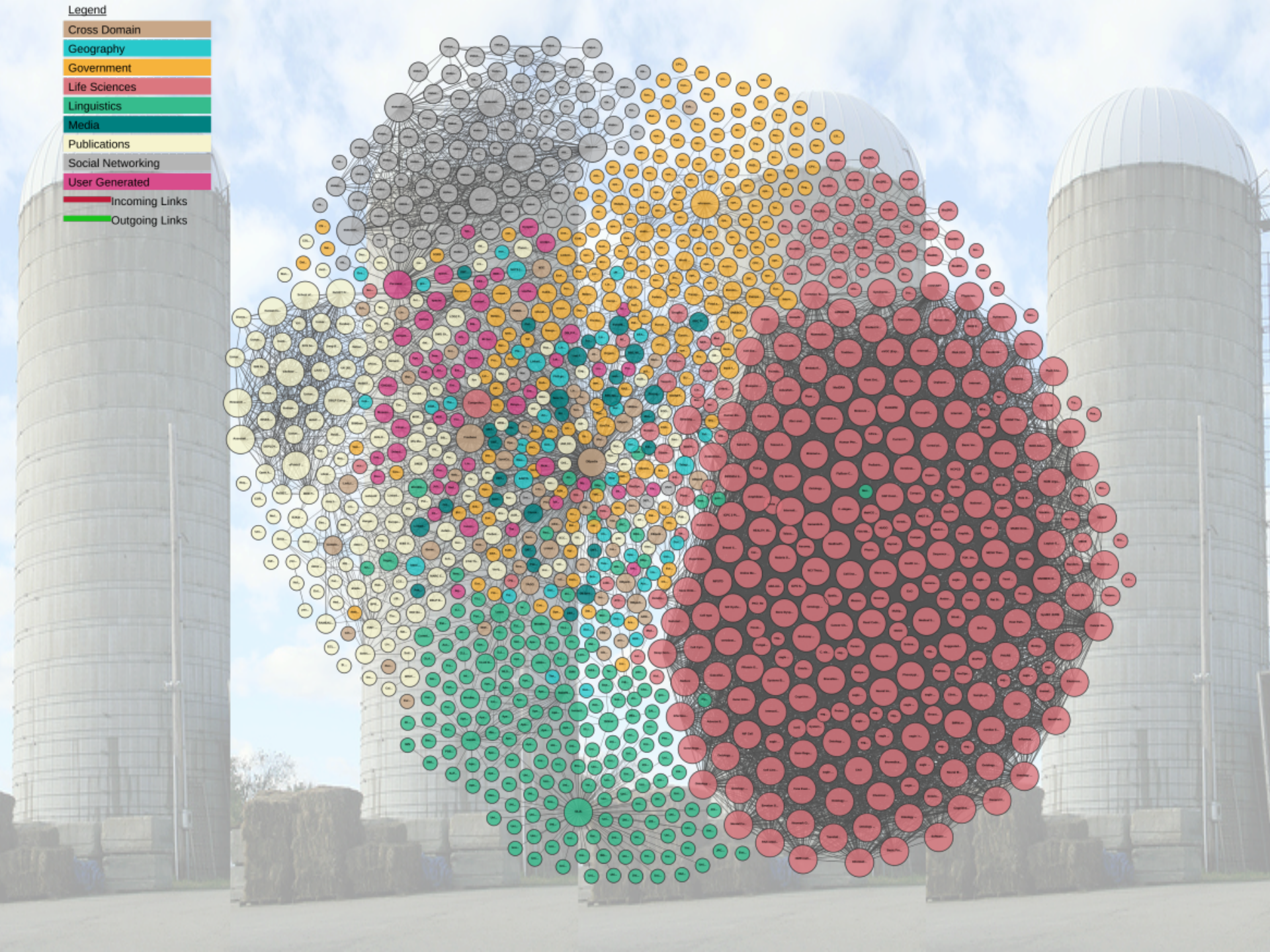
KNOWLEDGE MEDIA INSTITUTE

Thesis #2

The SW may well provide a solution to one of the classic AI challenges: how to acquire and manage large volumes of knowledge to develop truly intelligent problem solvers and address the brittleness of traditional KBS

From Enrico Motta's talk at SSSW 2007

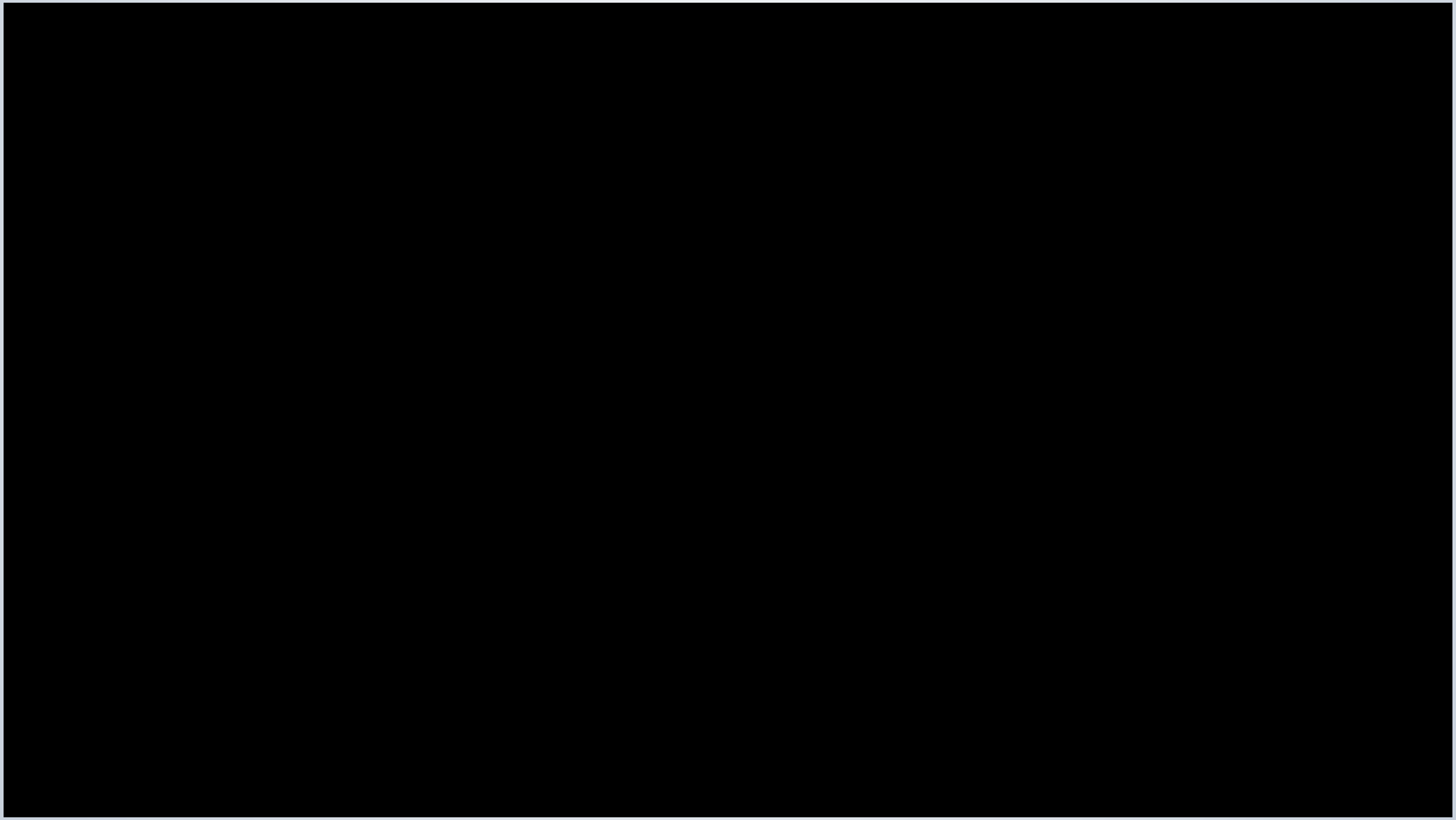

BBC Later + TOTP
NEW!
Onto-world
Sem-Web-Central
NEW!
ISWC ASWC 2007
Fresh-meat
ECS South-ampton
DBtune
Music-brainz
FOAF
SIOC
Open-Guides
Jamendo
Geo-names
Revyu
RDF Book Mashup
DBpedia
DBLP Berlin
US Census Data
World Fact-book
NEW!
lingvoj
Euro-stat
NEW!
flickr wrappr
Wiki-company
Open Cyc
DBLP Hannover
Gov-Track
W3C WordNet
Project Guten-berg


Legend
Cross Domain
Geography
Government
Life Sciences
Linguistics
Media
Publications
Social Networking
User Generated
Incoming Links
Outgoing Links

From Frank van Harmelen's talk at ISWS 2019

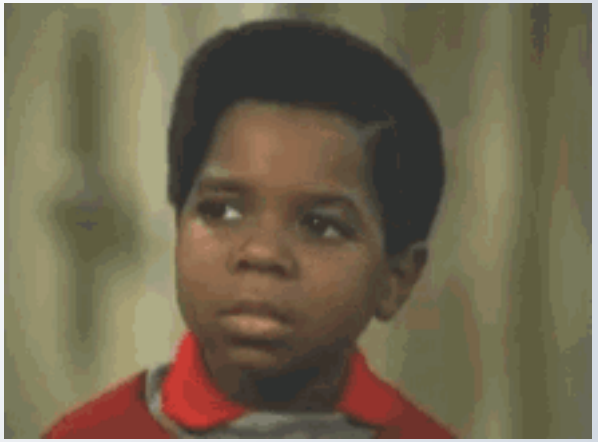
Siri

Expectations vs. Reality



They have no idea what they're talking about

- They do not reason
- They are not aware of the surrounding context
- They do not have any commonsense
- In the best case the right answer is built-in
- otherwise they borrow from Wikipedia or other (encyclopedic?) sources
- or they issue a query for you on Google
- in the worst case they are: *sorry, I think this is beyond my capability at the moment*




EASY
HARD

The lack of common sense knowledge

A main problem is the unavailability of common sense knowledge (and reasoning)

Existing knowledge graphs mainly encode domain-specific or encyclopedic knowledge (a special type of CSK)

This is a long-standing challenge of Artificial Intelligence

"Common-sense facts and methods are only very partially understood today, and extending this understanding is the key problem facing artificial intelligence."

John McCarthy, 1983

Some expert systems need common sense, Annals of the New York Academy of Sciences

"[AI needs a] formalization of a sizable portion of commonsense knowledge about the everyday physical world"

Patrick Hayes, 1989
The second naive physics manifesto

"We need common-sense knowledge – and programs that can use it. Common sense computing needs several ways of representing knowledge. It is harder to make a computer housekeeper than a computer chess-player, because the housekeeper must deal with a wider range of situations."

Marvin Minsky, 1998

The mind, artificial intelligence and emotions

Interview with Marvin Minsky

"To make real progress in A.I., we have to overcome the big challenges in the area of common sense"

Paul Allen, 2018

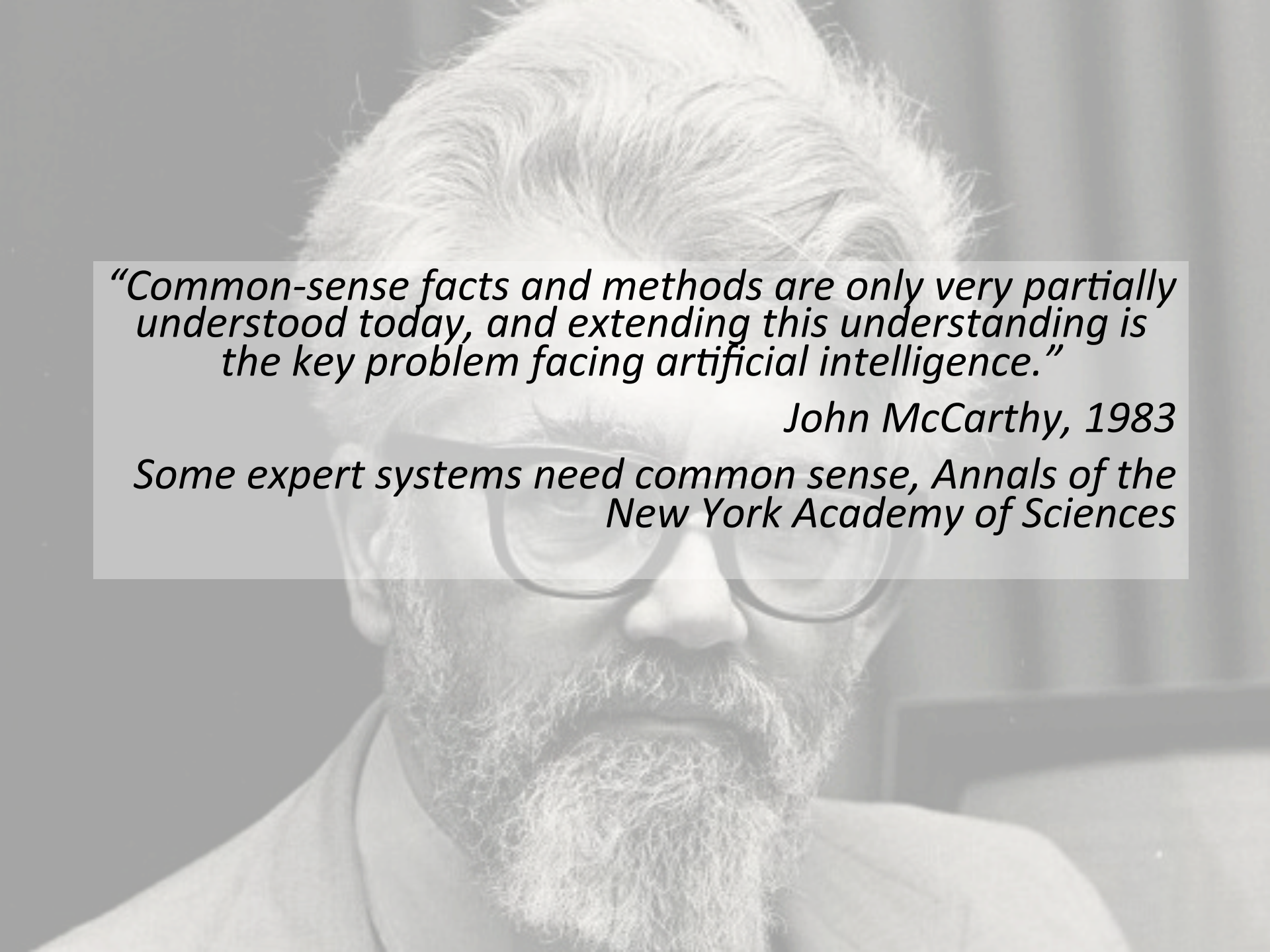
The New York Times

A *vague* definition of common sense

That type of knowledge that we all give for granted when we communicate with other humans, independently from the communication means (e.g. text, oral). We assume that (a group of) people share it. Commonsense also includes the type of reasoning that we perform with that knowledge.

Some examples are:

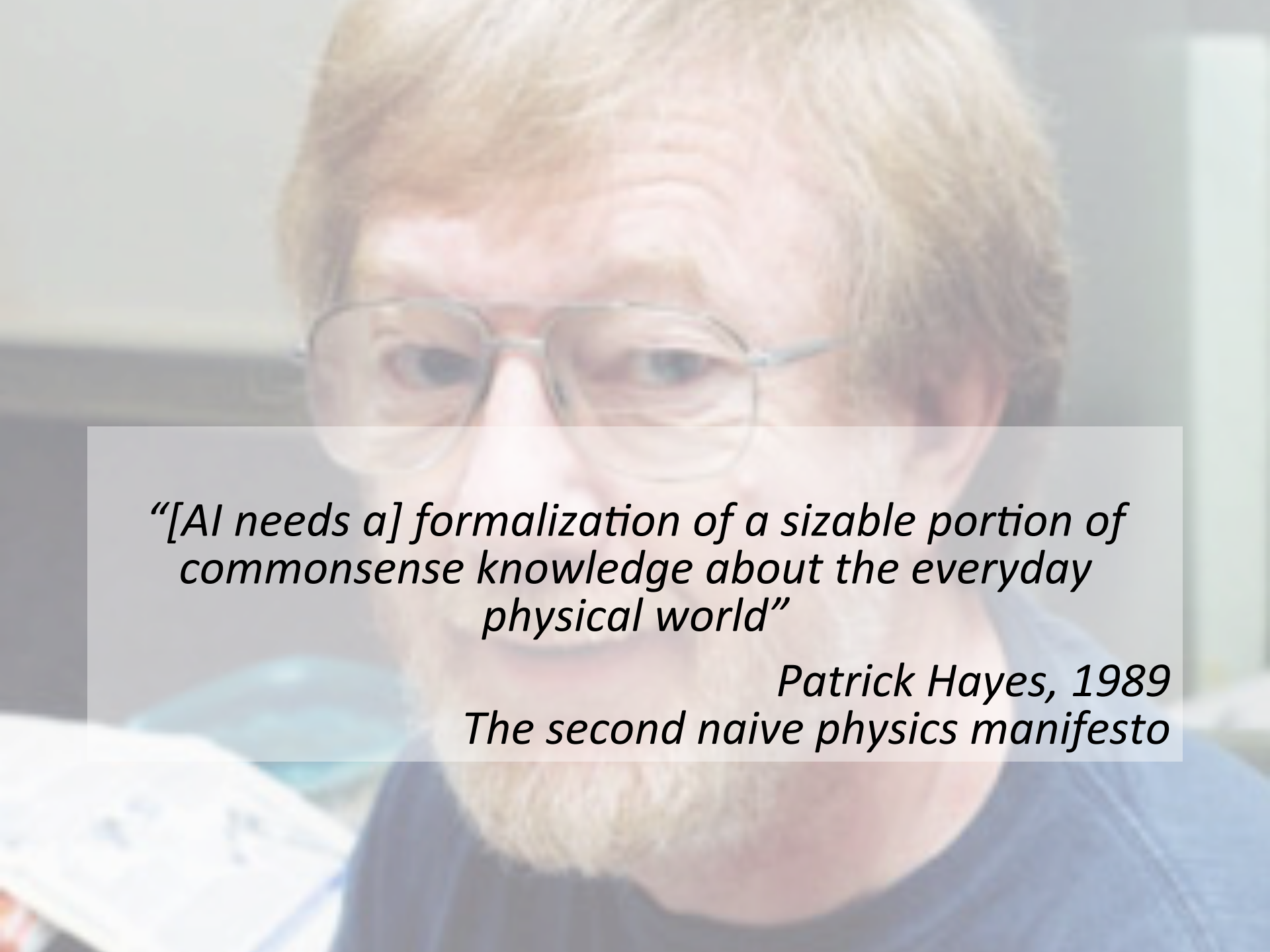
- Males cannot give birth
- A mother is older than her children
- A physical object cannot be in two places at the same time
- Liquids flow and have the shape of their container
- Cutlery is usually in a kitchen, in a drawer



"Common-sense facts and methods are only very partially understood today, and extending this understanding is the key problem facing artificial intelligence."

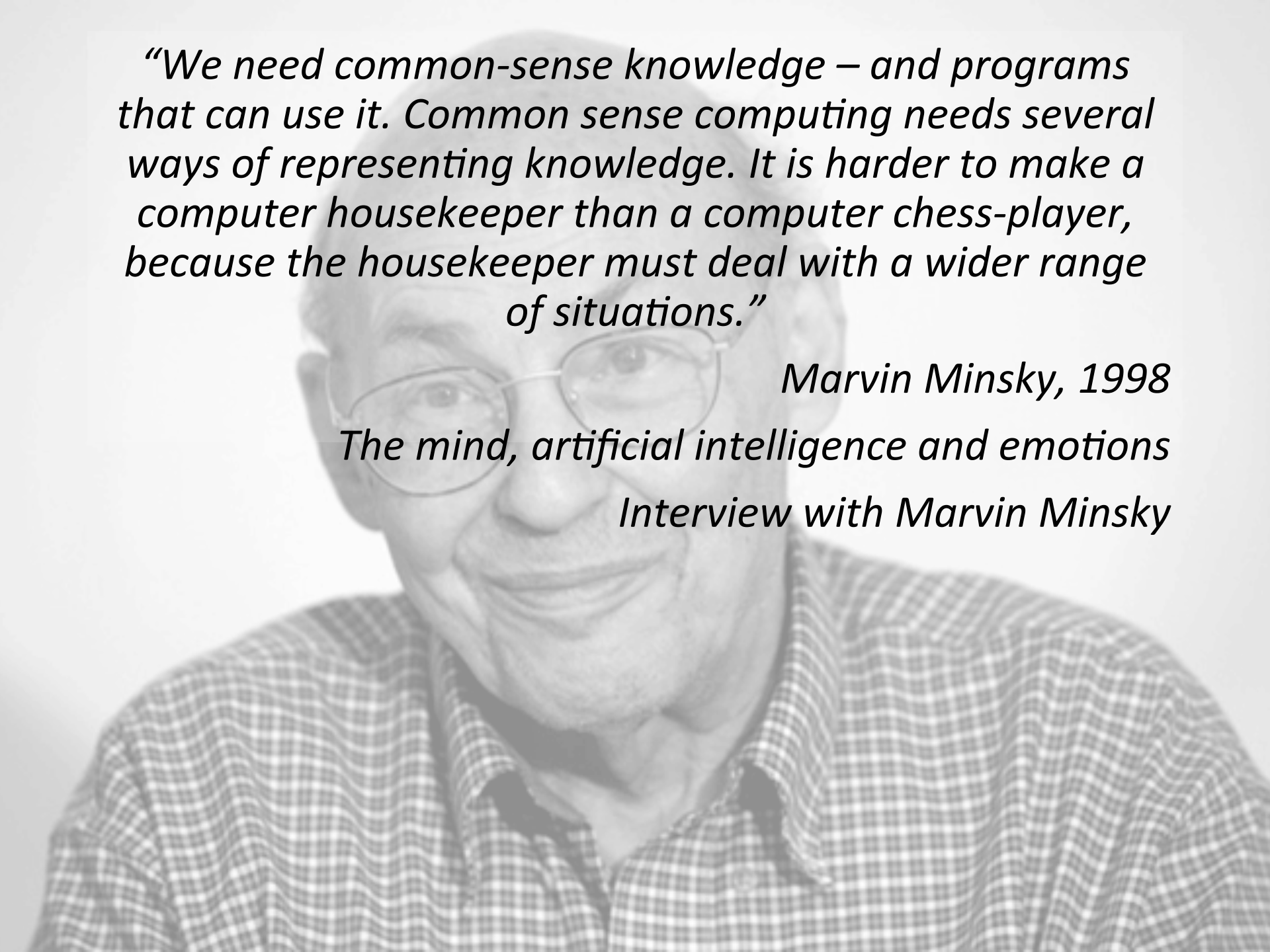
John McCarthy, 1983

Some expert systems need common sense, Annals of the New York Academy of Sciences



"[AI needs a] formalization of a sizable portion of commonsense knowledge about the everyday physical world"

Patrick Hayes, 1989
The second naive physics manifesto

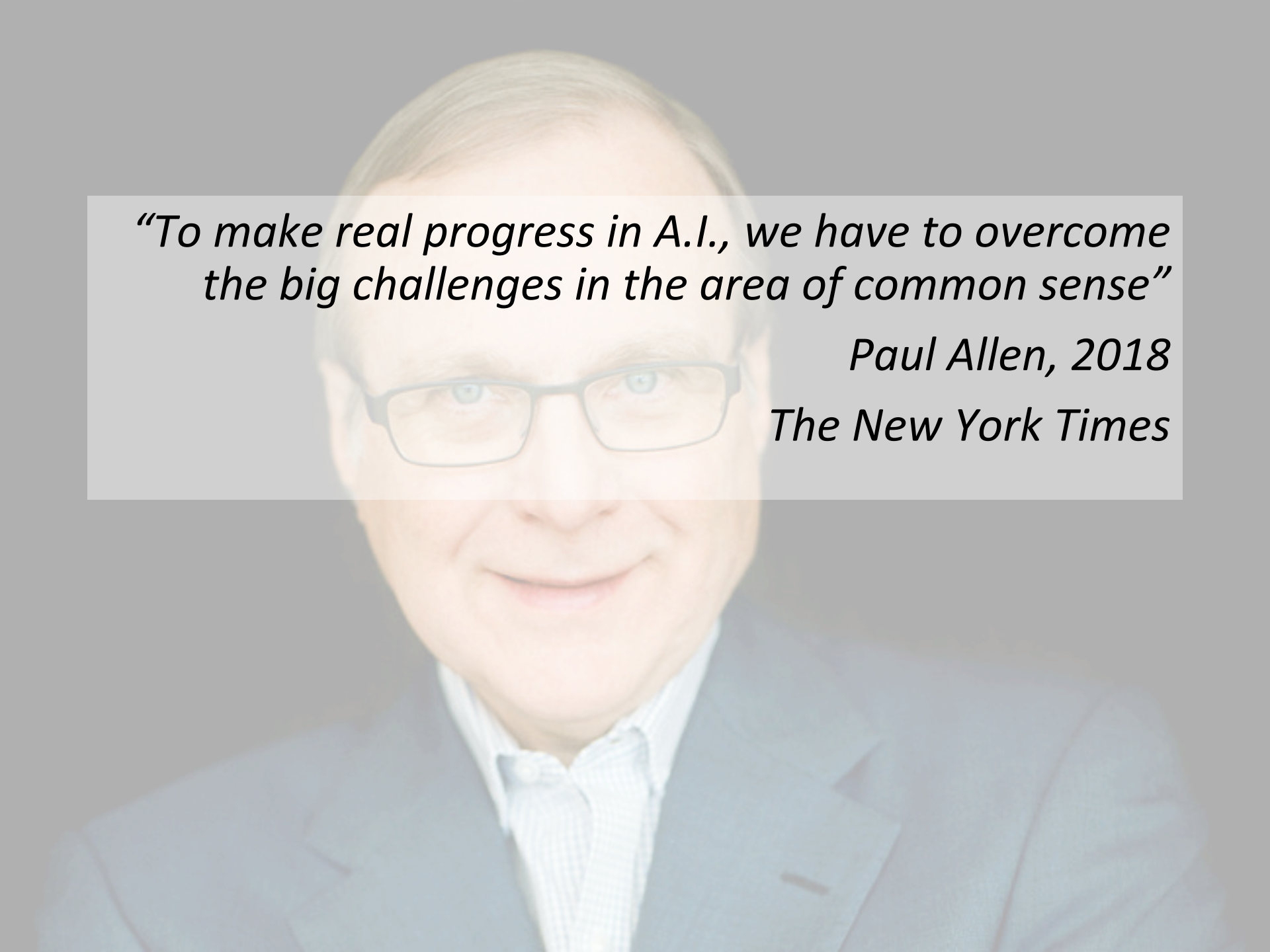


"We need common-sense knowledge – and programs that can use it. Common sense computing needs several ways of representing knowledge. It is harder to make a computer housekeeper than a computer chess-player, because the housekeeper must deal with a wider range of situations."

Marvin Minsky, 1998

The mind, artificial intelligence and emotions

Interview with Marvin Minsky



“To make real progress in A.I., we have to overcome the big challenges in the area of common sense”

Paul Allen, 2018

The New York Times

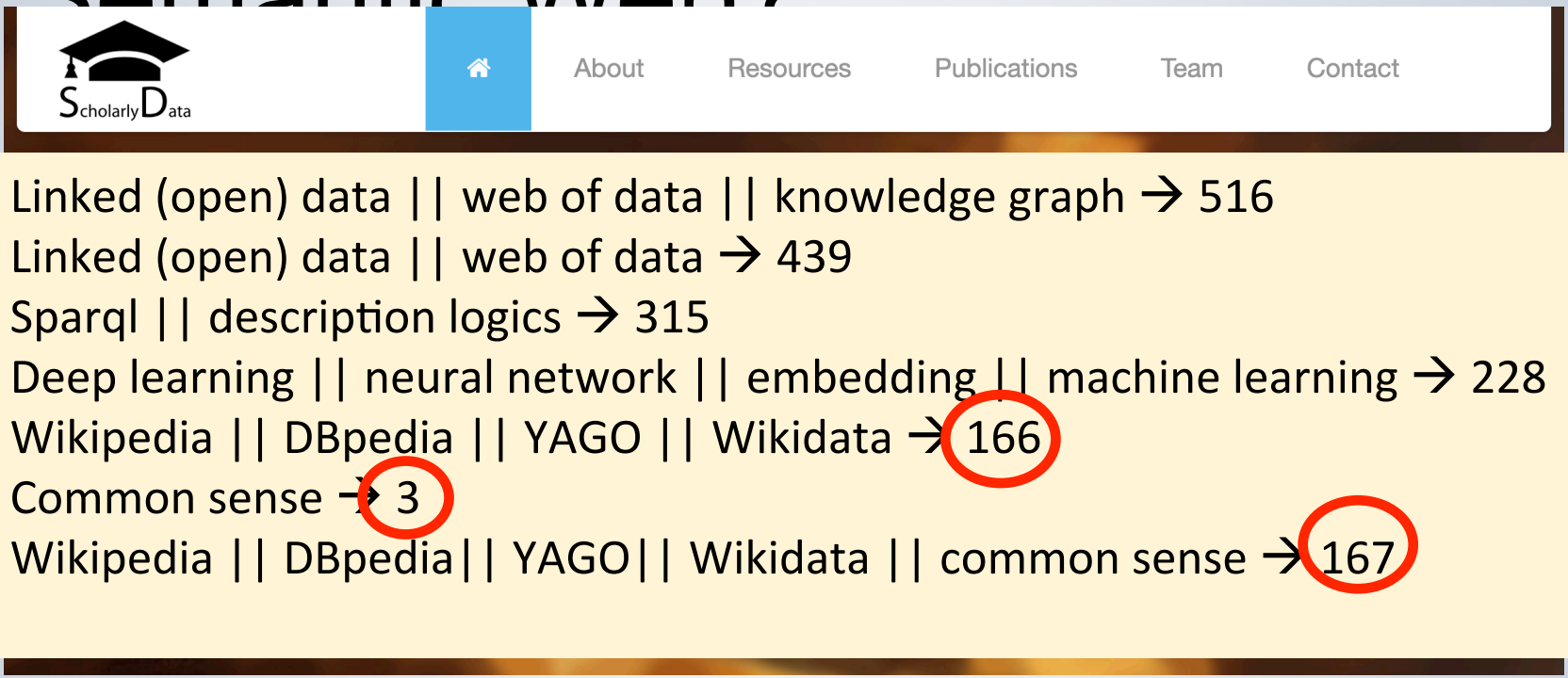
(A) Role of Semantic Web in Artificial intelligence

To create a robust, distributed, public* knowledge graph of *common sense knowledge*

* and F.A.I.R.

Piero Andrea Bonatti, Stefan Decker, Axel Polleres, Valentina Presutti: **Knowledge Graphs: New Directions for Knowledge Representation on the Semantic Web (Dagstuhl Seminar 18371).**

How much research on common sense in the Semantic Web?



http://www.scholarlydata.org/

2197 papers

ESWC 2006-2017
ISWC 2001-2018
EKAW 2010-2016
WWW 2007-2012

How can we build a knowledge graph of common sense?

Let us *look deeper* into what we already have

Ontological analysis and Empirical Semantics

Is there explicit commonsense knowledge in the (Semantic) Web?

Ontological analysis

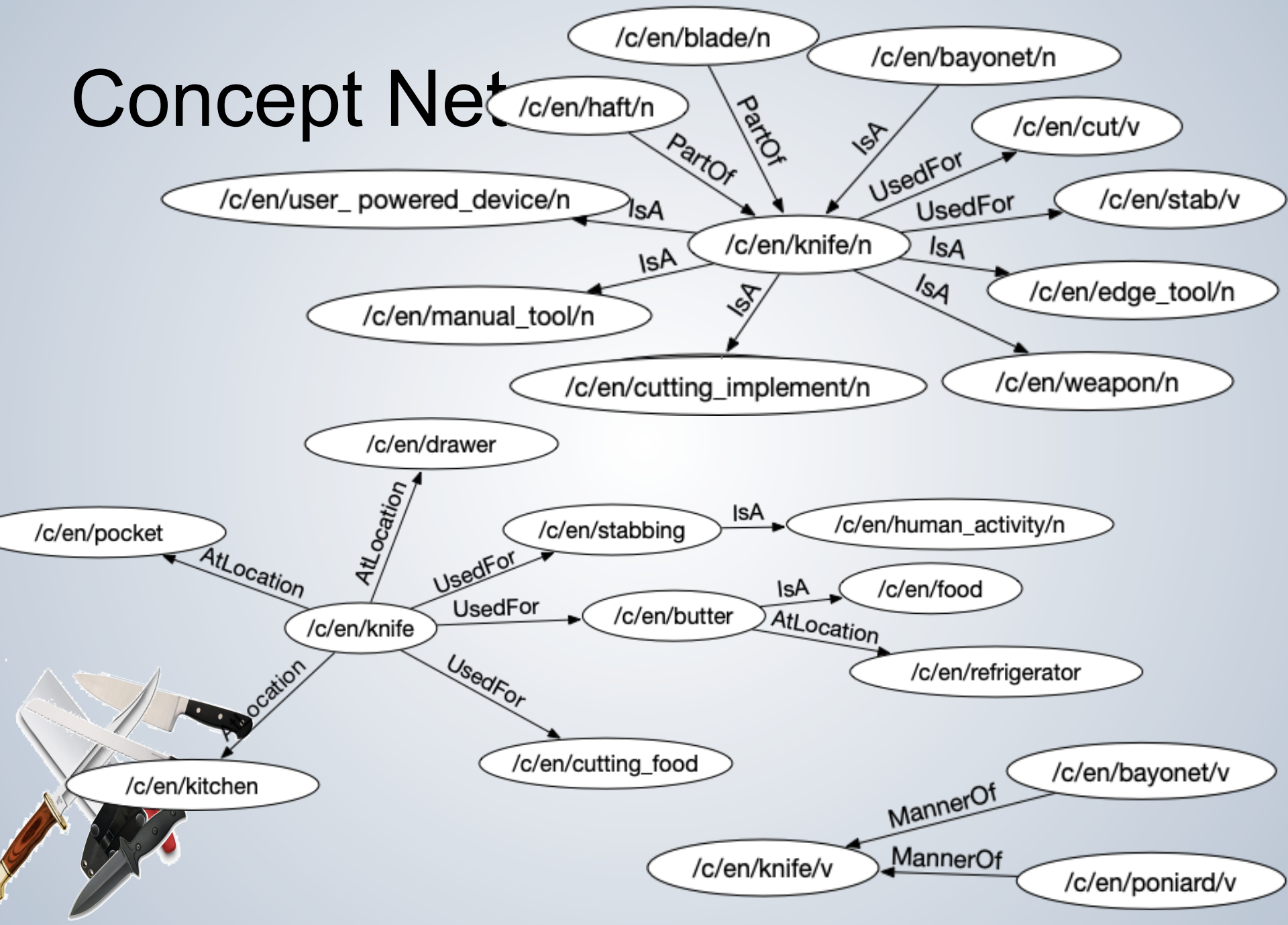
Let’s look at available knowledge about an everyday object such as a *knife*

ConceptNet http://conceptnet.io/

- A labeled graph (a semantic network) that targets text processing
  - It defines a closed set of relation-labels
- It has accumulated ~1M English facts
- Based on contribution from web users (Open Mind Common Sense, 1999) also through *games with a purpose*
- Reusing Wiktionary and WordNet and aligned partially to DBpedia
- It provides JSON-LD APIs

Robert Speer, Catherine Havasi: Representing General Relational Knowledge in ConceptNet 5. LREC 2012: 3679-3686

Concept Net


/c/en/blade/n
/c/en/bayonet/n
/c/en/haft/n
PartOf
PartOf
IsA
/c/en/cut/v
UsedFor
/c/en/user_ powered_device/n
IsA
/c/en/knife/n
UsedFor
/c/en/stab/v
IsA
IsA
/c/en/edge_tool/n
/c/en/manual_tool/n
IsA
IsA
/c/en/cutting_implement/n
/c/en/weapon/n
/c/en/drawer
/c/en/pocket
AtLocation
AtLocation
/c/en/stabbing
IsA
/c/en/human_activity/n
UsedFor
/c/en/knife
UsedFor
/c/en/butter
IsA
/c/en/food
AtLocation
/c/en/refrigerator
UsedFor
/c/en/cutting_food
/c/en/kitchen
/c/en/bayonet/v
MannerOf
/c/en/knife/v
MannerOf
/c/en/poniard/v

NELL http://rtw.ml.cmu.edu/

- A machine learning-based system that *reads* the web: it extracts facts from textual web documents
- Since 2010 it's been running continuously and has learned ~50M candidate beliefs
  - ~2.8M with high confidence
- Candidate beliefs are encoded as a KB of facts and an ontology of categories and relations
- NELL KB and Ontology are openly available also as LOD

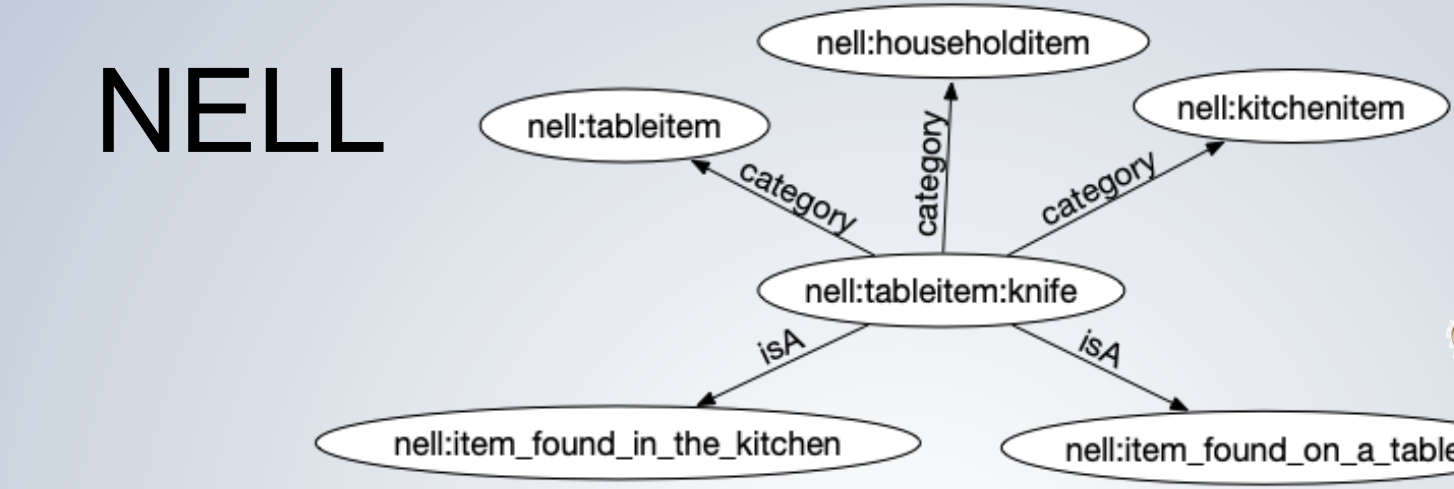
Tom M. Mitchell et al: **Never-Ending Learning.** AAAI 2015: 2302-2310

NELL

- 200 results
- Each associated with a category
- 35 are kitchenitem, tableitem, or householditem
- They are not related to each other (e.g. by taxonomical relations)

knife (tableitem)
bread_knife (kitchenitem)
knife_rack (kitchenitem)
dinner_knife (kitchenitem)
knife_set (kitchenitem)
wet_knife (kitchenitem)
knife_key (tableitem)
a_knife (kitchenitem)
dry_knife (tableitem)
knife_and (tableitem)
knife_with (tableitem)
diy_knife (householditem)
butter_knife (kitchenitem)
table_knife (tableitem)
wooden_knife (tableitem)
vegetable_knife (tableitem)
cheese_knife (tableitem)
disposable_knife (tableitem)
paring_knife (kitchenitem)
cutlery_knife (tableitem)
case_knife (kitchenitem)
steak_knife (householditem)
electric_knife (householditem)
carving_knife (kitchenitem)
exacto_knife (kitchenitem)

NELL


nell:householditem
nell:tableitem
nell:kitchenitem
category
category
category
nell:tableitem:knife
isA
isA
nell:item_found_in_the_kitchen
nell:item_found_on_a_table

nell:householditem
nell:tableitem
nell:kitchenitem
category
category
category
nell:kitchenitem:bread_knife
isA
nell:visualizable_object
isA
isA
nell:item_found_in_the_kitchen
nell:item_found_on_a_table

Atomic https://homes.cs.washington.edu/~msap/atomic/

- Textual descriptions of inferential knowledge (*if-then clauses*)
  - Based on 3 *if-then* relation types associated with 9 causal and inferential dimensions
- It has accumulated ~877k textual descriptions
- Crowdsourcing of "blank placeholders" put in 24k event phrases
  - Phrases extracted from Google Ngrams, Wiktionary, books, etc.

Maarten Sap, Ronan Le Bras, Emily Allaway, Chandra Bhagavatula, Nicholas Lourie, Hannah Rashkin, Brendan Roof, Noah A. Smith, Yejin Choi: **ATOMIC: An Atlas of Machine Commonsense for If-Then Reasoning.** AAAI 2019: 3027-3035

Atomic

PersonX grabs a knife

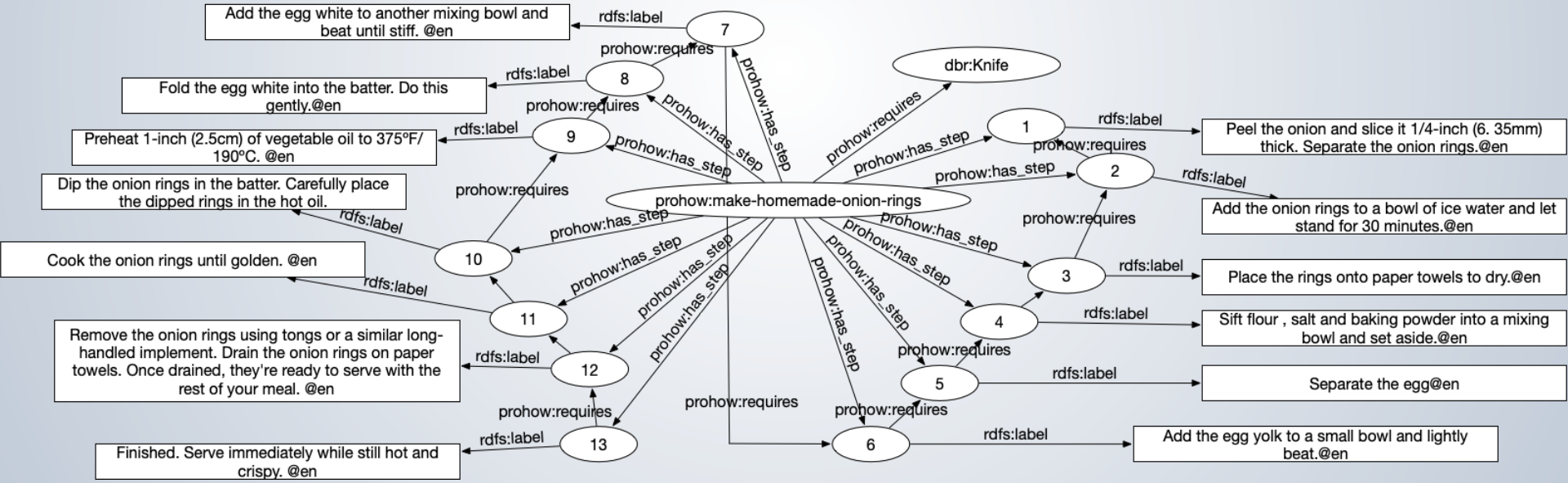
- Causes for PersonX
  - Because PersonX **wanted**
    - to cut the cake
    - to cut up some food
    - to cut something
  - Before, PersonX **needed**
    - to go to the kitchen
    - to open the drawer
    - go in kitchen
    - search for knife
- Attributes of PersonX
  - PersonX is **seen as**
    - Defensive
    - Alert
    - scary
    - defensive
- Effects on PersonX
  - As a result, PersonX **feels**
    - eager
    - excited to eat
    - useful
    - threatening
  - As a result, PersonX **wants**
    - to cut some food
    - to eat some food
    - cut vegitables
    - keep the knife
  - PersonX **then**
    - none
    - stabs someone
    - cuts themselves
    - starts cutting vegetables
    - holds it firm in his hand
    - checks if it is sharp enough.
- Effects on others
  - As a result, others **feel**
    - none
  - As a result, others **want**
    - none
  - Others **then**
    - none

The Human Know-How Dataset

https://datashare.is.ed.ac.uk/handle/10283/1985

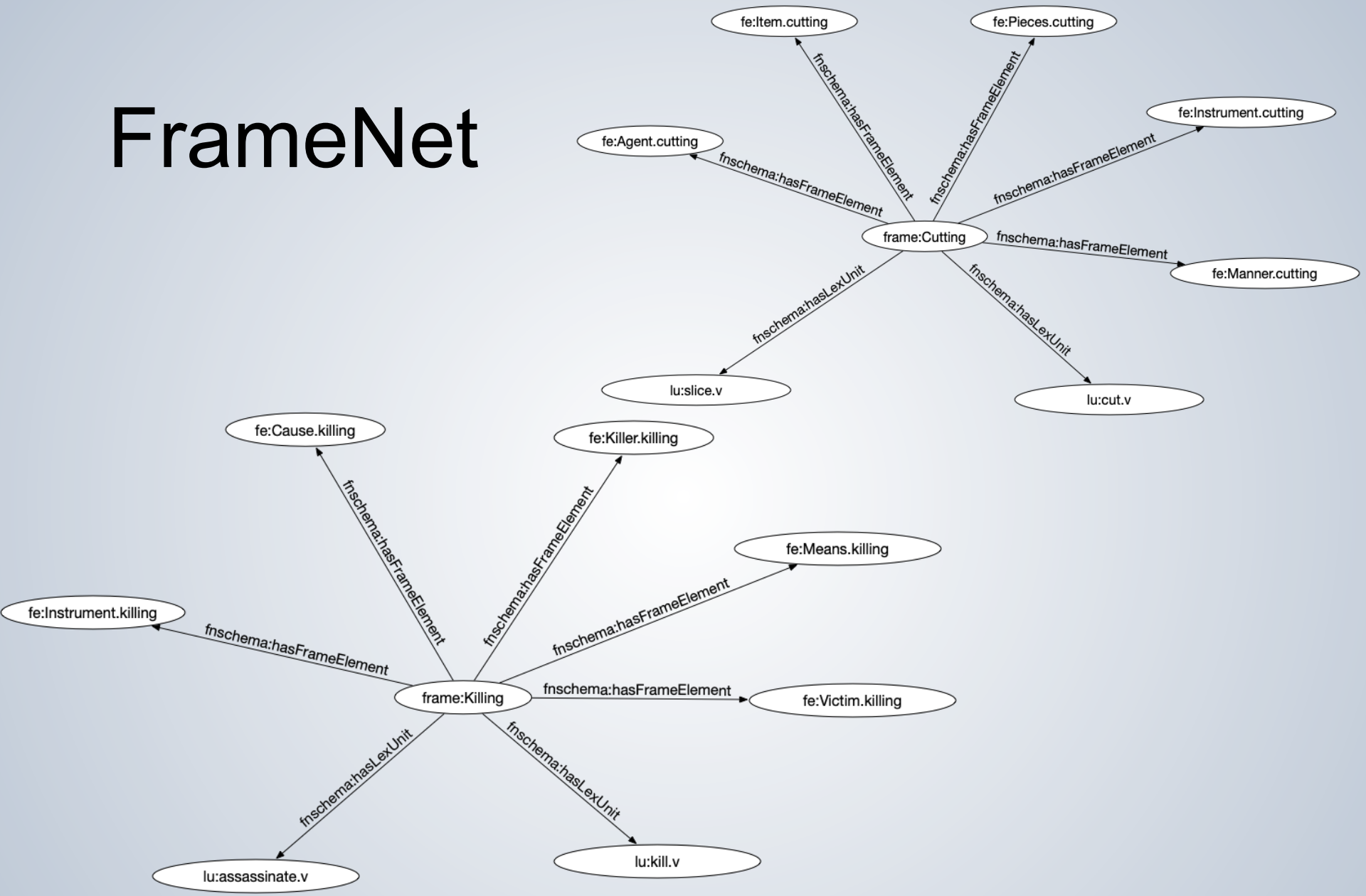
- A dataset and an ontology (PROHOW) derived from the WikiHow project
- It focuses on representing human activities, in particular “how to” processes
- It encodes ~220k such activities described through 2.6M entities
- ~250k entities are linked to DBpedia

The Human Know-How Dataset



Paolo Pareti, Ewan H. Klein (2016). The Human Know-How Dataset, 2014 [dataset]. https://doi.org/10.7488/ds/1394

FrameNet



Collin F Baker, Fillmore, Charles J, and Lowe, John B. 1998. The Berkeley FrameNet project. In COLING-ACL '98: Proceedings of the Conference, Montreal, Canada

Good News!

There are quite a number of valuable commonsense knowledge resources in the Semantic Web!

They are incomplete, mostly informal, encoded with conceptual heterogeneity, lacking *contextualisation*, scarcely linked

How can we reuse them?

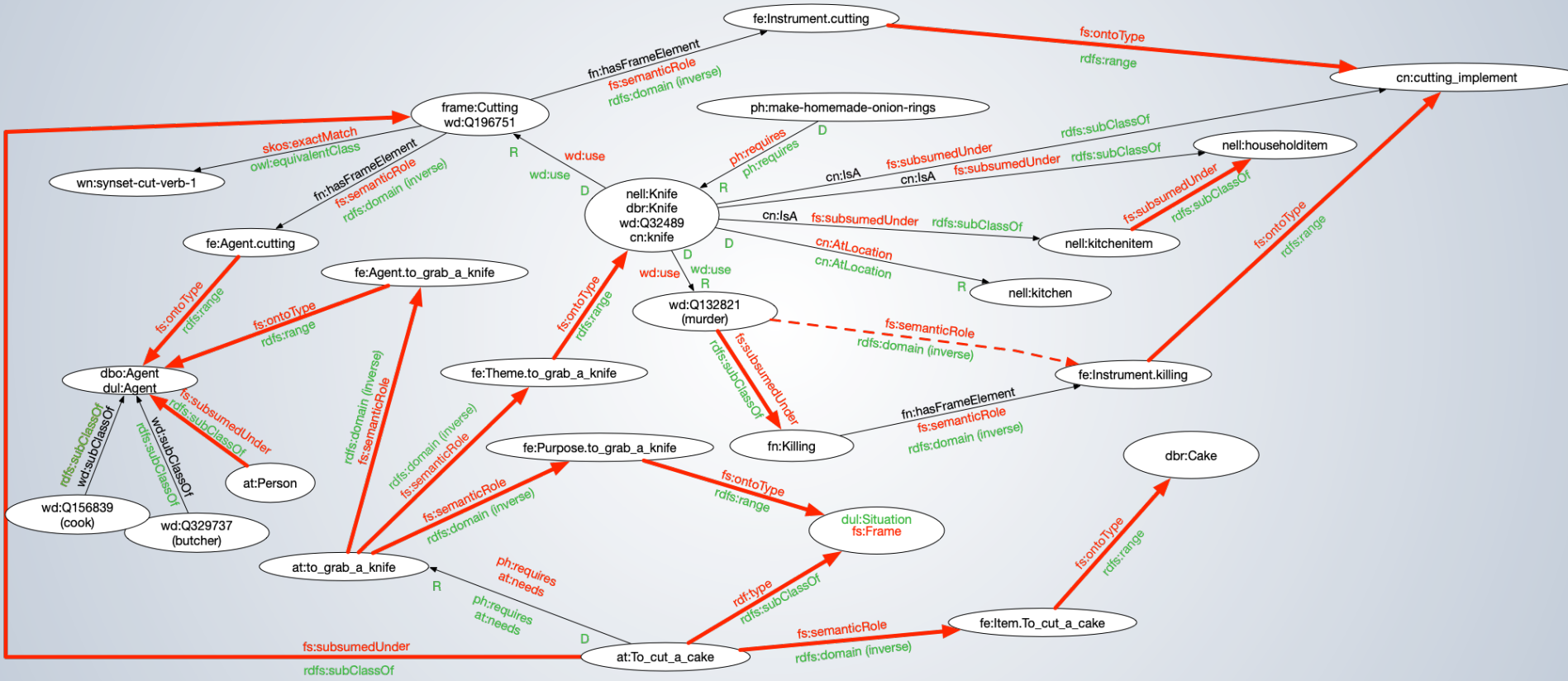
https://w3id.org/framester/

Framester

- Framester is a LOD resource that connects *linguistic* data with *factual* and *ontological* data
- It encodes 50M links between 21 resources
  - Resources include: DBpedia, WordNet, DOLCE, FrameNet, SentiWordNet, ConceptNet, etc.
  - Linking relations include: skos:closeMatch, skos: exactMatch, owl:equivalentClass, owl:sameAs, etc.
- Linking is based on a formal frame semantics: an attempt to provide a *unified semantics* to such diverse resources
- The core of Framester is FrameNet (its LOD version)

Aldo Gangemi, Mehwish Alam, Luigi Asprino, Valentina Presutti, Diego Reforgiato Recupero: Framester: A Wide Coverage Linguistic Linked Data Hub. EKAW 2016: 239-254

Knife


fe:Instrument.cutting
fs:ontoType
rdfs:range
cn:cutting_implement
fn:hasFrameElement
fs:semanticRole
rdfs:domain (inverse)
frame:Cutting
wd:Q196751
ph:make-homemade-onion-rings
skos:exactMatch
owl:equivalentClass
wn:synset-cut-verb-1
fn:hasFrameElement
fs:semanticRole
rdfs:domain (inverse)
R
wd:use
wd:use
D
ph:requires
ph:requires
D
R
nell:Knife
dbr:Knife
wd:Q32489
cn:knife
cn:IsA
fs:subsumedUnder
rdfs:subClassOf
cn:IsA
fs:subsumedUnder
rdfs:subClassOf
nell:householditem
cn:IsA
fs:subsumedUnder
rdfs:subClassOf
nell:kitchenitem
fs:subsumedUnder
rdfs:subClassOf
fe:Agent.cutting
D
D
cn:AtLocation
cn:AtLocation
R
nell:kitchen
fe:Agent.to_grab_a_knife
wd:use
wd:use
R
fs:ontoType
rdfs:range
fs:ontoType
rdfs:range
fs:ontoType
rdfs:range
wd:Q132821
(murder)
fs:semanticRole
rdfs:domain (inverse)
fs:ontoType
rdfs:range
dbo:Agent
dul:Agent
fe:Theta.to_grab_a_knife
fe:Instrument.killing
fs:subsumedUnder
rdfs:subClassOf
rdfs:subClassOf
wd:subClassOf
rdfs:subClassOf
wd:subClassOf
fs:subsumedUnder
rdfs:subClassOf
rdfs:domain (inverse)
fs:semanticRole
rdfs:domain (inverse)
fs:semanticRole
fn:Killing
fn:hasFrameElement
fs:semanticRole
rdfs:domain (inverse)
at:Person
fe:Purpose.to_grab_a_knife
dbr:Cake
wd:Q156839
(cook)
wd:Q329737
(butcher)
fs:ontoType
rdfs:range
dul:Situation
fs:Frame
at:to_grab_a_knife
ph:requires
at:needs
R
ph:requires
at:needs
D
rdf:type
rdfs:subClassOf
fs:ontoType
rdfs:range
fe:Item.To_cut_a_cake
fs:subsumedUnder
rdfs:subClassOf
at:To_cut_a_cake
fs:semanticRole
rdfs:domain (inverse)

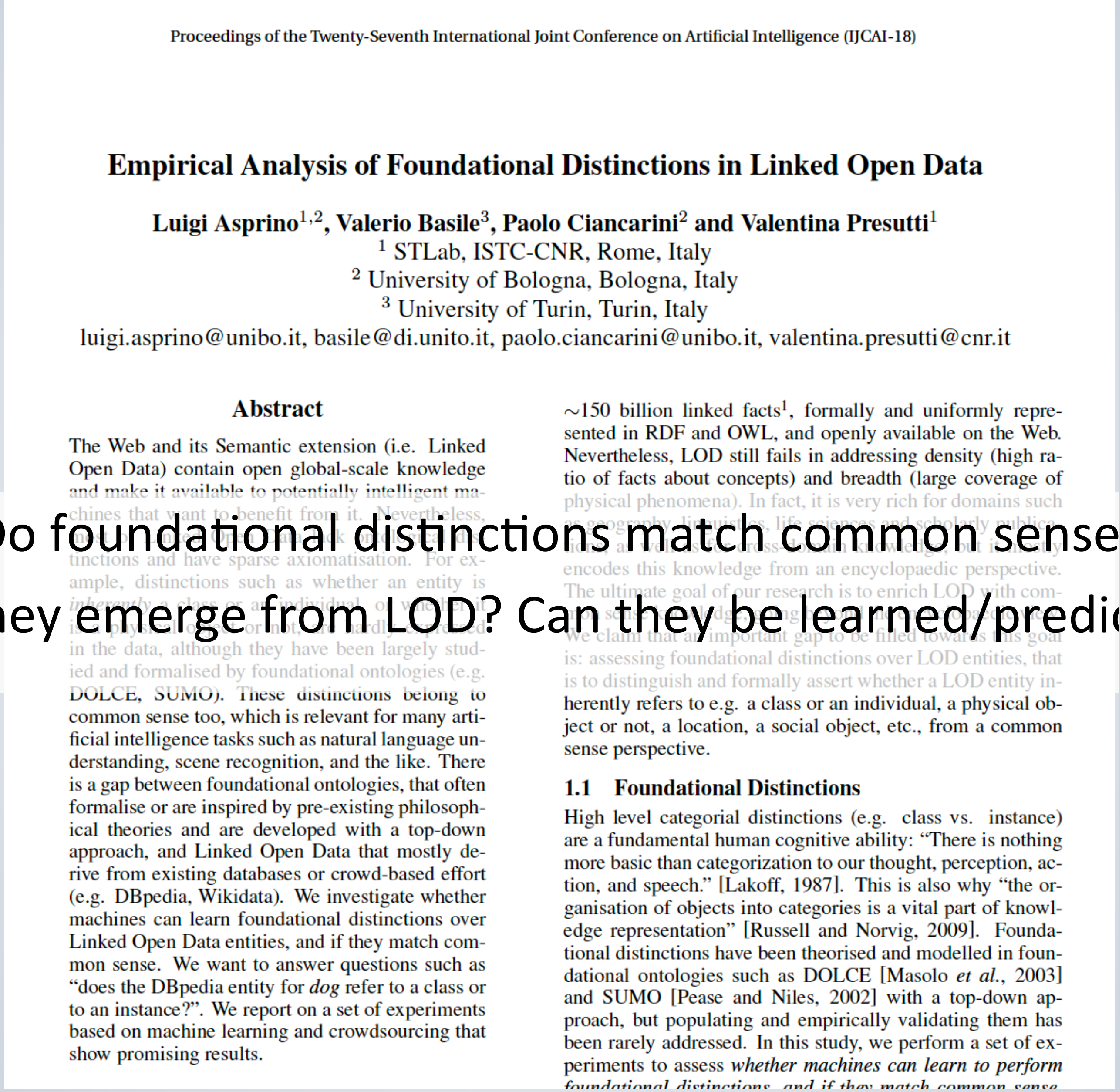
Many interesting research questions

- Criteria for systematic encoding and evaluation of commonsense
- Qualified self-describing commonsense statements
- How to identify explicit commonsense knowledge?
- How to represent common sense knowledge?
- How to evaluate commonsense knowledge quality?

Is there implicit commonsense knowledge in the Semantic Web?

Empirical Semantics

https://w3id.org/fox



Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18)

Empirical Analysis of Foundational Distinctions in Linked Open Data

Luigi Asprino[1,2], Valerio Basile[3], Paolo Ciancarini[2] and Valentina Presutti[1]
[1] STLab, ISTC-CNR, Rome, Italy
[2] University of Bologna, Bologna, Italy
[3] University of Turin, Turin, Italy
luigi.asprino@unibo.it, basile@di.unito.it, paolo.ciancarini@unibo.it, valentina.presutti@cnr.it

Abstract

The Web and its Semantic extension (i.e. Linked Open Data) contain open global-scale knowledge and make it available to potentially intelligent machines that want to benefit from it. Nevertheless, [illegible] tinctions and have sparse axiomatisation. For example, distinctions such as whether an entity is *inherently* a class or an individual, [illegible] in the data, although they have been largely studied and formalised by foundational ontologies (e.g. DOLCE, SUMO). These distinctions belong to common sense too, which is relevant for many artificial intelligence tasks such as natural language understanding, scene recognition, and the like. There is a gap between foundational ontologies, that often formalise or are inspired by pre-existing philosophical theories and are developed with a top-down approach, and Linked Open Data that mostly derive from existing databases or crowd-based effort (e.g. DBpedia, Wikidata). We investigate whether machines can learn foundational distinctions over Linked Open Data entities, and if they match common sense. We want to answer questions such as "does the DBpedia entity for *dog* refer to a class or to an instance?". We report on a set of experiments based on machine learning and crowdsourcing that show promising results.

~150 billion linked facts[1], formally and uniformly represented in RDF and OWL, and openly available on the Web. Nevertheless, LOD still fails in addressing density (high ratio of facts about concepts) and breadth (large coverage of physical phenomena). In fact, it is very rich for domains such [illegible] encodes this knowledge from an encyclopaedic perspective. The ultimate goal of our research is to enrich LOD with com[illegible] We claim that an important gap to be filled towards this goal is: assessing foundational distinctions over LOD entities, that is to distinguish and formally assert whether a LOD entity inherently refers to e.g. a class or an individual, a physical object or not, a location, a social object, etc., from a common sense perspective.

1.1 Foundational Distinctions

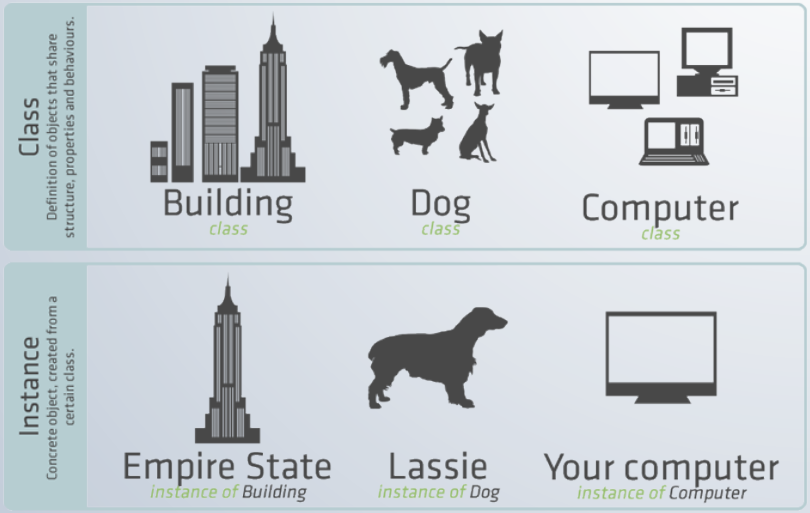
High level categorial distinctions (e.g. class vs. instance) are a fundamental human cognitive ability: "There is nothing more basic than categorization to our thought, perception, action, and speech." [Lakoff, 1987]. This is also why "the organisation of objects into categories is a vital part of knowledge representation" [Russell and Norvig, 2009]. Foundational distinctions have been theorised and modelled in foundational ontologies such as DOLCE [Masolo *et al.*, 2003] and SUMO [Pease and Niles, 2002] with a top-down approach, but populating and empirically validating them has been rarely addressed. In this study, we perform a set of experiments to assess *whether machines can learn to perform foundational distinctions, and if they match common sense*.

Do foundational distinctions match common sense?

Do they emerge from LOD? Can they be learned/predicted?

Class vs Instance

- Fundamental in formal ontology
- Basis of KR formalisms (RDF and OWL)
- Support taxonomic reasoning



Physical Object vs ¬Physical Object

- Essential to represent the physical world
- One of the primary distinctions in DOLCE

SEMANTiCS
Karlsruhe 2019

Class vs Instance

RESULT

- 1943 entities were classified as class (~44% of the dataset)
- 2519 entities were classified as instance (~56% of the dataset)

Cl_E: Agreement: 93,6 %

Cl_C: Agreement: 95,76 %

- 22510 judgments collected
- 117 different contributors

Experts-Crowd Agreement: 95,7%

Physical object

RESULT

- 3055 entities were classified as physical object
- 1447 entities were classified as not physical objects

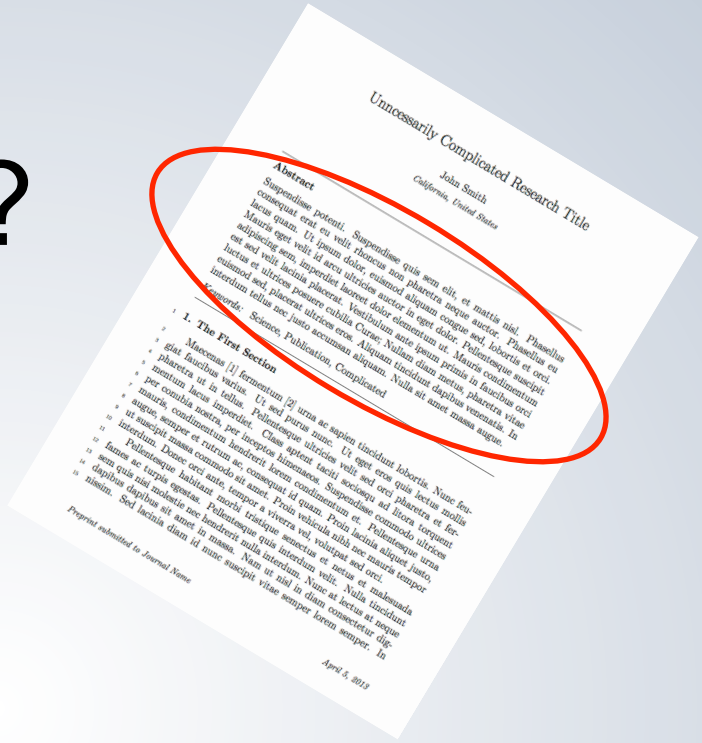
PO_E: Agreement: 93,9 %

PO_C: Agreement: 85,48 %

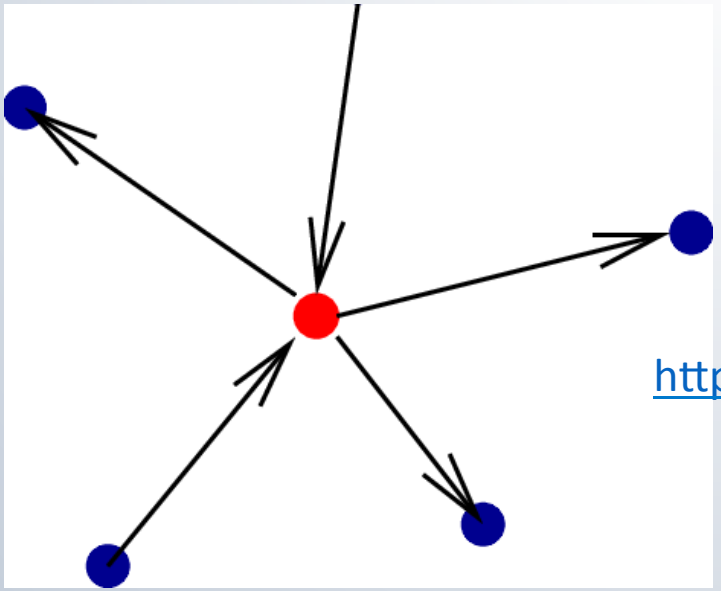
- 22510 judgments collected
- 117 different contributors

Experts-Crowd Agreement: 85,69%

Which features?

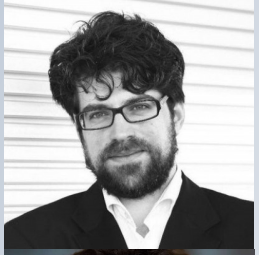


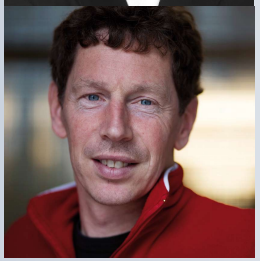
- Abstract
- Incoming/Outgoing properties



- URI id (for the class vs. instance classification)

http://dbpedia.org/page/Umbria_Jazz_Festival





Detecting Erroneous Identity Links on the Web Using Network Metrics

Joe Raad[1,3(✉)], Wouter Beek[2], Frank van Harmelen[2], Nathalie Pernelle[3], and Fatiha Saïs[3]

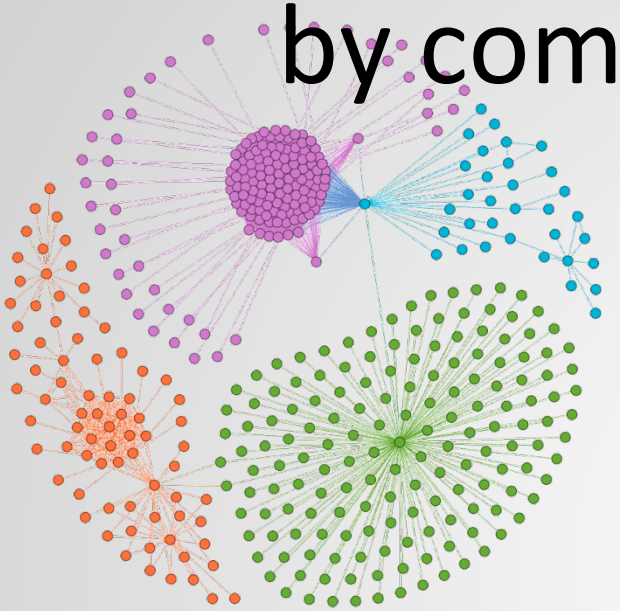
[1] UMR MIA-Paris, INRA, Paris-Saclay University, Paris, France
joe.raad@agroparistech.fr
[2] Department of Computer Science, VU University Amsterdam, Amsterdam, The Netherlands
{w.g.j.beek,frank.van.harmelen}@vu.nl
[3] LRI, Paris Sud University, CNRS 8623, Paris Saclay University, Orsay, France
{nathalie.pernelle,fatiha.sais}@lri.fr

Erroneous use of formal constructs may express a call for richer formal languages

Abstract. In the absence of a central naming authority on the Semantic Web, it is common that different names are used to refer to the same thing by different IRIs. Whenever multiple names are used to denote the same thing, `owl:sameAs` statements are needed in order to link the data and foster reuse. Studies that date back as far as 2009, have observed that the `owl:sameAs` property is sometimes used incorrectly. In this paper, we show how network metrics such as the community structure of the `owl:sameAs` graph can be used in order to detect such possibly erroneous statements. One benefit of the here presented approach is that it can be applied to the network of `owl:sameAs` links itself, and does not rely on any additional knowledge. In order to illustrate its ability to scale, the approach is evaluated on the largest collection of identity links to date, containing over 558M `owl:sameAs` links scraped from the LOD Cloud.

Keywords: Linked Open Data · Identity · owl:sameAs · Communities

Debugging identity by community detection



Communities correspond to roles:

- Person
- Senator
- President
- Government

Community 0

1. dbpedia.org/resource/B_hussein_obama
2. dbpedia.org/resource/Barack_H_Obama,_Jr
3. dbpedia.org/resource/Barak_hussein_obama
4. dbpedia.org/resource/President_Barack
5. dbpedia.org/resource/Senator_Barack_Obama
6. dbpedia.org/resource/Obama

...

99. dbpedia.org/resource/Hussein_Obama

Community 3

1. dbpedia.org/resource/Presidency_of_Barack_Obama
2. dbpedia.org/resource/Barack_Obama_Administration
3. dbpedia.org/resource/Barack_Obama_Cabinet
4. dbpedia.org/resource/Obama_White_House
5. dbpedia.org/resource/Obama_regime
6. dbpedia.org/resource/America_under_Obama

...

52. dbpedia.org/resource/Presidential_transition_of_Barack_Obama

From Frank van Harmelen's talk at ISWS 2019

Message from Empirical Semantics

It's not the users that got owl:sameAs wrong,
It's the formal semantics that got reality wrong

Challenge:
What alternative semantic model of equality would fit the empirically observed usage better?

From Frank van Harmelen's talk at ISWS 2019

http://www.ontologydesignpatterns.org/ekp/owl/

Encyclopedic Knowledge Patterns from Wikipedia Links

Andrea Giovanni Nuzzolese[1,2], Aldo Gangemi[1],
Valentina Presutti[1], and Paolo Ciancarini[1,2]

[1] STLab-ISTC Consiglio Nazionale delle Ricerche, Rome, Italy
[2] Dipartimento di Scienze dell'Informazione, Università di Bologna, Italy

The distribution of web links between wikipages tells us how to model entity classes

Abstract. What is the most intuitive way of organizing concepts for describing things? What are the most relevant types of things that people use for describing other things? Wikipedia and Linked Data offer knowledge engineering researchers a chance to empirically identifying invariances in conceptual organization of knowledge i.e. knowledge patterns. In this paper, we present a resource of Encyclopedic Knowledge Patterns that have been discovered by analyzing the Wikipedia page links dataset, describe their evaluation with a user study, and discuss why it enables a number of research directions contributing to the realization of a meaningful Semantic Web.

See references

Input data

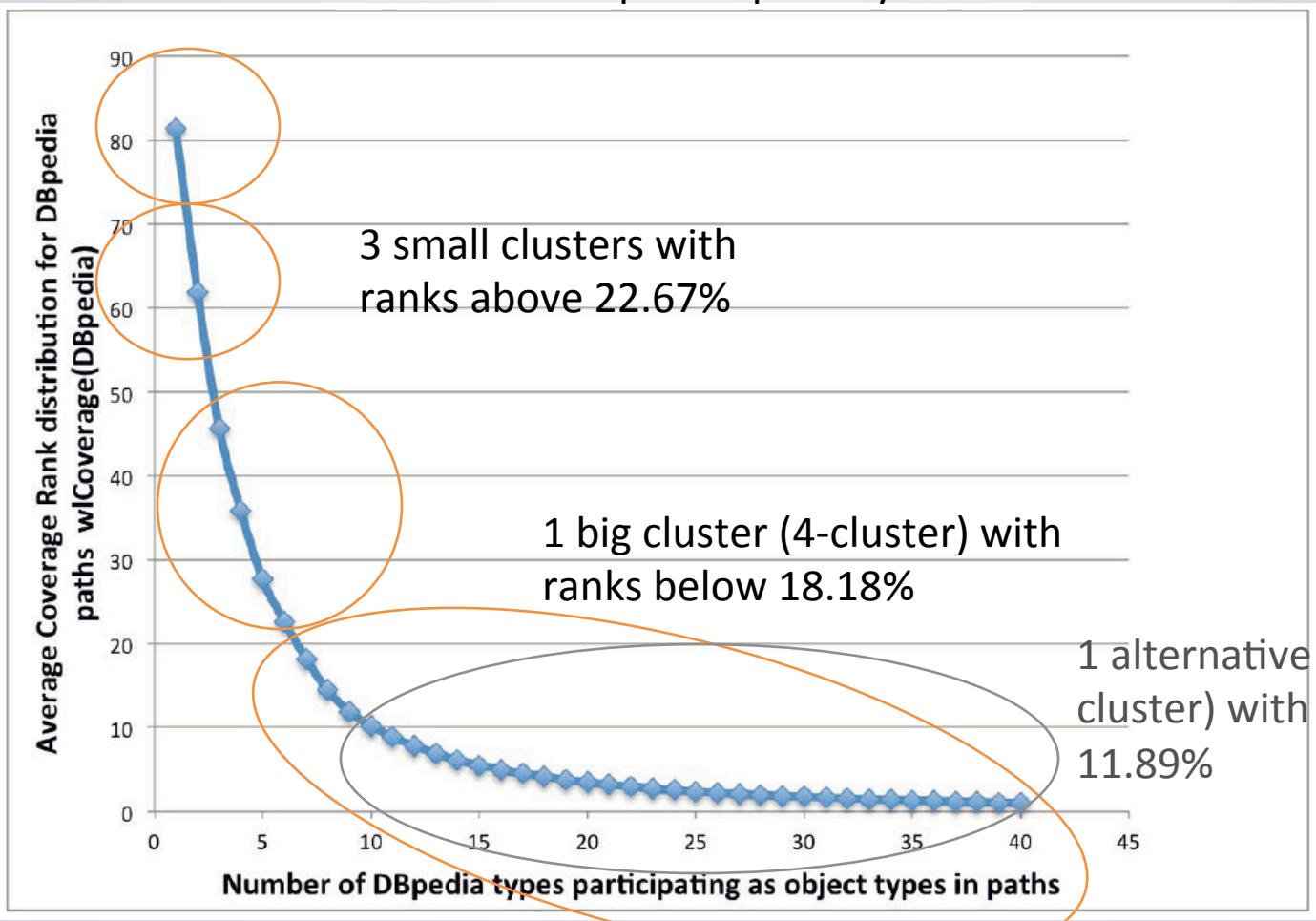
- Wikipedia page links generate 107.9M triples
- Infobox-based triples are 13.6M, including data value triples (9.4M)
- “Unmapped” object value triples are only 7% of page links

Paths are used to discover *Encyclopedic Knowledge Patterns*

Such patterns should make it emerge the most typical types of things that the Wikipedia crowd uses to describe a resource of a given type

k-means clustering on Path Popularity

Sample distribution of pathPopularity for DBpedia paths. The x-axis indicates how many paths (on average) are above a certain value *t* for pathPopularity



User study: Inter-rater agreement

Average coefficient of concordance for ranks Kendall's W

Group 1	**0.700**
Group 2	0.665

Kendall's W (for all values $p < 0.0001$)

Reliability test: Cronbach's alpha

DBPO class	Agreement	Reliability	DBPO class	Agreement	Reliability
Language	0.836	0.976	Philosopher	0.551	0.865
Writer	0.749	0.958	Ambassador	0.543	0.915
Legislature	0.612	0.888	Album	0.800	0.969
Radio Station	0.680	0.912	Administrative Region	0.692	0.946
Country	0.645	0.896	Insect	0.583	0.929
Disease	0.823	0.957	Aircraft	0.677	0.931

Kendall's W range [0,1]
0 = no agreement
1 = complete agreement

Take home message

- The Semantic Web as a global public knowledge graph of commonsense
- There are many commonsense resources that can be reused, and many open challenges
- It's time to address the hard problems
  - Unified formal semantics for commonsense
  - Criteria for maximising coverage
  - Evaluation methods
- Empirical semantics may uncover implicit commonsense in the Semantic Web
  - Perform observations on large scale


THAT'S ALL
FOLKS !

References 1/2

Mathieu d'Aquin, Enrico Motta, Marta Sabou, Sofia Angeletou, Laurian Gridinoc, Vanessa López, Davide Guidi:Toward a New Generation of Semantic Web Applications. IEEE Intelligent Systems 23(3): 20-28 (2008)

Piero Andrea Bonatti, Stefan Decker, Axel Polleres, Valentina Presutti:
Knowledge Graphs: New Directions for Knowledge Representation on the Semantic Web (Dagstuhl Seminar 18371). Dagstuhl Reports8(9): 29-111 (2018)

Maarten Sap, Ronan Le Bras, Emily Allaway, Chandra Bhagavatula, Nicholas Lourie, Hannah Rashkin, Brendan Roof, Noah A. Smith, Yejin Choi: ATOMIC: An Atlas of Machine Commonsense for If-Then Reasoning. AAAI 2019: 3027-3035

Cade Metz, “Paul Allen Wants to Teach Machines Common Sense,” The New York Times, 28 February 2018. https://www.nytimes.com/2018/02/28/technology/paul-allen-ai-common-sense.html

John McCarthy: Some expert systems need common sense. Annals of the New York Academy of Science, 1983. Invited presentation for the New York Academy of Sciences ScienceWeek Symposium on Computer Culture, April 5-8, 1983.

Doug Lenat, Mayank Prakash, and Mary Shepherd: CYC: Using Common Sense Knowledge to Overcome Brittleness and Knowledge Acquisition Bottlenecks. AI Magazine 6(4): 65-85 (1985)

Liang-Jun Zang, Cong Cao, Ya-Nan Cao, Yu-Ming Wu, and Cun-Gen CAO A survey of commonsense knowledge acquisition. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 28(4): 689-719 July 2013. DOI 10.1007/s11390-013-1369-6

Collin F Baker, Fillmore, Charles J, and Lowe, John B. 1998. The Berkeley FrameNet project. In COLING-ACL '98: Proceedings of the Conference, Montreal, Canada

References 2/2

Ernest Davis, Gary Marcus:Commonsense reasoning and commonsense knowledge in artificial intelligence. Commun. ACM 58(9): 92-103 (2015)

Joe Raad, Wouter Beek, Frank van Harmelen, Nathalie Pernelle, Fatiha Saïs: Detecting Erroneous Identity Links on the Web Using Network Metrics. International Semantic Web Conference (1) 2018: 391-407

Andrea Giovanni Nuzzolese, Aldo Gangemi, Valentina Presutti, Paolo Ciancarini:Encyclopedic Knowledge Patterns from Wikipedia Links. International Semantic Web Conference (1) 2011: 520-536

Luigi Asprino, Valerio Basile, Paolo Ciancarini, Valentina Presutti:Empirical Analysis of Foundational Distinctions in Linked Open Data. IJCAI 2018: 3962-3969. https://w3id.org/fox

Steven de Rooij, Wouter Beek, Peter Bloem, Frank van Harmelen, Stefan Schlobach:Are Names Meaningful? Quantifying Social Meaning on the Semantic Web. International Semantic Web Conference (1) 2016: 184-199

Robert Speer, Catherine Havasi: Representing General Relational Knowledge in ConceptNet 5. LREC 2012: 3679-3686

Tom M. Mitchell et al: **Never-Ending Learning.** AAAI 2015: 2302-2310

Fredo Erxleben, Michael Günther, Markus Krötzsch, Julian Mendez, Denny Vrandecic:Introducing Wikidata to the Linked Data Web. International Semantic Web Conference (1) 2014: 50-65

Paolo Pareti, Ewan H. Klein (2016). The Human Know-How Dataset, 2014 [dataset]. https://doi.org/10.7488/ds/1394

Heiko Paulheim and Aldo Gangemi Serving DBpedia with DOLCE – More than Just Adding a Cherry on Top. Proceedings of ISWC2015, the Thir.teenth International Semantic Web Conference, LNCS, Springer, 2015

DBpedia http://dbpedia.org/

- The Linked Open Data version of Wikipedia, mainly its inboxes
- Its English version describes ~4.6M things
- Localised versions in 125 languages
  - Altogether they describe 38.3M things
- Predominance of instances vs general concepts
- The DBpedia Ontology includes 685 classes and ~2.7k properties and 4.2M instances

About: Kettle

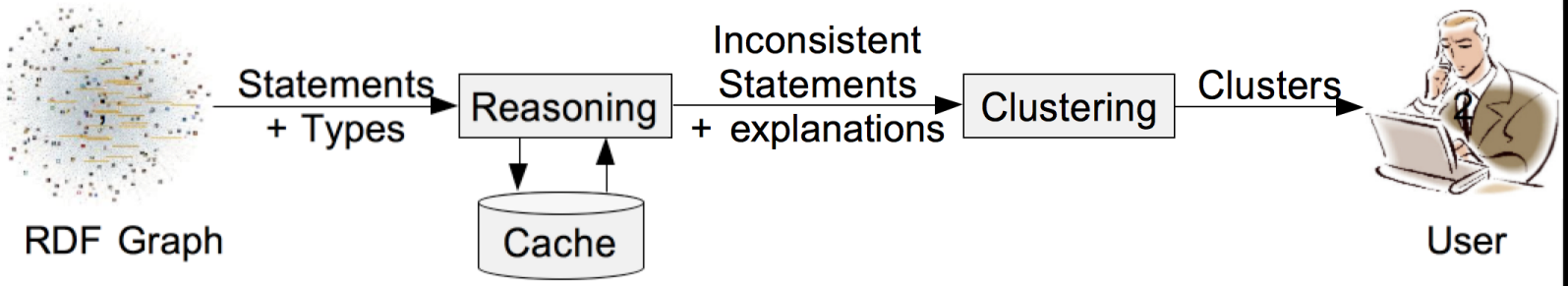
An Entity of Type : beverage, from Named Graph : http://dbpedia.org, within Data Space : dbpedia.org

A kettle, sometimes called a tea kettle or teakettle, or a water hotter is a type of pot, usually metal, specialized for boiling water, with a lid, sprouts and handle, or a small bathroom appliance of similar shape that functions in a self-pertained manner. Kettles can be heated either by placing on a wire brush, or by their own internal electric heating element in the appliance versions.

plated or solid), wood, porcelain or plastic.

Overall Workflow

- Overall, we find 3,654,255 inconsistent statements (24.4%)
  - cf.: only 97,749 (0.7%) without DOLCE
- Too much to inspect!
  - We are looking for *systematic* errors
  - Cluster explanations w/ DBSCAN
  - Each cluster represents a systematic error



Paulheim, H. and Gangemi, A. Serving DBpedia with DOLCE – More than Just Adding a Cherry on Top. Proceedings of ISWC2015, the *Thirteenth International Semantic Web Conference*, LNCS, Springer, 2015

Karlsruhe
Carlsruhe

Karlsruhe Palace, view over Karlsruhe, Schlossplatz, Konzerthaus, Crown of Baden

Flag

Coat of arms

Location of Karlsruhe [show]

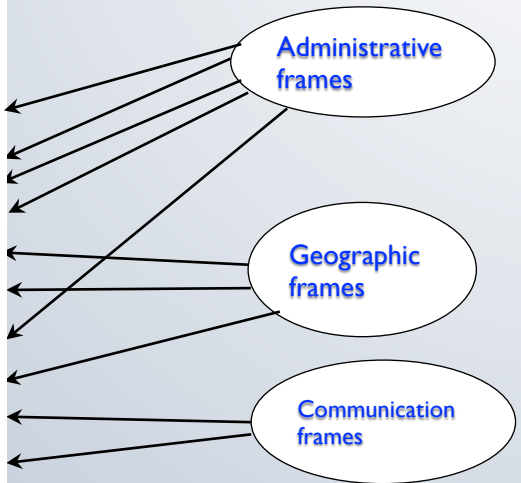
Show map of Germany
Show map of Baden-Württemberg
Show all
Coordinates: 49°00′33″N 8°24′14″E

Country	Germany
State	Baden-Württemberg
Admin. region	Karlsruhe
District	Urban district
Founded	1715
Subdivisions	27 quarters
Government	
· **Lord Mayor**	Frank Mentrup (SPD)
Area	
· **Total**	173.46 km² (66.97 sq mi)
Elevation	115 m (377 ft)
Population (2018-12-31)[1]	
· **Total**	313,092
· **Density**	1,800/km² (4,700/sq mi)
Time zone	CET/CEST (UTC+1/+2)
Postal codes	76131–76229
Dialling codes	0721
Vehicle registration	KA
Website	www.karlsruhe.de

dbo:areaCode	▪ 0721 ▪ KA
dbo:areaTotal	▪ 173460000.000000 (xsd:double)
dbo:country	▪ dbr:Germany
dbo:district	▪ dbr:Urban_districts_of_Germany
dbo:division	▪ dbr:Quarter_(country_subdivision)
dbo:elevation	▪ 115.000000 (xsd:dou
dbo:federalState	▪ dbr:Baden-Württem
dbo:foundingYear	▪ 1715-01-01 (xsd:date
dbo:leaderTitle	▪ Oberbürgermeister
dbo:populationAsOf	▪ 2007-12-31 (xsd:date
dbo:populationTotal	▪ 288917 (xsd:integer)
dbo:postalCode	▪ 76131–76229

DBpedia

geo:lat	▪ 49.000000 (xsd:float) ▪ 49.009209 (xsd:float)
geo:long	▪ 8.400000 (xsd:float) ▪ 8.403952 (xsd:float)
prov:wasDerivedFrom	▪ wikipedia-en:Karlsruhe?oldid=741639281
foaf:depiction	▪ wiki-commons:Special:FilePath/Karlsruhe_town_centre_air.jpg
foaf:homepage	▪ http://www.karlsruhe.de
foaf:isPrimaryTopicOf	▪ wikipedia-en:Karlsruhe
foaf:name	▪ Karlsruhe (en) ▪ (Carlsruhe) (en)
is dbo:administrativeDistrict of	▪ dbr:Brühl_(Baden)

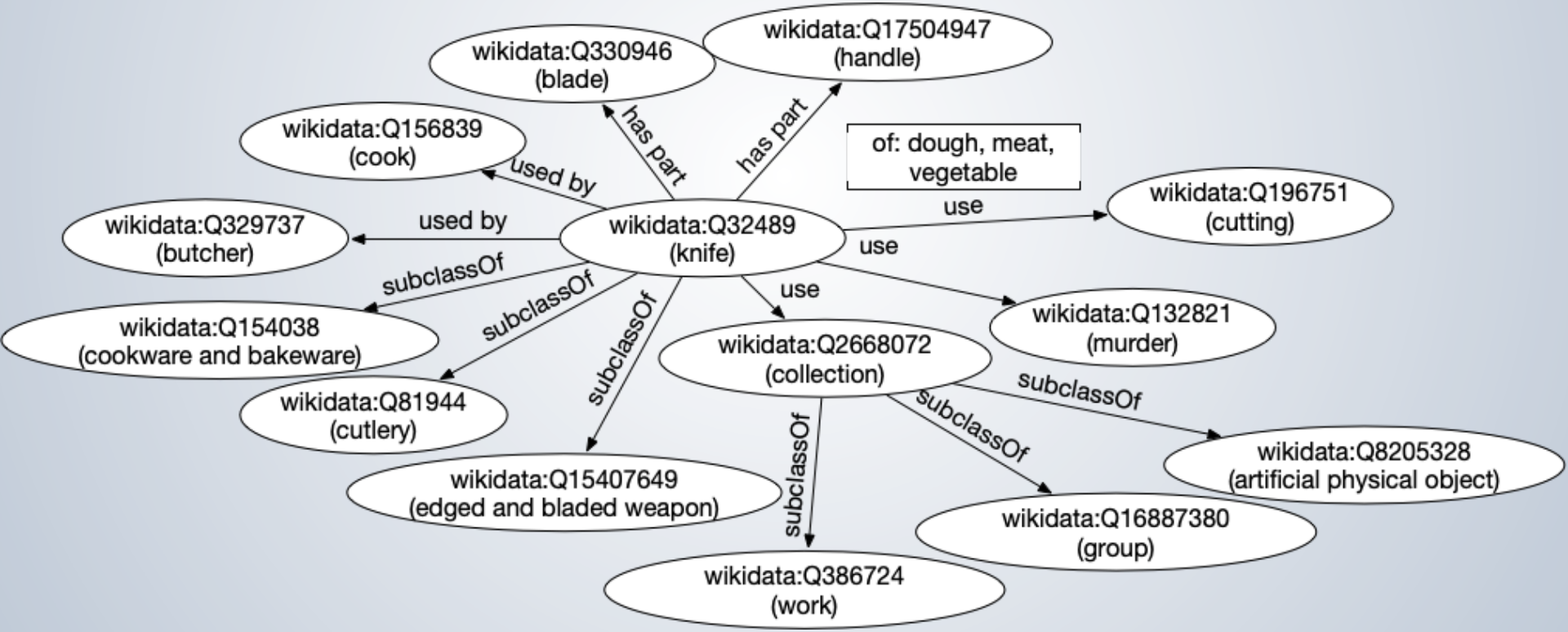


When triplifying Wikipedia infoboxes, its designers lost the framing of boxes and internal sub-boxes

Wikidata https://www.wikidata.org/

- The community-created knowledge base of Wikipedia
- Launched in 2012, it has collected ~15M entities, ~34M statements, ~80M labels and descriptions in 350+ languages, 40k+ registered contributors
- Statements provenance and *qualifiers:* which provide some *context* data
- Export service to LOD

Wikidata


wikidata:Q330946
(blade)
wikidata:Q17504947
(handle)
has part
has part
wikidata:Q156839
(cook)
used by
of: dough, meat,
vegetable
wikidata:Q329737
(butcher)
used by
wikidata:Q32489
(knife)
use
wikidata:Q196751
(cutting)
use
subclassOf
subclassOf
use
wikidata:Q154038
(cookware and bakeware)
wikidata:Q132821
(murder)
wikidata:Q2668072
(collection)
subclassOf
wikidata:Q81944
(cutlery)
subclassOf
subclassOf
subclassOf
wikidata:Q8205328
(artificial physical object)
wikidata:Q15407649
(edged and bladed weapon)
subclassOf
wikidata:Q16887380
(group)
wikidata:Q386724
(work)